

Modelado de series temporales con sistemas de inferencia borrosa multidimensionales

Por
Juan Fernando Rendón Sánchez

Director de tesis
Juan David Velásquez Henao, M.Sc.,Ph.D.

Tesis enviada a la
Escuela de Sistemas
como requerimiento parcial
para el grado de
Magister en Ingeniería - Ingeniería de Sistemas

Facultad de Minas
Universidad Nacional de Colombia

Junio 2009

A mi madre y a la memoria de mi padre.

Agradecimientos

Agradezco al profesor Juan David Velásquez por su apoyo y por la invaluable orientación durante la ejecución de este trabajo investigativo. A la Universidad Nacional de Colombia por su programa de becas para la realización de estudios de posgrado. Agradezco también a mis compañeros de trabajo y estudio por su constante apoyo.

Resumen

La predicción de series que exhiben características no lineales ha sido un problema vigente durante las últimas dos décadas. En este trabajo se explora este problema a partir del uso de sistemas de inferencia borrosa y, particularmente, de sistemas basados en conjuntos borrosos multidimensionales. Un paso fundamental para especificar este tipo de sistemas es la detección de clusters o grupos en los datos, ya que las técnicas de agrupamiento borroso han servido para agilizar el proceso de identificación de los modelos. Las aplicaciones de estos sistemas en el área de series temporales son escasas y su tratamiento formal estadístico es mínimo o inexistente. Más aún, existen tópicos concernientes a su construcción que no se han tratado en profundidad. En este trabajo se discuten los métodos existentes para el modelado de series temporales y los adelantos hechos en la identificación de sistemas de inferencia borrosa con conjuntos borrosos multidimensionales. Se realiza una propuesta metodológica para formalizar el uso de estos sistemas en el modelado y predicción de series temporales y se proponen mejoras a algunos tópicos de diseño. Finalmente se muestran algunas aplicaciones con series de tipo *benchmark* para probar la capacidad de generalización del modelo propuesto frente a otros y se utiliza una serie real de demanda eléctrica en Colombia para observar en qué medida ayuda el modelo en la comprensión de dicho problema.

Contenido

Resumen	iv
Lista de tablas	viii
Lista de figuras	ix
1 Antecedentes y definición del problema	1
1.1 Introducción	1
1.2 Aspectos generales de las series temporales	2
1.3 Aproximaciones al modelado con series temporales	3
1.3.1 Modelos ARIMA	3
1.3.2 Algunos modelos paramétricos no lineales de series temporales	4
1.3.3 Modelos no paramétricos pertenecientes a la inteligencia com- putacional	6
1.3.4 Evaluación de un modelo de predicción	8
1.4 Sistemas de inferencia borrosa	12
1.4.1 Introducción	12
1.4.2 Conceptos básicos	12
1.4.3 Tendencias en el modelado con sistemas de inferencia borrosa	14
1.5 Sistemas de inferencia borrosa multidimensionales	17
1.6 Requerimientos, objetivos e hipótesis de investigación	23
1.6.1 Requerimientos y limitaciones	23
1.6.2 Hipótesis de investigación	23
1.6.3 Objetivos	24
1.6.3.1 Objetivos específicos	24
1.6.3.2 Objetivo general	24
1.6.4 Aportes que se espera obtener	24
2 Aspectos metodológicos	26
2.1 Introducción	26
2.2 Estructura del modelo	27
2.2.1 Sistema de inferencia	27
2.2.2 Algoritmo	29

2.2.2.1	Algoritmo general para la obtención de modelos de inferencia	30
2.3	Metodología y entrenamiento	33
2.3.1	Identificación	33
2.3.1.1	Determinación de los regresores	33
2.3.1.2	Determinación de las propiedades estadísticas de la serie	33
2.3.1.3	Transformación de la serie temporal	35
2.3.2	Especificación del sistema de inferencia y estimación de parámetros	35
2.3.2.1	Criterio de parada	35
2.3.3	Diagnóstico	36
2.3.4	Metadiagnóstico	37
2.3.5	Construcción de pronósticos e intervalos de confianza	37
2.4	Propiedades del modelo	38
2.5	Conclusiones	40
3	Ejemplificación del uso de la metodología	41
3.1	Introducción	41
3.2	Identificación	41
3.2.1	Análisis exploratorio de los datos	41
3.2.2	Algunos estadísticos descriptivos	42
3.2.3	Indicaciones sobre la selección de variables	43
3.2.4	Estudio de linealidad en la serie	43
3.3	Especificación del sistema de inferencia y estimación de parámetros	44
3.4	Diagnóstico	44
3.5	Metadiagnóstico	46
3.6	Construcción de pronósticos e intervalos de confianza	49
3.7	Conclusiones	49
4	Pronóstico de series benchmark	50
4.1	Introducción	50
4.2	The Airline Passenger data set	51
4.3	The Sunspot data set	54
4.4	The Pollution Equipment data set	55
4.5	Conclusiones	59
5	Modelado de la demanda mensual de electricidad en Colombia	60
5.1	Introducción	60
5.2	Modelo MFIS	63
5.3	Contraste de resultados con otros modelos no lineales	67
5.4	Conclusiones	69
6	Conclusiones y trabajo futuro	70

Bibliografía	75
Anexos	81

Lista de tablas

3.1	Estadísticos descriptivos.	43
3.2	Modelos con menor AIC para la serie de datos log Airline.	45
3.3	Variaciones de modelos seleccionados.	46
3.4	Modelos seleccionados en la etapa de diagnóstico.	48
4.1	Desempeño de modelos para la serie de datos Airline.	53
4.2	Desempeño de modelos para la serie de datos Sunspot.	56
4.3	Desempeño de modelos para la serie de datos Pollution equipment.	58
5.1	Diferentes modelos para la demanda de energía eléctrica (tomado de [20]).	62
5.2	Desempeño de modelos MFIS para la serie de demanda de electricidad.	64
5.3	Especificación del modelo MFIS-7.	65
5.4	Desempeño de varios modelos para la serie de demanda de electricidad.	68

Lista de figuras

1.1	Partición para un sistema de inferencia borroso.	13
1.2	Sistema de inferencia borrosa tipo Takagi-Sugeno con varias reglas. . .	14
1.3	Una clasificación general de las tendencias en modelos de inferencia borrosa para el análisis de series temporales.	18
1.4	Función de pertenencia para un conjunto borroso multidimensional. . .	19
1.5	El agrupamiento como mecanismo de identificación.	20
2.1	Partición y funciones base asociadas.	32
3.1	Serie Airline.	42
3.2	Histograma del logaritmo de los datos en la serie Airline.	42
3.3	Autocorrelogramas para la serie Airline.	43
3.4	Herramientas gráficas de diagnóstico para el modelo 752-1.	47
3.5	Predicción para el modelo 162-2.	49
4.1	Serie Airline.	52
4.2	Predicción para la serie de datos Airline.	54
4.3	Serie Sunspot.	55
4.4	Predicción para la serie de datos Sunspot.	56
4.5	Serie Pollution equipment.	57
4.6	Predicción para la serie de datos Pollution equipment.	58
5.1	Demanda mensual de electricidad (miles de GWh-mes).	63
5.2	Predicción para la serie de datos de la demanda eléctrica mensual en Colombia. Modelo MFIS-7.	65
5.3	Agrupamiento de datos realizado por el modelo MFIS-7.	66

Capítulo 1

Antecedentes y definición del problema

1.1 Introducción

Las justificaciones de la utilidad del pronóstico son claras y abundantes en la literatura. Todas ellas convergen en la ventaja que representa el conocimiento previo de la ocurrencia de un evento. Tal conocimiento se hace muchas veces no sólo útil, sino necesario. Su uso beneficia diversas áreas, como las finanzas, las ciencias sociales, la ingeniería y la economía.

El pronóstico, según [40], consiste en la generación de información sobre el posible desarrollo futuro de un proceso, con base en datos de su comportamiento pasado y presente. Se distingue de la labor de modelamiento (que se centra en el entendimiento del comportamiento histórico de un proceso), ya que su propósito es la obtención de valores futuros desconocidos. La labor de modelamiento ofrece beneficios como la profundización en el conocimiento del sistema estudiado y la posibilidad de confrontar el conocimiento de los expertos con los resultados de los modelos y los datos. También se facilita la aproximación sistémica y estructurada al problema y la mejora de los procesos decisorios. El pronóstico, por su lado, ofrece ventajas en la visualización, con cierto grado de certidumbre, del comportamiento venidero de la variable bajo estudio con el consecuente beneficio para las organizaciones o los individuos.

En las tareas de modelamiento o de pronóstico se pueden encontrar fácilmente situaciones en las que no se posea el conocimiento experto suficiente para formular un modelo con el cual sea plausible describir la dinámica del sistema de interés¹. Cuando esto sucede y se dispone solamente de datos observados del sistema a lo

¹En otras situaciones, como aquellas en las que los datos disponibles no son adecuados para un análisis cuantitativo o en los que la información cualitativa tiende a incrementar la precisión, la relevancia o la aceptación de los pronósticos, son convenientes otras herramientas de pronóstico, tales como el *Juicio sin Ayuda*, el *Juego de Roles* o el *Análisis Conjunto*, entre otros. De ellas puede encontrarse un resumen en [5].

largo del tiempo, es de gran ayuda poder construir modelos empíricos que ayuden a describir la dinámica que gobierna el proceso. Las series temporales han mostrado ser una herramienta de suma utilidad en la construcción de tales modelos empíricos (véase [66]).

El modelado a través de series temporales se ha realizado con diversas herramientas, entre las cuales están los métodos estadísticos, las redes neuronales, los sistemas de inferencia borrosa y algunos métodos combinados. Las redes neuronales y los sistemas de inferencia borrosa han recibido una atención considerable en años recientes, dada su versatilidad, su reconocida capacidad de aproximación de funciones no lineales y los resultados alentadores en el área de pronóstico (véase por ejemplo [1], [14], [13]). En estos modelos, la inexistencia de limitantes como la necesidad de especificar una forma funcional a priori en problemas de regresión los torna útiles en situaciones en las que se carece de la información suficiente sobre el sistema o fenómeno bajo estudio como para proponer una forma funcional adecuada.

En el campo de los sistemas de inferencia borrosa existen variantes específicas, entre las cuales están los sistemas de inferencia basados en conjuntos borrosos multidimensionales. Estos modelos han surgido como respuesta a algunos inconvenientes que se han detectado en la identificación de sistemas de inferencia a partir de datos, entre los cuales se encuentra el aumento exponencial del número de reglas (y el consecuente aumento del número de parámetros) a medida que se incrementa la cantidad de variables de entrada. El advenimiento de estos modelos específicos ha abierto varios interrogantes con respecto a su construcción y con respecto a su uso en el modelado y pronóstico de series temporales. En ellos se centra la atención de esta investigación.

En las siguientes secciones se darán unos conceptos generales de series temporales. Se mostrarán posteriormente las aproximaciones al modelamiento de las mismas y las metodologías que han ayudado en el uso de los modelos. Se discutirán los sistemas de inferencia borrosa y se examinarán los modelos específicos basados en conjuntos borrosos multidimensionales. Se plantearán algunos interrogantes sobre su uso en el área de series temporales y, finalmente, se establecerá un conjunto de objetivos que ayudarán a delimitar y responder algunos de los interrogantes de interés.

1.2 Aspectos generales de las series temporales

Una serie temporal está definida como un conjunto de observaciones ordenadas en el tiempo

$$\dots, y_{-1}, y_0, y_1, \dots, y_T, y_{T+1} \tag{1.1}$$

de las cuales sólo se dispone de la realización $\{y\}_1^T$. Dichas observaciones son el resultado de la aplicación de un mecanismo generador a un conjunto de variables causales que explican el comportamiento de y_t .

La hipótesis de la modelación es que dicho mecanismo puede ser representado por una función computable y desconocida $f(\cdot)$ que es aplicada al conjunto de regresores $\mathbf{x}_t = \{x_t^{(1)}, \dots, x_t^{(U)}\}$:

$$Y_t = f(\mathbf{x}_t) + e_t \quad (1.2)$$

donde el término e_t representa una serie de ruidos aleatorios independientes e idénticamente distribuidos, que siguen una distribución con media cero y varianza σ_v^2 . Los regresores x_t en 1.2 están formados por un subconjunto de las variables causales visibles y cuantificables, que influyen el mecanismo generador de las observaciones. La serie de ruidos estocásticos, $\{e_t\}_1^T$, representa el efecto de todas aquellas variables causales, cuantitativas o cualitativas, visibles o no visibles, no incluidas en el conjunto de regresores.

Partiendo de la premisa de que $f(\cdot)$ existe y es computable, el objetivo fundamental de la modelación de series temporales es construir una función $\hat{f}(\cdot)$ que la aproxime con algún grado de precisión valiéndose de los datos disponibles $\{y_t, \mathbf{x}_t\}_1^T$, tal que

$$Y_t = \hat{f}(\mathbf{x}_t) + \hat{e}_t \quad (1.3)$$

Así, dicho modelo empírico puede ser entendido como una hipótesis, teoría o ley que define la probabilidad de transición de una observación y_t a la siguiente o y_{t+1} , de tal forma que la dinámica de la serie temporal estudiada es completamente especificada por 1.3. Como un modelo probabilístico, es caracterizado por sus propiedades estadísticas tales como su densidad de probabilidad, su correlación o su espectro de energía. Se supone que la información disponible es suficiente para estimar la función $\hat{f}(\cdot)$ con la precisión requerida.

1.3 Aproximaciones al modelado con series temporales

El conjunto de modelos utilizados para el pronóstico de series temporales comprende los lineales y los no lineales. Y dentro de los no lineales se encuentran los paramétricos y los no paramétricos (véase [27]). Dada la gran cantidad de modelos existentes, especialmente no lineales, se mostrarán aquí sólo algunos y su discusión será general.

1.3.1 Modelos ARIMA

Estos modelos tienen una trayectoria que puede extenderse hasta los años 50, cuando George Box y Gwilym Jenkins integraron sus elementos teóricos y los popularizaron. Se han estudiado ampliamente y las teorías estadística y matemática en las que se basan están bastante desarrolladas y entendidas (la teoría de secuencias Gaussianas

y la inferencia estadística para modelos Gaussianos lineales, en la estadística, y las ecuaciones lineales en diferencias en el campo matemático). Estos modelos tienen una componente autorregresiva (AR) que permite describir un comportamiento estocástico en el que y_t se modela como una suma ponderada de observaciones pasadas. Por otro lado, la componente de medias móviles (MA) describe comportamientos en los que los choques aleatorios que ingresan al modelo, persisten y afectan el valor de y_t durante uno o varios períodos. La componente de integración (I) especifica la necesidad de diferenciar la serie cuando las fluctuaciones observadas se deben a efectos acumulados de choques aleatorios (véase [50]).

A pesar de que estos modelos han sido exitosos como herramientas prácticas de análisis, son una aproximación lineal y la mayoría de los fenómenos de interés presentan características no lineales. Además, dado que tienen distribuciones conjuntas simétricas, no son ideales para datos que presentan mucha asimetría. Tampoco son ideales para datos que tienen cambios repentinos de gran amplitud, de manera irregular.

Metodología general para la utilización de modelos ARIMA

La construcción de modelos ARIMA de series temporales se realiza con la metodología formulada por Box y Jenkins (véase [50]). En ella se realizan los siguientes pasos:

1. Identificación de las especificaciones preliminares del modelo.
2. Estimación de los parámetros del modelo.
3. Diagnóstico de la bondad del modelo.
4. Metadiagnóstico del modelo.
5. Uso del modelo.

En la etapa de identificación se escoge el modelo particular que se utilizará. Luego se estiman los parámetros y se realiza un diagnóstico para determinar si el modelo cumple con algunos supuestos básicos. Si el modelo no es satisfactorio, se debe realizar un nuevo ciclo de identificación, estimación de parámetros y chequeo. Aquellos modelos que superen la etapa de diagnóstico se someten a un examen adicional, para verificar la calidad de los resultados que arroja. Finalmente, el modelo que supere los chequeos se utiliza para realizar pronósticos.

1.3.2 Algunos modelos paramétricos no lineales de series temporales

A continuación se mencionan algunos modelos paramétricos no lineales (véase por ejemplo [66]).

Modelos de umbral (*threshold models*) El propósito de estos modelos es realizar aproximaciones locales de una serie temporal. Esto se realiza descomponiéndolo en un grupo de modelos paramétricos no lineales. El cambio de régimen en una serie de tiempo es algo plausible, de manera que estos modelos dan una buena flexibilidad en tales casos. Dentro de esta clase de modelos se tienen los SETARMA y los STAR.

Modelos bilineales Están inspirados en los sistemas dinámicos y son útiles en situaciones en las que la entrada y salida de un sistema se conocen, pero el estado interno se desconoce. Este tipo de modelo es muy específico.

EXPAR (modelos exponenciales autorregresivos dependientes de la amplitud)

La dependencia de la amplitud es un término derivado de los modelos contruidos para explicar algunos fenómenos físicos (como las vibraciones) y refleja la importancia de considerar no linealidades en los modelos, cuando estas son importantes. La forma funcional de estos modelos es bastante específica y debe ajustarse para aplicaciones en campos diferentes de las vibraciones.

En general estos modelos, y otros que se clasifican dentro de los paramétricos, se pueden aplicar exitosamente si se tiene el suficiente conocimiento experto para elegir cuál debe usarse. Sin embargo, en muchos casos, la ley subyacente que gobierna la dinámica de la situación bajo estudio no se conoce y la utilidad práctica de estos modelos queda limitada.

Metodología para la utilización de modelos paramétricos no lineales

La metodología general para la aplicación de modelos paramétricos no lineales está basada en la propuesta por Box y Jenkins y ha sido modificada por algunos autores para su utilización en el caso no lineal (véase [64] y [66]). Los pasos fundamentales son los siguientes:

1. Identificación.
 - Selección de los regresores.
 - Propiedades estadísticas de la serie temporal.
 - Transformación de la serie temporal.
 - Especificación de la forma funcional del modelo.
2. Estimación de parámetros.
3. Diagnostico y análisis de la dinámica de la serie.
4. Metadiagnóstico.

5. Construcción de pronósticos e intervalos de confianza.

Las prácticas metodológicas involucradas en estos pasos incluyen aspectos de suma importancia, como las pruebas de linealidad, cuyo propósito es guiar al modelador en la decisión de usar un modelo no lineal o uno lineal. Se incluyen también prácticas que ayudan en la selección de modelos, dado que la amplitud de opciones paramétricas puede desembocar en un conjunto de numerosos modelos candidatos, que a su vez pueden tener diversas configuraciones de parámetros.

Dada la naturaleza empírica de los modelos de series temporales, se hace necesario disponer de herramientas que permitan juzgar objetivamente la calidad de los mismos. Al respecto, las metodologías mencionadas, para los modelos lineales y los no lineales, agregan un valor significativo para el modelador.

1.3.3 Modelos no paramétricos pertenecientes a la inteligencia computacional

Se mencionan a continuación los modelos básicos que han sido usados en el área de la inteligencia computacional para el modelado y predicción de series temporales no lineales. Su gran ventaja consiste en que tanto su estructura como sus parámetros se pueden ajustar. Características como esta son deseables en la construcción de modelos empíricos para los cuales no se dispone del suficiente conocimiento que permita formular una forma funcional *a priori*.

Redes neuronales Una red neuronal es un procesador masivo y paralelo compuesto por unidades simples de procesamiento, que tiene una propiedad natural para almacenar conocimiento experimental y hacerlo disponible para su uso (véase [30]). Esta tecnología se basa en áreas como las neurociencias, las matemáticas, la estadística, la física, las ciencias de la computación y la ingeniería. Sus campos de aplicación incluyen el análisis de series temporales, el reconocimiento de patrones y el procesamiento de señales. Existen varios tipos de redes neuronales:

Perceptrones de una capa Son la forma más simple de red neuronal. Se usan en la clasificación de patrones, pero tienen limitaciones computacionales con respecto a otros tipos de redes.

Perceptrones multicapa Han recibido considerable atención en el modelado y predicción de series temporales y se han realizado trabajos importantes en cuanto a la utilización de métodos estadísticos para realizar la selección de modelos (véase [3]).

Redes de funciones de base radial Han sido formuladas y estudiadas desde el contexto de la regresión no lineal en [30]. Allí se da un enfoque de tipo estadístico al estudio de este tipo de redes y a sus algoritmos de entrenamiento. Además, se han realizado trabajos que indirectamente involucran el uso de este tipo de redes en procesos de identificación de sistemas

no lineales (véase [58]). En estos procesos se han utilizado criterios estadísticos para la construcción del modelo.

Maquinas de soporte vectorial Al igual que los perceptrones multicapa y las redes de funciones de base radial, estas redes son apropiadas para la clasificación de patrones y para la regresión no lineal. Su fundamentación teórica es de tipo estadístico, como puede encontrarse en [30].

Máquinas de comité son redes neuronales que apelan al principio de *divide y vencerás* para realizar la labor de aprendizaje. Su uso incluye problemas de reconocimiento de patrones y de regresión. En algunos de sus esquemas de modelamiento se utilizan conceptos estadísticos para la estimación de los parámetros del modelo.

Mapas auto-organizados Estas redes neuronales se basan en aprendizaje competitivo (entre neuronas). Se caracterizan por la formación de un mapa topográfico de los patrones de entrada, en el cual las posiciones espaciales de las neuronas en la red indican características estadísticas intrínsecas contenidas en los patrones. Estas redes pueden usarse para tareas de clasificación, exploración de datos y monitoreo remoto.

Los pasos para el diseño de un modelo de pronóstico de redes neuronales son los siguientes (según [49]):

1. Selección de Variables.
2. Recolección de Datos.
3. Preprocesamiento de Datos.
4. Conjuntos de entrenamiento, validación y verificación.
5. Selección del modelo.
 - Número de capas ocultas.
 - Número de neuronas ocultas.
 - Número de neuronas de salida.
6. Criterio de Evaluación.
7. Entrenamiento de la Red Neuronal.
8. Implementación.

El creciente interés en las redes neuronales y su uso amplio en el pronóstico de series temporales ha motivado el surgimiento de aproximaciones metodológicas como ésta, que benefician la tarea de modelamiento al tratar particularidades de dichos modelos. Otro aporte valioso de tales indicaciones metodológicas es la disminución del esfuerzo que debe realizarse por el modelador, en ausencia de un conocimiento profundo, cuando desea incorporar las redes neuronales a su conjunto de herramientas.

Sistemas de inferencia borrosa Son sistemas que pueden capturar conocimiento experto en forma reglas vagas. También tienen la propiedad de realizar mapeos no lineales entre un espacio de entrada y un espacio de salida. En su formulación se utilizan conceptos de teoría de conjuntos borrosos, reglas borrosas del tipo *si-entonces* y razonamiento borroso. Además de la predicción de series temporales, se han aplicado en campos como el reconocimiento de patrones.

Sistemas ANFIS (*Adaptive neuro fuzzy inference system*) Son redes adaptativas funcionalmente equivalentes a los sistemas de inferencia borrosa. Fueron propuestos en [34] y han sido aplicados, entre otros, a problemas de control, modelado de series temporales, reconocimiento de patrones y procesamiento de señales. Algunos trabajos específicos en el modelado de series temporales con estos sistemas se encuentran en [52], [60] y [57].

1.3.4 Evaluación de un modelo de predicción

Dado que la cantidad de modelos para el análisis de series temporales ha crecido a través de los años, como se aprecia en las tendencias mencionadas en secciones anteriores, se han tornado importantes las herramientas que ayudan a su evaluación. A través de ellas es posible establecer qué tan bien reproduce el modelo las características de los datos y esto resulta esencial en la construcción de modelos empíricos: en ausencia del conocimiento profundo sobre el sistema estudiado, y frente a varios posibles modelos construidos, debe juzgarse minuciosamente cada uno de ellos.

A continuación se enumeran algunos criterios importantes para evaluar un modelo de predicción basado en series temporales. Algunos de ellos se han incorporado en las metodologías mencionadas para el uso de modelos de series temporales.

Verificación de variaciones alternativas del modelo Las variaciones alternativas de un modelo se obtienen ya sea agregando regresores que no son considerados actualmente o eliminando algunos de los regresores requeridos para representar la dinámica de la serie temporal. En la verificación debe demostrarse que los regresores adicionales no tienen una contribución significativa en el desempeño del modelo, por lo que deben ser eliminados o que la eliminación de algunos de los regresores actuales degrada de forma importante el desempeño del modelo, por lo que deben ser incluidos. Si el modelo analizado es no paramétrico, las variaciones alternativas pueden ser obtenidas aumentando o disminuyendo su complejidad; en ambos casos, debe demostrarse que el nuevo modelo obtenido tiene un desempeño inferior respecto al original. El desempeño puede ser determinado usando medidas como los criterios de información de Akaike, Bayes o Schwartz.

Verificación de la estabilidad de los parámetros La estabilidad de los parámetros en el tiempo se puede establecer a través de su estimación recursiva. Particularmente para el caso de modelos no lineales, tales como las redes neuronales

artificiales, existen muchos puntos de mínima de la superficie de error, por lo que la estimación recursiva de los parámetros debe realizarse a partir sus valores óptimos calculados sobre la muestra de calibración. De esta forma, la estimación se realiza para la muestra menos la última observación, luego para la muestra menos las dos últimas y así sucesivamente. La detección de los cambios de los parámetros en el tiempo, se puede realizar indirectamente por el monitoreo de las variaciones en la distribución de los residuales estimados recursivamente, mediante las pruebas CUSUM (véase [55]) y CUSUMQ (véase [9]). Igualmente, ambas pruebas permiten detectar cambios en la estructura de la serie temporal.

Determinación de las Propiedades del Modelo Uno de los objetivos de la determinación de las propiedades estadísticas de los modelos es establecer qué tan bien se reproducen las propiedades homólogas en los datos. Aunque algunos autores recomiendan el estudio de estas propiedades, en los trabajos revisados sobre sistemas de inferencia borrosa en el análisis de series temporales no se encontraron ejemplos de su uso.

- Propiedades estáticas: Se determinan las propiedades estadísticas de \hat{Y}_t , las cuales están resumidas en su distribución de probabilidades. Para ello se estiman la esperanza y la varianza condicionales de \hat{Y}_t , $E \left[\hat{Y}_t \mid \hat{Y}_{t+k} \right]$ y $V \left[\hat{Y}_t \mid \hat{Y}_{t+k} \right]$ respectivamente.
- Propiedades Dinámicas: Tradicionalmente ha sido de interés el estudio de las propiedades dinámicas, (que son discutidas en el contexto general de los modelos no lineales por [66]), ya que ellas dan una idea del comportamiento local del modelo ante distintos regímenes ([27]). A partir de la formulación general definida en 1.3, las propiedades de largo plazo se pueden obtener estimando el valor esperado de \hat{Y}_t en el largo plazo. Para ello, se realiza la solución numérica de la siguiente ecuación, en donde todas las ocurrencias de \hat{Y}_{t-j} son reemplazadas por su valor esperado $E \left[\hat{Y}_t \right] = y_*$:

$$E \left[\hat{Y}_t \right] = \hat{f}(x_t)$$

Para ello, se asumen diferentes valores iniciales de y_* muestreados en el dominio de \hat{Y}_t , y se realiza un proceso de recursión en donde el término aleatorio ha sido fijado en cero; dicho proceso puede converger a un único punto de equilibrio estable independientemente del valor inicial; puede presentar un ciclo límite, en donde una secuencia de valores se repite de forma infinita; puede presentar un comportamiento caótico, en el que pequeñas variaciones en las condiciones iniciales pueden causar comportamientos considerablemente diferentes, aunque la realización como tal no diverge a un valor infinito. Finalmente, se puede presentar la divergencia del proceso a un valor infinito, siendo esta última situación una causal suficiente

para descartar el modelo obtenido. Igualmente pueden presentarse combinaciones de los anteriores, donde, por ejemplo, ciertos valores iniciales generan la convergencia a un punto de equilibrio estable, mientras otros causan un ciclo límite.

Estabilidad Local La estabilidad local del modelo está determinada por el efecto que produce en el tiempo una perturbación aleatoria ocurrida en el instante t , y para su análisis se utiliza la Función Generalizada de Respuesta al Impulso (véase [43]), la cual mide el cambio que se produce en el pronóstico determinístico para el periodo $t + k$, cuando se aplica una perturbación u_t conocida en el instante t al pronóstico de \hat{Y}_t con la información conocida en $t - 1$, X_{t-1} ; dicha función es definida como:

$$GIR(k, x_{t-1}, u_t) = E \left[\hat{Y}_{t+k} \mid X_t = x_t, u_t \right] - E \left[\hat{Y}_{t+k} \mid X_t = x_t, u_t = 0 \right]$$

Se considera que el modelo obtenido es inestable cuando la perturbación aleatoria se amplifica en el tiempo. En el caso no lineal, [43] han mostrado que la respuesta producida es dependiente del valor actual de la serie temporal, la magnitud de la perturbación (a la que se responde de forma no lineal), y a su signo. Consecuentemente, el resultado de evaluar la función GIR es una cantidad aleatoria, por lo que para su representación se utilizan gráficos de caja. Al igual que sucede con el estudio de las propiedades dinámicas de los modelos, no se ha encontrado que el análisis de la estabilidad local se realice en los trabajos revisados.

Precisión de la Predicción Los criterios de selección del modelo final basados en la precisión de la predicción, implican la separación de la información disponible en dos muestras de datos, de las cuales la primera se usa para la estimación de los parámetros y la determinación del ajuste de la predicción a los datos históricos (*in-sample*), mientras que la segunda muestra se usa para establecer la capacidad de predicción del modelo por fuera de la muestra de calibración (*out-of-sample*) y, particularmente para los modelos no lineales, diagnosticar si se ha presentado la memorización de la información. En la literatura de series temporales, la precisión de la predicción *in-sample* ha sido usualmente determinada a través de las siguientes medidas de error:

- El porcentaje de la varianza explicada o coeficiente de determinación:

$$R^2 = 1 - \sum_{t=1}^T \frac{\hat{e}_t^2}{y_t^2}, \text{ donde } \hat{e}_t = y_t - \hat{y}_t$$

el cual es ajustado usualmente por la cantidad de parámetros, N , para que pueda ser usado como un criterio de comparación entre modelos:

$$adj - R^2 = R^2 - \frac{N-1}{T-N} (1 - R^2)$$

– La sumatoria de errores cuadráticos (SSE):

$$SSE = \sum_{t=1}^T \hat{e}_t^2$$

– La sumatoria de errores absolutos (SAE):

$$SAE = \sum_{t=1}^T |\hat{e}_t|$$

– El error absoluto porcentual medio (MAPE):

$$MAPE = \sum_{t=1}^T \left| \frac{\hat{y}_t}{y_t} - 1 \right|$$

En las definiciones anteriores, T representa el tamaño de la muestra para realizar la estimación de los parámetros. Por otra parte, el estadístico U propuesto por [65] es una medida relativa de la precisión de la predicción respecto a un pronóstico ingenuo que usa como pronóstico el último valor conocido:

$$U^2 = \frac{\sum_{t=2}^T \hat{e}_t^2}{\sum_{t=2}^T (y_t - y_{t-1})^2}$$

La precisión de la predicción fuera de la muestra de estimación (*out-of-sample*) es estimada usualmente (véase por ejemplo a [29], [27], entre otros) ya sea usando la suma de errores cuadráticos extrapolativos:

$$ESS = \sum_{t=T+1}^{T+k} \hat{e}_t^2$$

O la suma de errores absolutos extrapolativos:

$$ESAV = \sum_{t=T+1}^{T+k} |\hat{e}_t|$$

cuyos promedios sobre el horizonte de predicción de k periodos corresponden al error cuadrático medio predictivo (MSPE) y el error absoluto porcentual medio (MAPE). El MSPE se define como

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

donde n es el tamaño de la muestra de datos de validación.

1.4 Sistemas de inferencia borrosa

1.4.1 Introducción

Herramientas como las redes neuronales y los sistemas de inferencia borrosa se han convertido en opciones importantes para el modelado de series temporales no lineales, dadas las ventajas ya mencionadas con respecto a los modelos lineales y los modelos no lineales de tipo paramétrico. Esta tendencia es comprensible dada la fuerte conexión que existe entre el área de regresión y el área de series temporales (véase por ejemplo [50]). Es razonable que se examine las herramientas cuyo poder de aproximación con base en datos sea importante y se quiera determinar su utilidad en el campo de las series temporales. Más aún, autores como [27], manifiestan que la experimentación con modelos no lineales, en campos como la economía, es poca y que el aprendizaje no ha sido el suficiente como para determinar cuáles modelos son mejores que otros. De manera que la exploración de modelos es una actividad vigente. Y en el área de los sistemas de inferencia borrosa, por ejemplo, esta situación se manifiesta, parcialmente, en la ramificación de enfoques en cuanto a su uso y construcción. A continuación se dará una revisión de estos sistemas y de sus aplicaciones en el área de las series temporales.

1.4.2 Conceptos básicos

Un conjunto borroso, que es la base para la construcción de las reglas en los sistemas de inferencia, se define de la siguiente manera:

Si X es una colección de objetos denotados por x , entonces un conjunto borroso A en X se define como un conjunto de pares ordenados

$$A = \{(x, \mu_A(x)) \mid x \in X\} \tag{1.4}$$

Donde $\mu_A(x)$ se denomina función de pertenencia para el conjunto borroso A . Dicha función de pertenencia mapea cada elemento de X a un grado de pertenencia entre 0 y 1.

Como ya se ha mencionado, los sistemas de inferencia borrosa pueden capturar conocimiento expresado de forma vaga. Los conjuntos borrosos ayudan a describir estos conceptos vagos en forma de etiquetas lingüísticas, como se observa en la Figura 1.1. Allí se establecieron los conceptos de *alto* y *bajo* para las variables temperatura (T).

y presión (P). El concepto de alto o bajo para cada variable se describe mediante las funciones correspondientes de T y P . La regla en 1.5 expresa el comportamiento local del sistema de inferencia en una región del espacio $T \times P$ donde tanto la temperatura como la presión se consideran bajas. Para cada región del espacio de entrada se pueden definir reglas y obtener la salida final del sistema para un vector (T, P) mediante una agregación de las salidas para las diferentes reglas.

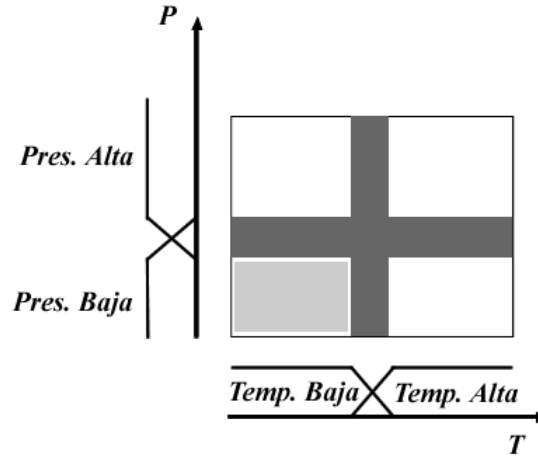


Figura 1.1: Partición para un sistema de inferencia borroso.

$$\text{Si } T \text{ es Baja y } P \text{ es Baja entonces } z_1 = f_1(T, P) \quad (1.5)$$

$$z(\mathbf{x}) = \sum_{i=1}^L \bar{w}_i f_i(\mathbf{x}) \quad (1.6)$$

En términos más detallados, el sistema de inferencia borroso se compone de una base de reglas borrosas; una base de datos, que contiene las definiciones de las funciones de pertenencia usadas en las reglas y un mecanismo de razonamiento, que realiza el procedimiento de inferencia sobre las reglas y hechos para obtener una salida o conclusión. La entrada y la salida del sistema pueden estar en forma de conjuntos borrosos o de valores concretos. Sin embargo, algunas veces (por ejemplo al usar el sistema como un controlador) se requiere una salida concreta, como se observa en 1.6, donde esta se presenta en forma numérica y no en forma de conjunto borroso.

Cuando se utiliza un sistema de inferencia borroso para la extracción de conocimiento a partir de un conjunto de datos, puede ser interesante observar las funciones de pertenencia que se obtuvieron y su significado en términos del problema tratado. En otras situaciones, como cuando se utiliza el sistema en forma de caja negra y sólo

hay interés en la calidad del mapeo que este realiza entre el espacio de entrada y el espacio de salida, puede darse menos importancia al significado de los conjuntos y reglas obtenidos. Tal situación sucede en aplicaciones de regresión, como se ilustra en la Figura 1.2, donde los conjuntos borrosos pueden no estar asociados a un concepto del dominio del problema.

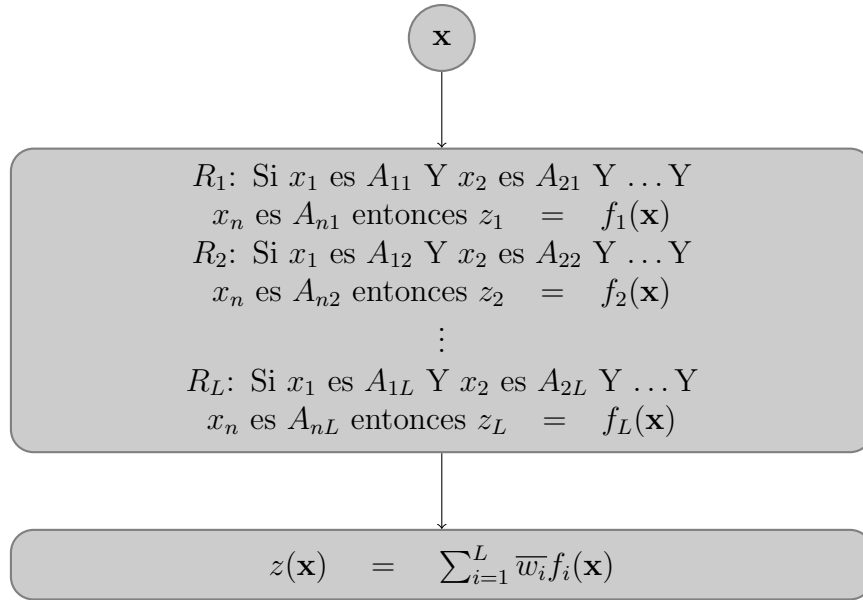


Figura 1.2: Sistema de inferencia borrosa tipo Takagi-Sugeno con varias reglas.

Además de los sistemas de tipo Takagi-Sugeno (TSK), ilustrados anteriormente, existen los sistemas de tipo Mamdani, que se caracterizan porque su salida es borrosa y, por tanto, se requiere usar métodos de concreción para hallar salidas cuando se desee valores numéricos concretos. También se tiene los sistemas de inferencia borrosa tipo Tsukamoto, que son similares a los TSK con algunas modificaciones en la forma de calcular los consecuentes de las reglas de inferencia. Se utilizan, por ejemplo, funciones monótonas para transformar la salida de cada regla. Para un tratamiento detallado de temas como los conjuntos borrosos, las reglas borrosas y los métodos de inferencia borrosa puede consultarse [35].

1.4.3 Tendencias en el modelado con sistemas de inferencia borrosa

La versatilidad de los sistemas de inferencia borrosa ha generado una gama de aplicaciones amplia, dentro de las cuales el modelado y pronóstico es sólo uno de los tópicos tratados. Son abundantes las opciones en cuanto a los modelos y algoritmos de en-

trenamiento estudiados. En esta sección se realiza una revisión de trabajos relevantes en el modelamiento o predicción de series temporales.

Entre los trabajos recientes puede mencionarse el de [38], que se constituye en una referencia relevante ya que formula explícitamente estructuras de sistemas de inferencia borrosa tipo feedforward y recursivos que son útiles en problemas de series temporales y que son usados en trabajos de otros autores. El modelo feedforward es:

$$\text{Regla } i : \text{ Si } x_1(t) \text{ es } A_{i1} \text{ Y } \dots \text{ Y } x_n(t) \text{ es } A_{in} \text{ entonces } u(t+1) = a_{i0} + \sum_{j=1}^n a_{ij}x_j(t) \quad (1.7)$$

Donde $x_j(t)$ es una variable de entrada, u es la variable de salida y A_{ij} es un conjunto borroso. La salida del modelo es una combinación de la salida de cada regla, utilizando el método de defuzificación de promedio ponderado:

$$u = \frac{\sum_{i=1}^r \mu_i(\mathbf{x}) f_i(\mathbf{x})}{\sum_{i=1}^r \mu_i(\mathbf{x})} \quad (1.8)$$

El modelo recursivo es

$$\begin{aligned} \text{Regla } i : \text{ Si } x_1(t) \text{ es } A_{i1} \text{ Y } x_2(t) \text{ es } A_{i2} \text{ Y } \dots \text{ Y } x_n(t) \text{ es } A_{in} \text{ Y } h_i(t) \text{ es } G \\ \text{ entonces } y(t+1) \text{ es } a_{i0} + \sum_{j=1}^n a_{ij}x_j(t) + a_{in+1}h_i \end{aligned} \quad (1.9)$$

$$\text{ Y } h_1(t+1) \text{ es } w_{i1} \text{ Y } h_2(t+1) \text{ es } w_{i2} \text{ Y } \dots \text{ Y } h_r(t+1) \text{ es } w_{ir}$$

Donde A_{ij} y G son conjuntos borrosos, w_{ij} , a_{ij} son los parámetros del consecuente para la salida h_i e y , respectivamente; $h_i(t)$ se utiliza para memorizar historia temporal.

La salida de este sistema se describe en las siguientes ecuaciones:

$$y(t+1) = \frac{\sum_{i=1}^r \phi_i(\mathbf{x}(t), h_i(t)) f_i(t)}{\sum_{i=1}^r \phi_i(\mathbf{x}(t), h_i(t))} \quad (1.10)$$

$$h_i(t+1) = \sum_{k=1}^r \phi_k(x(t), h_i(t)) w_{ik} \quad (1.11)$$

$$f_i(t) = a_{i0} + \sum_{j=1}^n a_{ij}x_j(t) + a_{in+1}h_i(t) \quad (1.12)$$

$$\phi_i(x(t), h_i(t)) = \frac{1}{1 + e^{-h_i(t)}} \mu_i(\mathbf{x}(t)) \quad (1.13)$$

En dicho trabajo se utilizan técnicas de agrupamiento borroso para determinar la estructura del sistema con base en 1.7 ó 1.9 y posteriormente se optimiza los parámetros. En la especificación de la estructura se utilizan conjuntos borrosos unidimensionales², lo cual da la posibilidad de realizar interpretación de los mismos una vez obtenidos. Trabajos como estos han avanzado en la dirección de modificar la estructura de inferencia de tal manera que se adecue más a su uso en el modelamiento de series temporales. Un enfoque similar puede encontrarse en [22].

En trabajos como [31] se utiliza particionamiento a través de rejillas, en lugar de agrupamiento, para facilitar la determinación de la estructura del sistema de inferencia. El modelo de inferencia borroso obtenido finalmente posee también conjuntos unidimensionales dado que se persigue conservar la interpretabilidad de los mismos.

En [53] se formula un modelo muy similar al descrito en 1.7, generalizando aspectos como la forma en que se determina el grado de concordancia o de disparo de las reglas para un vector de salida. En el modelo se considera cada entrada en forma de vector, y se establece el grado de disparo de las reglas para dicho vector, no para las componentes individuales del mismo, lo cual se persigue en otros esquemas que usan agrupamiento. Para el entrenamiento del modelo, se lo representa en forma de red y se aplica un algoritmo de entrenamiento híbrido.

Los modelos que utilizan algoritmos genéticos tienen estructuras similares a los modelos mencionados anteriormente. Sus diferencias principales se encuentran en el tipo de algoritmos usados para la determinación de los parámetros. Como ejemplos se tiene los trabajos de [4], [11], [33], [41], [63] y [59]. En ellos se explora la utilización de algoritmos evolutivos para la especificación de sistemas de inferencia borrosa. En algunos se asume un número dado de reglas y se procede a encontrar la estructura desde dicho punto de partida. En otros trabajos se determina de manera automática.

Por otro lado, en trabajos como [18], se utilizan sistemas de inferencia borrosa tipo Takagi-Sugeno para realizar combinaciones de pronósticos. En este caso, el uso de estos sistemas con conjuntos unidimensionales brinda la capacidad de tornar más claros los resultados para efectos de interpretación y análisis por parte de expertos. No se estudia en este trabajo el uso directo de los sistemas de inferencia en la generación de pronósticos o el modelado de la dinámica de una serie temporal. En [45] se tiene un ejemplo de modelos combinados en el cual el sistema de inferencia borrosa se usa para capturar conocimiento experto. La captura de la dinámica de la serie, a través de datos numéricos, se deja a una red neuronal. Otros modelos combinados pueden encontrarse en [54], [71] y [73], [15], [56] y [42]. Se combina sistemas de inferencia borrosa y redes neuronales (redes neurodifusas).

En las orientaciones sobre el uso de conjuntos borrosos en los sistemas de inferencia se tienen la adopción de conjuntos unidimensionales y de conjuntos multidimensionales. Estos últimos pueden expresarse, para problemas con varias variables de entrada, de una manera más compacta. Autores como [44], [25], [26], [2] han realizado avances

²Por conjunto borroso unidimensional se entiende aquel cuya función de pertenencia es del tipo $\mu : \mathbf{R} \rightarrow [0, 1]$

en este tipo de sistemas de inferencia. Entre sus ventajas se encuentra la posibilidad de construirlos sin que el número de reglas (y el número de parámetros de los consecuentes) crezca exponencialmente a medida que se aumenta el número de variables de entrada, como sucede con los sistemas de inferencia cuya estructura se determina con base en conjuntos borrosos unidimensionales y con esquemas de particionamiento de tipo rejilla. Sobre estos sistemas de inferencia se dará más detalle en la siguiente sección.

Los conjuntos borrosos que se han utilizado en los trabajos mencionados hasta ahora son de tipo I, que se definen como aquellos para los cuales el grado de pertenencia de un elemento al conjunto se califica con número en el intervalo $[0, 1]$. Otros autores (véase por ejemplo [39]) han experimentado con el uso de conjuntos borrosos de tipo II, que son aquellos para los cuales el grado de pertenencia se especifica con una función borrosa, es decir, aquella donde el valor de pertenencia es un valor borroso, en lugar de un valor concreto, en $[0, 1]$. Este tipo de orientaciones parecen favorecer el manejo de datos con ruido y la obtención de resultados similares a los intervalos de confianza, a la hora de realizar pronósticos.

La revisión anterior evidencia la variedad de orientaciones en la utilización de sistemas de inferencia borrosa (SIB) para el modelamiento y pronóstico de series temporales. Dicha variedad podría resumirse con ayuda de una clasificación breve, como se ilustra en la Figura 1.3. Allí se muestra que tanto para su especificación estructural (conjuntos borrosos y reglas) como para su especificación algorítmica (que determina la manera en que se construye el sistema) existen diversas variantes. Estas opciones dan origen, al combinarlas, a nuevas opciones que son exploradas para distintos propósitos, tal como sucede en el campo de los modelos paramétricos. Algunos modelos de SIB se han usado de manera directa en el modelamiento y predicción de series temporales, mientras que otros se han utilizado como complemento en investigaciones con otros modelos, sin que estén involucrados en el modelamiento o predicción. Por otro lado, los problemas en que se han aplicado estos sistemas son de tipo académico, para estudiar propiedades de los modelos, o de tipo práctico, para establecer bondades de los modelos en la solución de un problema. De esta manera, el enfoque de los autores tiende a ser más particular que general y pareciera que los sistemas de inferencia borrosa muestran su utilidad en el área de las series temporales cuando se alejan de los modelos simples (tipo TSK, por ejemplo) y se complementan con otros avances, como las redes neuronales y los algoritmos evolutivos.

1.5 Sistemas de inferencia borrosa multidimensionales

El enfoque de los sistemas de inferencia borrosos con conjuntos multidimensionales es el de interés en esta investigación por varias razones. Como consideración general, los sistemas de inferencia borrosa poseen la capacidad de ser aproximadores universales

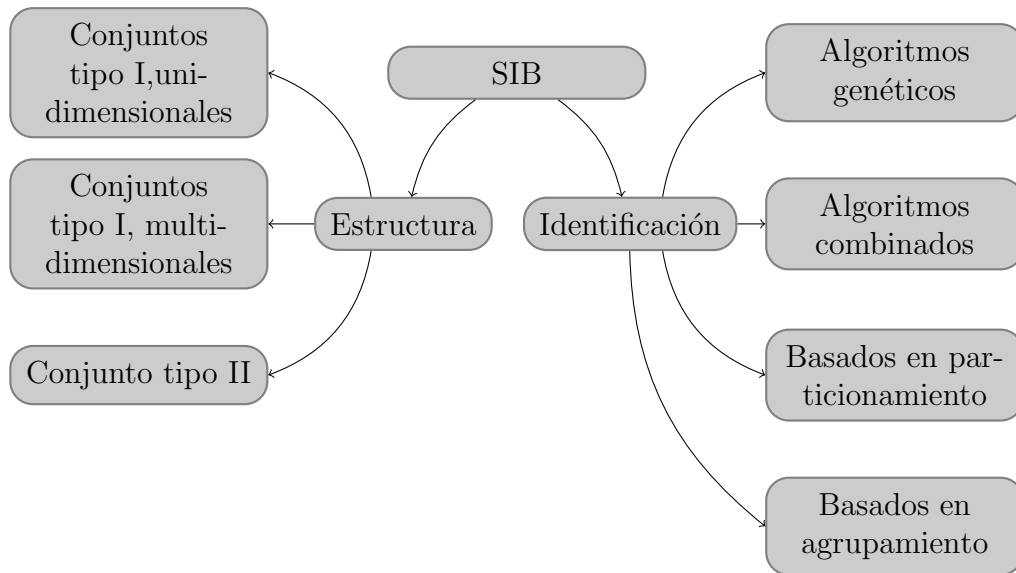


Figura 1.3: Una clasificación general de las tendencias en modelos de inferencia borrosa para el análisis de series temporales.

de funciones (véase [70]), lo cual resulta atractivo en problemáticas donde se tenga muy factiblemente dinámicas no lineales. Por otro lado, existe la dificultad de la especificación de las reglas en los sistemas de inferencia borrosa cuando la cantidad de variables aumenta. Por ejemplo, para un problema con dos particiones en cada variable y 6 variables de entrada se tiene $2^6 = 64$ reglas con $(6 + 1) \times 64 = 448$ parámetros lineales. Los SIB multidimensionales simplifican la estructura del sistema de manera que se puede controlar la cantidad de reglas y se reduce el número de parámetros de los consecuentes. Otra consideración particular es que existen aplicaciones promisorias de los SIB multidimensionales en regresión y modelamiento de sistemas no lineales (véase por ejemplo [28], [37], [56]). Además, se requiere ganar más experiencia en la construcción y aplicación de los SIB multidimensionales. En autores como [44], [25], [26] se mencionan tópicos de investigación abiertos como:

- La robustez de los algoritmos de identificación y de los modelos resultantes son aspectos bajo investigación.
- El afinamiento de los sistemas de inferencia es un tópico bajo estudio.
- El establecimiento de cuál es el mejor mecanismo de agrupamiento de datos (que guía la obtención de las reglas borrosas) está en investigación.

Finalmente, la exploración de modelos para el análisis de series temporales es plausible, como ya es expuso, y los sistemas de inferencia multidimensionales han sido poco explorados dicha área. Por estas razones, se estudió tales sistemas y se

examinaron algunos tópicos en cuanto a su uso en el modelado y pronóstico de series temporales. A continuación se profundiza un poco en estos sistemas de inferencia.

A través de los conjuntos borrosos multidimensionales se establece el grado de pertenencia para un elemento X que está definido en un espacio de dimensión $n > 1$ (véase la Figura 1.4). A su vez pueden construirse sistemas de inferencia en los cuales las reglas borrosas estén expresadas en términos de estos conjuntos, como se muestra en 1.14.

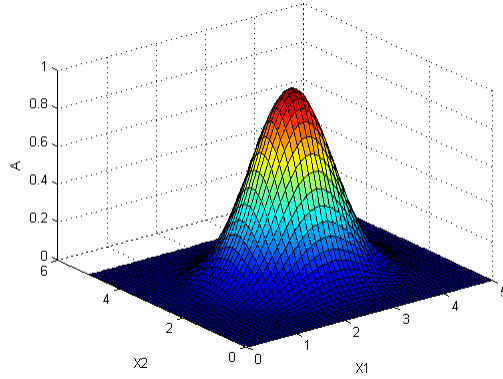
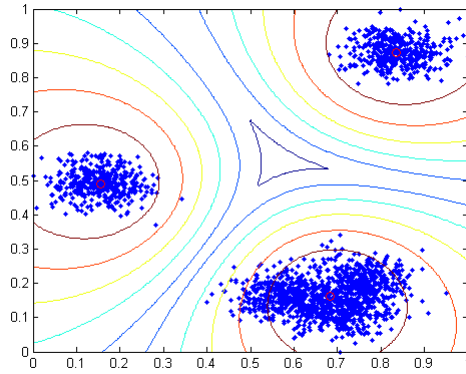


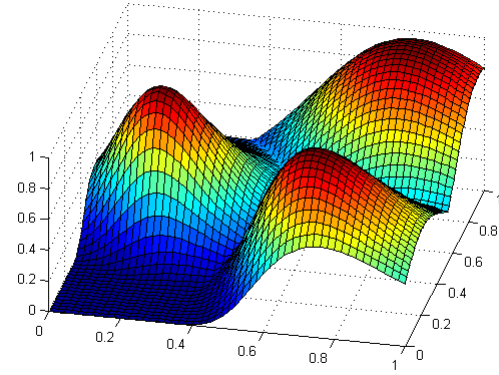
Figura 1.4: Función de pertenencia para un conjunto borroso multidimensional.

$$\begin{aligned}
 &\text{Si } (\mathbf{X} \text{ es } A) \text{ entonces } z_1 = f_1(\mathbf{X}) \\
 &\mathbf{X} = (X_1, X_2) \\
 &A : \mathbf{R}^2 \rightarrow [0, 1]
 \end{aligned}
 \tag{1.14}$$

El uso de este tipo de conjuntos borrosos, para simplificar la estructura de las reglas en el sistema de inferencia, ha sido tratado por varios autores, que han realizado importantes aportes (véase [44], [25], [26], [2] y [28]). Estos trabajos se caracterizan por el uso de los sistemas en los que se extrae conocimiento a partir de datos numéricos, sin requerir que los conjuntos borrosos obtenidos sean interpretables para el experto en el problema tratado (uso en forma de caja negra). La tendencia en cuanto a la determinación de la estructura del sistema es utilizar agrupamiento borroso. A través de algoritmos de agrupamiento se pueden construir prototipos iniciales de las reglas borrosas a partir de los datos. Esto se ilustra en la Figura 1.5(a), donde los datos dan origen a tres grupos borrosos, con base en los cuales se construyen las funciones de pertenencia de los conjuntos multidimensionales a usar en las reglas del sistema. Las líneas de nivel de colores cercanos al rojo indican mayor pertenencia de los datos al grupo. Las funciones de pertenencia derivadas de estos grupos se ilustran



(a) Grupos en 2D.



(b) Funciones de pertenencia.

Figura 1.5: El agrupamiento como mecanismo de identificación.

a través de las superficies indicadas en la Figura 1.5(b). Otros autores como [37] utilizan agrupamiento para facilitar el particionamiento del espacio de entrada y utilizan conjuntos borrosos multidimensionales, pero siguen el enfoque de [62] en el cual se trata de encontrar conjuntos interpretables en una dimensión, a partir de conjuntos borrosos multidimensionales.

Los pasos generales de las metodologías usadas en este enfoque, se mencionan a continuación, teniendo en cuenta que hay variaciones en el tratamiento particular de cada autor.

- Preprocesamiento de datos: Algunos autores realizan un filtrado de los datos para disminuir la influencia de ruido y de datos espurios. Este paso también puede incluir la normalización de los datos entrada-salida y la eliminación de los datos repetidos.
- Agrupamiento: Se utiliza un algoritmo de agrupamiento (p.ej. el Fuzzy C-Means) para generar k grupos a partir de los datos entrada-salida del sistema. Puede realizarse agrupamiento en el espacio producto de las variables de entrada y salida o separadamente en cada espacio.
- Se utiliza la información de los k grupos y sus centroides para construir un conjunto de reglas iniciales de las cuales se puede obtener luego el sistema afinado. En cada regla borrosa se hace referencia a un conjunto borroso en varias dimensiones y no a varios conjuntos de una sola variable. Esto ayuda a disminuir la complejidad en la determinación de la estructura del sistema de inferencia.
- Se utiliza un mecanismo de inferencia. Se ha probado con el método de inferencia de Mizumoto, que utiliza consecuentes constantes (grupos borrosos tipo

singleton). También se ha probado con consecuentes funcionales como en los sistemas Takagi-Sugeno. Estos últimos tienen mejores resultados al permitir modelar sistemas con grados altos de complejidad.

Además de las particularidades de la metodología en cada autor, estos sistemas de inferencia tienen varios aspectos que están sujetos a variación. Uno de ellos es el algoritmo de agrupamiento escogido, el cual a su vez tiene elementos importantes que influyen el resultado final del sistema de inferencia, como son: los puntos de arranque del algoritmo, la función de distancia (o grado de similitud) utilizada, la función de costo usada y el número inicial de grupos. Otro aspecto es el espacio usado para realizar el agrupamiento: puede ser el espacio producto de las variables de entrada y salida, el espacio de entrada solamente o ambos espacios por separado. Se tiene también el algoritmo de optimización usado y la forma de implementarlo. En general se realiza una optimización para encontrar los grupos borrosos y sus centros, luego de lo cual se determina los coeficientes en los consecuentes de las reglas y, al final, se realiza una optimización global tanto de los centros de los grupos como de los coeficientes en los consecuentes.

De esta diversidad de opciones surgen varios interrogantes. ¿Cuál es la mejor técnica de agrupamiento para estos sistemas? Dado que es posible que una sola técnica no sea satisfactoria para la diversidad de problemas que pueden encontrarse, ¿hay alguna metodología que pueda guiar la selección de dicha técnica? ¿Realmente es necesario realizar agrupamiento para conseguir los resultados que se han obtenido a través de los sistemas de inferencia en los que se utilizan las técnicas? Por otro lado, se observa que es común el uso de algoritmos de agrupamiento como el Fuzzy C-Means o Gustafson Kessel, que utilizan puntos de partida aleatorios o dictados por el modelador. Esto puede dar lugar a múltiples configuraciones de grupos y, dado que dichos grupos son el insumo para la construcción de las reglas del sistema de inferencia, se tendrá a su vez varias configuraciones de modelos para distintos puntos de partida. Además, la utilización del agrupamiento se ha visto como un paso de las metodologías, cuyo resultado se afina una vez se han construido las reglas del sistema. No se ha explorado la construcción de los grupos de manera que la inclusión de uno u otro este dictada por su aporte en la descripción de la dinámica de la serie. Una situación similar se tiene con la determinación del número óptimo de grupos (que a su vez corresponde generalmente con el número de reglas del sistema de inferencia): los criterios para determinar el número apropiado de grupos utilizan procedimientos de validación de grupos o procedimientos de particionamiento sobre los datos. No se ha observado que el número adecuado de grupos se asocie a la capacidad del sistema para describir adecuadamente la dinámica de la serie.

Finalmente, en la revisión realizada de estos sistemas de inferencia borrosa (como en el caso de los de estructura unidimensional) se han encontrado indicios de algunas prácticas metodológicas para su uso en el modelamiento y pronóstico de series temporales:

- El uso de diagramas de errores para visualizar la calidad del ajuste del modelo a la serie.
- La utilización de histogramas de errores.
- La utilización de diagramas de dispersión para los rezagos, con el fin de visualizar los posibles grupos de datos.
- La utilización (aunque en muy pocos autores) de autocorrelogramas para guiar la selección de variables de entrada.
- La realización de discusiones de resultados y selección de modelos con base en medidas de desempeño para conjuntos de datos de entrenamiento y validación.

Sin embargo, se encuentra que algunos aspectos metodológicos de suma importancia no tienen un tratamiento formal. Entre ellos cabe mencionar los siguientes:

- No se mencionan pruebas formales de linealidad de la serie, para guiar al modelador en la toma de decisiones en cuanto al uso de modelos lineales o no lineales.
- No es generalizado el uso de herramientas para la selección de variables.
- No se especifican transformaciones de datos más allá de las requeridas por el modelo para su funcionamiento.
- No se aplican criterios de selección de modelos.
- No se examina formalmente los residuales del modelo, de manera que pueda evaluarse la capacidad del mismo para explicar adecuadamente de la dinámica de la serie.

Estos hallazgos muestran la carencia de una metodología explícita fundamentada en herramientas de la estadística para la construcción de los modelos (para predicción o modelamiento de series temporales). Por otro lado, como ya se mencionó, la experiencia en el área de series temporales es poca (algunos trabajos pueden encontrarse en [35], [39], [13], [14]) y es aún más restringida la cantidad de trabajos con series temporales de tipo económico. También se observa que en las aplicaciones de los modelos se enfatizada el buen ajuste a los datos, pero no es común el uso de los criterios mencionados en la sección 1.3.4, que ayudan a evaluar un modelo de predicción.

1.6 Requerimientos, objetivos e hipótesis de investigación

1.6.1 Requerimientos y limitaciones

Como se estableció en la sección 1.5, existen varios aspectos por resolver, concernientes al uso de los sistemas de inferencia en cuestión para el modelado y predicción de series temporales no lineales. A continuación se enumeran aquellos en los que se concentra la atención en esta investigación:

1. La utilización del agrupamiento se ha visto como un paso de las metodologías, cuyo resultado se afina una vez se han construido las reglas del sistema. No se ha explorado la construcción de los grupos de manera que la inclusión de uno u otro este dictada por su aporte en la descripción de la dinámica de la serie. Más aún, se desconoce cuál metodología de agrupamiento sería más adecuada para las series temporales.
2. En los algoritmos de agrupamiento que se usan comúnmente hay dependencia de los puntos iniciales escogidos. Esto ocasiona que haya diversas estructuras posibles del modelo para diversos puntos de partida.
3. Los criterios para determinar el número apropiado de grupos (y de reglas asociadas) utilizan procedimientos de validación de grupos o procedimientos de particionamiento sobre los datos. No se ha observado que el número adecuado de grupos se asocie a la capacidad del sistema para describir adecuadamente la dinámica de la serie.
4. Las metodologías no están vinculadas al conocimiento propio de la Econometría y el modelamiento estadístico de series temporales.
5. El uso de dichos sistemas para el modelado y la predicción de series temporales no lineales es muy escaso y existe poca experiencia con series temporales de tipo económico.

1.6.2 Hipótesis de investigación

La tesis defendida en esta investigación es que los sistemas de inferencia borrosa del tipo Takagi-Sugeno, que usan conjuntos borrosos multidimensionales, son aplicables al modelado y predicción de series temporales no lineales, y esta aplicación puede potenciarse al enmarcar el uso de los modelos dentro de una metodología formal desde el punto de vista estadístico.

A continuación se enumeran los objetivos específicos de investigación que permitirán resolver las limitantes descritas en la sección anterior.

1.6.3 Objetivos

1.6.3.1 Objetivos específicos

1. Proponer un mecanismo de particionamiento o agrupamiento que incluya o descarte grupos con base en el aporte del modelo en la descripción de la dinámica de la serie.
2. Proponer un mecanismo de particionamiento o agrupamiento que mitigue la dependencia de puntos iniciales.
3. Proponer un criterio para establecer la complejidad del modelo con base en el desempeño global del sistema y que sea independiente de criterios de validez del algoritmo de agrupamiento.
4. Formular una metodología para el modelamiento y predicción de series temporales no lineales, que haga uso de los sistemas de inferencia borrosa en cuestión. Dicha metodología deberá estar basada en el modelamiento estadístico de series temporales.
5. Utilizar el modelo y la metodología obtenidos para analizar series *benchmark*, que permitan comparar el desempeño del modelo.
6. Utilizar el modelo y la metodología obtenidos para analizar una o más series reales, de tipo económico.

1.6.3.2 Objetivo general

Proponer una aproximación metodológica para la especificación de modelos de inferencia borrosa del tipo Takagi-Sugeno, que usen conjuntos difusos multidimensionales, para el modelado y la predicción de series temporales no lineales, y aplicarla al modelado de una o más series temporales reales.

1.6.4 Aportes que se espera obtener

1. Obtención de un modelo en el que las reglas de inferencia multidimensionales, construidas a partir de grupos borrosos, realicen un aporte a la descripción de la dinámica de la serie que se está modelando. Además el modelo deberá mitigar en alguna medida la dependencia de los puntos iniciales del algoritmo de agrupamiento.
2. Obtención de un aporte de tipo metodológico para el modelado y predicción de series temporales no lineales con sistemas de inferencia borrosa basados en CBM (conjuntos borrosos multidimensionales). El sustento estadístico que se utilizará para formular la metodología permitirá un uso más acertado de los

modelos. Dentro de la metodología se incluirá un criterio que permita establecer el número adecuado de grupos (y, por tanto, de reglas borrosas) de tal manera que no sea dependiente del algoritmo de agrupamiento.

3. Se construirá una aplicación que permita validar la propuesta metodológica utilizando un problema de aplicación real, del área de la economía.

Capítulo 2

Metodología para el modelado de series temporales con sistemas de inferencia borrosa multidimensionales

2.1 Introducción

En el modelado de series temporales se han utilizado diversas herramientas, mencionadas en el capítulo introductorio. Los sistemas de inferencia borrosa multidimensionales, una de dichas herramientas, han sido usados con resultados alentadores. Sin embargo, el modelado de series temporales con estos sistemas es un campo poco explorado y existen aspectos en cuanto a los modelos y a las metodologías para su uso que merecen un tratamiento más detallado. Como respuesta a algunos de estos aspectos y a los interrogantes planteados en el capítulo anterior, se proponen en este capítulo un modelo de sistemas de inferencia multidimensionales y una metodología que tienden a solucionar tales interrogantes.

El avance realizado se enmarca dentro varias áreas del conocimiento. Dada la naturaleza de los modelos usados, el área que proporciona los cimientos principales para la investigación es la inteligencia computacional, dando los puntos de partida en cuanto a la estructura de los sistemas de inferencia multidimensionales y en cuanto a la inferencia borrosa. Por otro lado, dentro de la búsqueda de la solución a los interrogantes planteados como fundamento de la investigación, se encontró que los modelos de regresión tales como MARS poseían características sumamente deseables en tópicos como el particionamiento de un espacio de entrada para un modelo de regresión. De manera que el campo de la regresión se constituye en una fuente importante de conocimiento para el avance planteado en el presente trabajo. En cuanto al modelamiento y pronóstico de series temporales, son estos dos temas los que guiaron la aplicación de los modelos propuestos y los que a su vez mostraron la

necesidad de explorar el uso de herramientas estadísticas para formular indicaciones metodológicas de uso de los mismos. De manera que el campo de la estadística es otro punto de apoyo esencial.

En las secciones que siguen se muestran las partes estructural y algorítmica de los modelos propuestos y se exponen los aspectos metodológicos para el uso de los mismos. En la parte final se examinan algunas propiedades de los modelos.

2.2 Estructura del modelo

2.2.1 Sistema de inferencia

La estructura del sistema de inferencia borrosa construido involucra varias decisiones de diseño. Esta sección muestra la estructura del modelo, explica dichas decisiones y muestra los puntos en los cuales hay diversidad de opciones probadas por otros autores. Entre los aspectos relevantes de diseño se tiene el uso de conjuntos borrosos multidimensionales, en lugar de los conjuntos unidimensionales. Esta estrategia, a su vez, crea la posibilidad de tener distintos esquemas de mapeos desde el espacio de las variables de entrada al espacio de la variable de salida (véase [26]). Inicialmente se mostrará la estructura de inferencia y luego la estructura de agrupamiento por medio de la cual se construye los antecedentes de las reglas.

El modelo tiene como base fundamental la formulación de los sistemas de inferencia borrosa tipo Takagi-Sugeno. Las reglas de un sistema de inferencia tipo Takagi-Sugeno tienen la forma indicada en la ecuación 2.1, donde A_{ij} es un conjunto borroso en la regla i para la variable de entrada j .

$$R_i : \text{ Si } x_1 \text{ es } A_{i1} \text{ Y } x_2 \text{ es } A_{i2} \text{ Y } \dots \text{ Y } x_n \text{ es } A_{in} \text{ entonces } z_i = f_i(x_1, x_2, \dots, x_n) \quad (2.1)$$

La salida de un sistema de este tipo se calcula a través de 2.2 y 2.3, donde w_i es el grado de cumplimiento de la regla i .

$$y = \sum_{i=1}^L \overline{w}_i f_i(x_1, x_2, \dots, x_n) \quad (2.2)$$

$$\overline{w}_i = \frac{w_i}{\sum_{j=1}^L w_j} \quad (2.3)$$

Este sistema tiene conjuntos borrosos para calificar las variables de entrada separadamente. En un sistema con conjuntos borrosos multidimensionales, por el contrario, se usan conjuntos borrosos para grupos de variables, lo cual simplifica la forma de las reglas y conserva, del modelo unidimensional, la propiedad de modelar comportamientos locales de una función o sistema en distintas regiones del espacio de entrada. Cada conjunto borroso multidimensional puede obtenerse, en el proceso de entrenamiento,

a través de algoritmos de agrupamiento, con el objetivo de facilitar su construcción. En esta línea de trabajo varían el algoritmo de agrupamiento usado, las variables incluidas en el agrupamiento (bien sea del espacio de entrada o del espacio de salida) y la forma en que se construyen las reglas a partir de los grupos encontrados (véase por ejemplo [26], [25], [44]). De los varios enfoques existentes, se adoptó en este modelo la construcción de grupos con base en las variables del espacio de entrada (el algoritmo de agrupamiento se presenta más adelante en la sección sobre los algoritmos usados). En cuanto a la construcción de las reglas, se crea una por cada conjunto borroso multidimensional encontrado y el grado de cumplimiento de las reglas para un dato se determina con base la distancia del dato a los centroides de los grupos. A continuación se detallan estos aspectos del modelo.

La forma general de una regla del sistema de inferencia se expresa en 2.4. Esta forma de construcción implica que existe un conjunto borroso multidimensional para el antecedente de cada regla y la pertenencia a dicho conjunto estará determinada por un grupo encontrado en el espacio de las variables de entrada. Existen otros enfoques en los cuales se tienen varios grupos en el antecedente de una regla (véase [2]).

$$R_i : \text{Si } \mathbf{x} \text{ es } B_i \text{ entonces } z_i = f_i(\mathbf{x}) \quad (2.4)$$

La salida del sistema está dada por 2.5. En este modelo, w_i es el factor de pertenencia normalizado del elemento \mathbf{x} al conjunto borroso multidimensional i y μ_i es el grado de pertenencia del elemento \mathbf{x} al conjunto borroso multidimensional i .

$$y(\mathbf{x}) = \sum_{i=1}^L w_i f_i(\mathbf{x}) \quad (2.5)$$

$$w_i(\mathbf{x}) = \frac{\mu_i(\mathbf{x})}{\sum_{j=1}^L \mu_j(\mathbf{x})} \quad (2.6)$$

$$\mu_i(\mathbf{x}) = e^{-\frac{D_i^2(\mathbf{x})}{\sum_{j=1}^L D_j^2(\mathbf{x})}} \quad (2.7)$$

$$\mathbf{x} = (x_1, x_2, \dots, x_n) \quad (2.8)$$

$$f_i(x_1, x_2, \dots, x_n) = \phi_{i0} + \phi_{i1}x_1 + \phi_{i2}x_2 + \dots + \phi_{in}x_n \quad (2.9)$$

Sea G_i el grupo en el espacio de las variables de entrada que determina el grado de pertenencia, $\mu_i(\mathbf{x})$, del elemento \mathbf{x} al conjunto borroso multidimensional B . Dicho grado de pertenencia, como puede verse en la expresión 2.7, se determina a través de $D_i(\mathbf{x})$, que mide la distancia entre el elemento \mathbf{x} y el centro de G_i . Esta función de pertenencia se construyó con base en la utilizada en [44] y [2] y tiene la propiedad de que su valor máximo ocurre cuando $D_i(x) = 0$, donde el grado de pertenencia es 1, y su valor decae a medida que la distancia entre el centro y el elemento crece. Los consecuentes de las reglas del sistema de inferencia, según la ecuación 2.9, son funciones

lineales de las variables de entrada, lo cual corresponde con los sistemas de inferencia borrosa tipo Takagi-Sugeno de primer orden. Cada parámetro ϕ_{ij} corresponde al coeficiente en la regla i para la variable de entrada j .

Dentro de esta formulación, la variación importante con respecto a enfoques de otros autores es el uso de una función de pertenencia que resultó ser más apropiada por su continuidad. Sin embargo, la estructura general de las reglas borrosas es un punto de partida para los aportes realizados.

2.2.2 Algoritmo

Como se menciona en el capítulo introductorio, los algoritmos de agrupamiento usados en los modelos basados en conjuntos borrosos multidimensionales, como el Fuzzy C-Means, Gustafson Kessel, utilizan, en la construcción de grupos, puntos de partida aleatorios o dictados por el modelador. Esto puede dar lugar a múltiples configuraciones de grupos y, dado que dichos grupos son el insumo para la construcción de las reglas del sistema de inferencia, se tendrá a su vez varias configuraciones de modelos para distintos puntos de partida. Por otro lado, una propiedad esencial del modelo de inferencia es la de contar con reglas que correspondan de manera aproximada con las regiones que merecen tener un tratamiento separado en el espacio de entrada. En esta sección se mostrará el algoritmo construido para la obtención de los modelos, que incluye la obtención de grupos borrosos de una manera que mitiga la dependencia de puntos de arranque aleatorios y que cuenta con la propiedad de ajuste mencionada.

Para que el modelo construya reglas borrosas que tengan una influencia significativa en la explicación de la función modelada se debe tener grupos que representen adecuadamente las regiones importantes del espacio de entrada, ya que las reglas se establecen con base en los grupos. Para alcanzar este objetivo, es conveniente tener un mecanismo que evalúe las regiones (o grupos) y favorezca o penalice su inclusión en el modelo de acuerdo con el aporte que realice su presencia en la descripción de la función a modelar.

Debido a estas necesidades, se ha adaptado el algoritmo de particionamiento recursivo descrito en [21] y parte del algoritmo MARS descrito allí mismo. Estos algoritmos son reconocidos por su capacidad de aproximar formas funcionales complejas y por su propiedad de ajustar simultáneamente la estructura y los coeficientes del modelo. Esta forma de ajuste corresponde con nuestra necesidad de favorecer o penalizar regiones y no parte de puntos aleatorios, lo cual se ajusta a nuestro otro requerimiento. Más aún, al revisar la estructura de un sistema de inferencia borroso multidimensional y compararlo con las técnicas de particionamiento en las que se basa MARS, se encuentran varias analogías y equivalencias que favorecen la asimilación de dichas técnicas en el modelo de inferencia borrosa considerado (en la sección 2.4 se realiza dicha comparación.) Por tanto, en el enfoque adoptado se parte de algoritmos de particionamiento para encontrar los grupos borrosos, lo cual difiere del enfoque usual basado en agrupamiento borroso.

2.2.2.1 Algoritmo general para la obtención de modelos de inferencia

Las partes generales del procedimiento se describen en el algoritmo 2.1. La rutina de particionamiento produce un conjunto de funciones base, B , y un conjunto de coeficientes, Φ , para los consecuentes de las reglas. El concepto de función base pertenece al desarrollo de [21] sobre particionamiento recursivo y es una manera algebraica de representar una partición en un espacio n-dimensional. En la Figura 2.1 se ilustra una partición y su conjunto de funciones base. Cada una de dichas funciones delimita una región del espacio de las variables de entrada. En 2.10 se muestra la forma de una función base. Para un vector de entrada \mathbf{x} , la función indica si dicho vector pertenece a la partición (cuando toma el valor de 1) o si no pertenece (cuando toma el valor de cero). K_m indica el número de particiones que dieron lugar a la función; s_{km} es una constante que toma el valor de 1 o -1 e indica si la partición es a la derecha o a la izquierda del valor t_{km} . Finalmente, \mathbf{x}_v es la componente de \mathbf{x} en la cual se está haciendo la partición (que corresponde a x_{va} , x_{vb} ó x_{vc} en la Figura 2.1). En el algoritmo 2.1, H_{max} es el número máximo de funciones base, que corresponde con el número máximo de reglas admitidas para el sistema de inferencia.

$$B_m(\mathbf{X}) = \prod_{k=1}^{K_m} H(s_{km} \cdot (x_{v(k,m)} - t_{km})) \quad (2.10)$$

$$H(\eta) = \begin{cases} 1 & \text{si } \eta \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Con las funciones base se obtiene un conjunto de centros, C , para los conjuntos multidimensionales y con dichos centros se construyen las reglas del sistema de inferencia borroso. Los centros se obtienen de manera similar a la utilizada en el algoritmo Fuzzy C-Means: se calcula el promedio de los datos de entrada que caen dentro de cada región. Con los parámetros Φ_0 , producidos a medida que se encuentra una partición óptima, se realiza una optimización final para obtener Φ , que contiene los parámetros definitivos de los consecuentes de las reglas. Esta optimización, denotada por LOF_1 , calcula la falta de ajuste a los datos y se implementó usando un algoritmo de optimización no lineal. Utilizando C y Φ se construye el sistema de inferencia definitivo, de acuerdo con 2.5. El criterio de parada se expone detalladamente en la sección 2.4.

La rutina específica de particionamiento (algoritmo 2.2) comprende un paso hacia adelante para la división en regiones y un paso hacia atrás para eliminar regiones poco significativas. Dado que el cálculo de la pertenencia de los datos a las regiones se hace a través de un sistema de inferencia borroso, habrá grados de pertenencia en cada grupo para cada dato y no quedarán huecos en el espacio de entrada. Por tanto es suficiente el procedimiento de poda del algoritmo 2.4, que es relativamente simple.

Las modificaciones sobre el algoritmo original de particionamiento recursivo se encuentran en el ciclo más anidado del paso hacia adelante en el algoritmo 2.3. La función que se minimiza para este modelo es y , que representa la salida del sistema

Algoritmo 2.1: Pasos generales para la construcción de modelos de inferencia

```

 $h \leftarrow h_{min}$ 
mientras ( $\sim$  CriterioParada ( $e_1, e_2, h, H_{max}$ ))
     $\{\mathbf{B}, \Phi_0\} \leftarrow$  Particionamiento ( $\mathbf{X}, h$ )
     $\mathbf{C} \leftarrow$  Centros ( $\mathbf{B}$ )
     $\Phi \leftarrow \operatorname{argmin}_{\{\phi_{ij}\}_{i=1,\dots,L;j=0,\dots,n}} LOF_1 (y (\mathbf{X} \mid \Phi_0, \mathbf{C}))$ 
     $\{e_1, e_2\} \leftarrow$  Error ( $y (\mathbf{X} \mid \Phi, \mathbf{C})$ )
     $h \leftarrow h + 1$ 
fin mientras

```

Algoritmo 2.2: Particionamiento del espacio de las variables de entrada

```

Particionamiento ( $\mathbf{X}, M_{max}$ ):
 $\mathbf{B} \leftarrow$  PasoHaciaAdelante ( $\mathbf{X}, M_{max}$ )
 $\mathbf{J} \leftarrow$  PasoHaciaAtras ( $\mathbf{X}, \mathbf{B}, M_{max}$ )
 $\Phi_0 \leftarrow \operatorname{argmin}_{\{\phi_{ij}\}_{i \in J^*; j=1,\dots,n}} LOF_2 (y (\mathbf{X} \mid \Phi_r, \text{centros} (B_{i \in J^*})))$ 
 $\mathbf{B} \leftarrow \mathbf{B}_{i \in J^*}$ 

```

Algoritmo 2.3: Paso hacia adelante del particionamiento recursivo

```

PasoHaciaAdelante ( $\mathbf{X}, M_{max}$ ):
 $B_1(\mathbf{x}) \leftarrow 1$ 
Para  $M = 2 : M_{max}$ 
     $lof^* \leftarrow \infty$ 
    Para  $m = 1 : M - 1$ 
        Para  $v = 1 : n$ 
            Para  $t \in \{x_{jv} \mid B_m(\mathbf{x}_j) > 0\}$ 
                 $b_1 \leftarrow B_m[+(x_v - t)]$ 
                 $b_2 \leftarrow B_m[-(x_v - t)]$ 
                 $G \leftarrow \{B_k \mid k \neq m\} \cup b_1 \cup b_2$ 
                 $g \leftarrow y (\mathbf{x} \mid \Phi_r, \text{Centros} (G))$ 
                 $lof \leftarrow \min_{\{\phi_{ij}\}_{i=1,\dots,L;j=0,\dots,n}} LOF_2(g)$ 
                Si  $lof < lof^*$ , entonces  $lof^* \leftarrow lof; m^* \leftarrow m; v^* \leftarrow v; t^* \leftarrow t$  Fin Si
            Fin Para
        Fin Para
    Fin Para
     $B_M(\mathbf{x}) \leftarrow B_{m^*}(\mathbf{x}) H[-(x_{v^*} - t^*)]$ 
     $B_{m^*}(\mathbf{x}) \leftarrow B_{m^*}(\mathbf{x}) H[+(x_{v^*} - t^*)]$ 
Fin Para

```


Algoritmo 2.4: Paso hacia atrás del algoritmo de particionamiento

PasoHaciaAtras (B):

$J^* \leftarrow \{1, 2, \dots, M_{max}\}; K^* \leftarrow J^*$

$lof^* \leftarrow \min_{\{\phi_{ij}\}_{i \in J^*; j=1, \dots, n}} LOF_2(y(\mathbf{X} \mid \Phi_r, \text{centros}(B_{i \in J^*})))$

Para $M = M_{max} : 2$

$b \leftarrow \infty; L \leftarrow K^*$

Para $m = 2 : M$

$K \leftarrow L - \{m\}$

$lof \leftarrow \min_{\{\phi_{kp}\}_{k \in K; p=1, \dots, n}} LOF_2(y(\mathbf{X} \mid \Phi_r, \text{centros}(B_{k \in K})))$

Si $lof < b$, **entonces** $b \leftarrow lof; K^* \leftarrow K$ **Fin Si**

Si $lof < lof^*$, **entonces** $lof^* \leftarrow lof; J^* \leftarrow K$ **Fin Si**

Fin Para

Fin Para

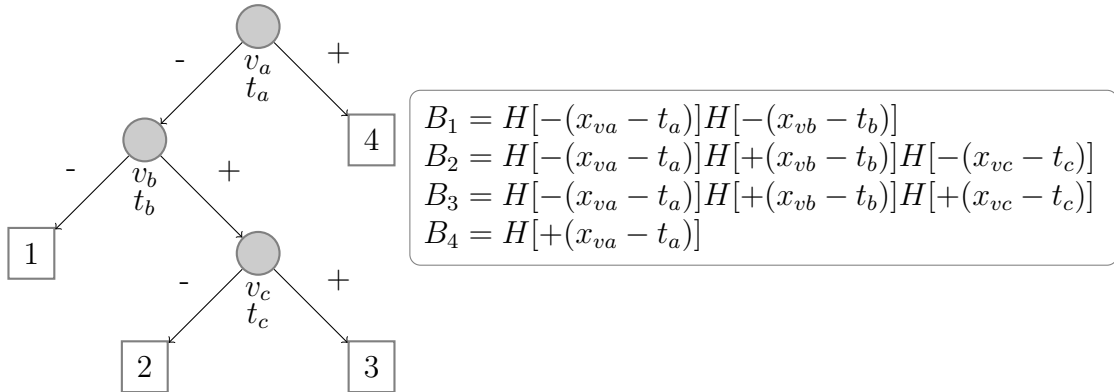


Figura 2.1: Partición y funciones base asociadas.

de inferencia borrosa. Sus parámetros, ϕ_{ij} , corresponden a los coeficientes de las variables en los consecuentes de las reglas. En este algoritmo, \mathbf{X} es una matriz de $m \times n$ donde m es el número de datos de entrada usados en el ajuste de las regiones y parámetros y n es el número de componentes de cada vector de entrada. A su vez, \mathbf{x}_j es el j -ésimo dato de entrada y corresponde a la fila j -ésima de \mathbf{X} y x_{ij} es la componente de \mathbf{X} en la fila i y columna j . Para la optimización (LOF_2) se utilizó mínimos cuadrados ordinarios.

2.3 Metodología y entrenamiento

La metodología seguida para la formulación de modelos se basó en la propuesta originalmente por [7], que ha sido ajustada por diferentes autores (principalmente [66], [27] y [29]) para ser aplicada en la construcción de modelos no lineales. A continuación se especifican los pasos teniendo en cuenta las prácticas mencionadas por [68].

2.3.1 Identificación

Durante la fase de identificación se realiza un análisis exploratorio de la serie que tiene como objetivos la determinación de los regresores con los que se construirá el modelo de inferencia borrosa, la realización de una transformación de los datos que permita al modelo capturar sus propiedades estadísticas y, finalmente, la determinación de las principales propiedades estadísticas de la serie. Como resultado de este proceso, se especifican varios conjuntos posibles de variables de entrada (regresores) al modelo de inferencia borrosa.

2.3.1.1 Determinación de los regresores

La elección del conjunto de regresores se basa profundamente en decisiones subjetivas del modelador y su apreciación del ambiente donde se genera la serie temporal ([10]). En consecuencia, este proceso tiene una alta componente de juicio experto que puede ser viciado por las desviaciones y errores sistemáticos de no realizar un proceso formal de razonamiento, ya que se encuentra sometido a las limitaciones propias de la mente humana (véase [32]). Para mitigar este efecto, se realiza un examen de los autocorrelogramas y autocorrelogramas parciales de la serie temporal, para encontrar información de utilidad en la especificación de los regresores.

2.3.1.2 Determinación de las propiedades estadísticas de la serie

El análisis de las propiedades estadísticas de la serie permite entender de una manera más clara la distribución de los datos y su estructura de dependencia. Las propiedades se estudian con las siguientes herramientas:

1. Los estadísticos descriptivos: media, varianza, curtosis, coeficiente de asimetría de los datos, valores máximo y mínimo, cantidad de observaciones atípicas.
2. El autocorrelograma serial, que mide la relación entre y_t y y_{t+k} :

$$\rho_k \left(\{y_t\}_{t=1}^T \right) = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \cdot \frac{T}{T-k} \quad (2.11)$$

3. El autocorrelograma parcial, que permite determinar la dependencia lineal entre y_t y y_{t+k} , después de que la influencia de los rezagos intermedios se ha eliminado parcialmente. La autocorrelación parcial se calcula a través del sistema de ecuaciones de Yule-Walker.
4. Algunos comportamientos caóticos se pueden determinar a partir de los diagramas de dispersión entre y_t y y_{t+k} . Estos diagramas se construyen uniendo los puntos (y_t, y_{t-j}) y (y_{t+1}, y_{t+1-j}) mediante líneas.
5. Los métodos anteriores se complementan, siguiendo la recomendación en [66], con pruebas formales de linealidad. En el caso general se considera como hipótesis nula que la serie temporal puede ser representada como una combinación lineal del vector de regresores $\mathbf{x}_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})$:

$$H_0 : y_t = \alpha_0 + \mathbf{x}_t \alpha + e_t \quad (2.12)$$

Los errores e_t son independientes e idénticamente distribuidos. Se considera como hipótesis alternativa que existe una relación no lineal remanente en los residuales, que permite una mejor explicación del comportamiento de y_t , que puede detectarse adicionando alguna función no lineal al modelo de regresión lineal en H_0 . En caso de que realmente exista una relación de orden lineal, el modelo aumentado será incapaz de explicar el comportamiento más allá del modelo lineal.

En particular, se utiliza la prueba propuesta en [67], que consta de los siguientes pasos:

- Estimar los parámetros del modelo $y_t = \alpha_0 + \mathbf{x}_t \alpha + e_t$ y calcular la sumatoria del cuadrado de los residuales $SSR_0 = \sum \hat{e}_t^2$
- Realizar la regresión de $\hat{e}_t = \beta_0 + \mathbf{x}_t \beta + \mathbf{w}_t \gamma_t + \vartheta_t$, donde \mathbf{w} es un vector cuyas componentes se obtienen como los productos $x_t^{(i)} \times x_t^{(j)}$, para $i = 1, \dots, p; j = 1, \dots, p$
- Calcular la cantidad $SSR = \sum \hat{\vartheta}_t^2$
- El estadístico

$$F = \frac{(SSR_0 - SSR) / (p(p+1)/2)}{SSR / (T - p - p(p+1)/2)} \quad (2.13)$$

Sigue una distribución F con $p(p+1)/2$ y $T - p - p(p+1)/2 - 1$ grados de libertad.

2.3.1.3 Transformación de la serie temporal

La selección de una transformación adecuada de la serie puede ayudar a minimizar el ruido e indicar relaciones importantes. Además puede preparar los datos en una forma que facilite la estimación del modelo. La mayoría de los modelos son desarrollados bajo la suposición de homogeneidad en el sentido estadístico, lo que implica que la serie temporal debe tener una varianza estable y que la tendencia, el patrón estacional y los ciclos de largo plazo han sido eliminados antes del modelamiento. Para hacer la varianza estable se usan transformaciones como $\log(u)$, \sqrt{u} ó $1/u$ (para más detalle véase [51]). La eliminación de componentes estructurales se realiza a través de diferenciaciones simples y estacionales; sin embargo, este procedimiento asume que dichas componentes son fijas y no son influenciadas por los regresores.

2.3.2 Especificación del sistema de inferencia y estimación de parámetros

Inicialmente se especifica un sistema de inferencia que tiene la estructura establecida en la ecuación 2.5. El número inicial de centros es usualmente $h_{min} = 1$. Dadas las propiedades del modelo, este utiliza funciones lineales de las variables de entrada para explicar el comportamiento de la variable de salida en distintas regiones del espacio de entrada. Se parte de la hipótesis de que en las distintas regiones del espacio de entrada tienen influencia las mismas variables y cambia solamente el valor del coeficiente en cada una de las reglas. En cada paso del algoritmo general, la complejidad del sistema aumenta, al incrementarse el número de centros, hasta que se cumple un criterio de parada y en cada iteración, según la hipótesis mencionada, las funciones lineales de los consecuentes de las reglas tienen el mismo número de variables. Las pruebas realizadas con un número creciente de variables en los consecuentes no han arrojado resultados satisfactorios. Sin embargo su exploración a través de una optimización combinatoria podría mostrar beneficios hasta ahora no encontrados.

En el Algoritmo 2.1 se realiza simultáneamente la determinación de la estructura del sistema y la optimización de sus parámetros. Desde la consecución de las regiones que dan lugar a los centros se lleva a cabo una optimización que establece los parámetros de los consecuentes para cada configuración de regiones. Los detalles de la optimización se mencionan en los Algoritmos 2.1, 2.2, 2.3, y 2.4.

2.3.2.1 Criterio de parada

El algoritmo de entrenamiento presentado (Algoritmo 2.1), debe evitar que el modelo obtenido no esté ajustado de manera suficiente a los datos o que, por el contrario, esté sobre ajustado. Para tal propósito se utilizaron los criterios de parada establecidos en las ecuaciones 2.14 y 2.15, mencionados en [23]. El primero de ellos, e_1 , permite determinar si hay o no poco ajuste. El segundo, e_2 , previene el sobre ajuste. Para el cálculo de e_1 se obtiene la suma de los errores al cuadrado (SSE) en la k -ésima iteración

y se lo compara con el correspondiente en la próxima iteración. Para el cálculo de e_2 es necesario dividir el conjunto de datos de entrenamiento en dos grupos. Uno para el entrenamiento mismo, con el cual se calcula MSE_T (error cuadrático medio de entrenamiento), y el otro para calcular el error de validación (MSE_v). También existe un límite para el número de reglas posibles en un modelo, H_{max} . El modelo se considera entrenado por completo cuando el nivel especificado de precisión se alcanza y, a su vez, se satisfacen los criterios de parada establecidos.

$$e_1 = (SSE_k - SSE_{k+1}) / SSE_k \leq e_1^* \quad (2.14)$$

$$e_2 = |MSE_T - MSE_v| / MSE_T \leq e_2^* \quad (2.15)$$

Los valores e_1^* y e_2^* así como los tamaños de los subconjuntos de datos para entrenamiento y validación, se establecen de manera experimental para cada problema.

2.3.3 Diagnóstico

El proceso de diagnóstico consiste en verificar el cumplimiento de las hipótesis del modelo una vez ha sido ajustado a la serie de datos. Se utilizan dos tipos de pruebas: la emisión de juicios informados basados en la inspección visual de ajuste del sistema de inferencia a la serie y el análisis estadístico de los resultados, que permitirá corroborar los juicios y detectar inconsistencias que no son fácilmente visibles. El análisis de los errores puede dar indicaciones sobre problemas en la especificación del modelo o la selección de los regresores. Sin embargo, este proceso requiere de juicio subjetivo para determinar si el modelo se ajusta apropiadamente a los datos. Este proceso se realiza comúnmente en el modelado estadístico de series temporales, pero aparentemente no es una práctica usual en el modelamiento con herramientas de inteligencia computacional.

Durante esta fase se examinan los residuos del modelo, $\{e_t\}_1^T$, para juzgar si ellos son ruido blanco. El proceso de análisis consta de dos partes: una inspección visual de diferentes gráficos que permiten visualizar el comportamiento de los residuales a lo largo de la serie temporal y una realización y de pruebas estadísticas para verificar que sean efectivamente ruido blanco.

El procedimiento para examinar el comportamiento de los residuales es graficarlos contra el tiempo. Sin embargo, una gráfica de los residuos estandarizados puede ser más útil. Otros procedimientos de utilidad son examinar la gráfica de la suma acumulada de residuos estandarizados y examinar la gráfica de la suma acumulada del cuadrado de los residuos estandarizados. Estas tres gráficas pueden ayudar a detectar cambios estructurales en la serie temporal; regiones con errores persistentemente positivos o negativos o muy altos o bajos en magnitud respecto al resto del gráfico son de interés.

El modelo descrito parte de la hipótesis de normalidad de los errores. Su falta de cumplimiento es un indicativo de que el sistema de inferencia no ha capturado

adecuadamente las propiedades de $\{y_t\}_1^T$. El procedimiento básico para verificar la normalidad es la construcción del histograma de los errores estandarizados y compararlo contra la distribución normal estándar.

Para verificar la normalidad de los errores, se utiliza la prueba propuesta por [36]. El estadístico es

$$N = \frac{T}{6} \sum \frac{\hat{e}_t^6}{T^2} + \frac{T}{24} \left(\left[\sum \frac{\hat{e}_t^4}{T} \right] - 3 \right)^2 \quad (2.16)$$

Que sigue una distribución χ_2^2

2.3.4 Metadiagnóstico

En este proceso se realizan pruebas para determinar la calidad del ajuste en términos de precisión. Los estadísticos comúnmente aceptados son:

- El porcentaje de varianza explicada por el modelo:

$$R^2 = 1 - \sum_{t=1}^T \frac{\hat{e}_t^2}{y_t^2} \quad (2.17)$$

- La varianza del predictor:

$$\sigma_{\hat{e}_t}^2 \sim \frac{1}{T} \sum_{t=1}^T \hat{e}_t \quad (2.18)$$

- La desviación media absoluta (MAD):

$$MAD = \frac{1}{T} \sum_{t=1}^T |\hat{e}_t| \quad (2.19)$$

- También se considera el error absoluto porcentual medio (MAPE):

$$MAPE = \sum_{t=1}^T \left| \frac{\hat{y}_t}{y_t} - 1 \right| \quad (2.20)$$

2.3.5 Construcción de pronósticos e intervalos de confianza

Cuando se ha finalizado el proceso de especificación y validación del sistema de inferencia, se procede a la construcción de pronósticos e intervalos de confianza. El pronóstico para un período adelante, y_{T+1} , sigue usualmente una distribución normal, y se calcula usando 1.3. Cuando x_t contiene uno o más valores rezagados de y_t ,

la construcción del pronóstico j períodos adelante requiere del uso de simulación de Montecarlo (véase [19]), ya que en el sistema de inferencia se realiza una transformación no lineal de los datos de entrada.

Un factor crítico de este proceso es la incorporación de información contextual o cualitativa en el proceso de pronóstico. Una vez construidos los pronósticos, el experto emite un juicio sobre su validez basado en su apreciación subjetiva del proceso generador de la serie temporal. Si el experto considera que el pronóstico no refleja adecuadamente la realidad, se realiza una nueva configuración del modelo. Esta configuración podrá tener un conjunto diferente de regresores o unos niveles de errores e_1^* y e_2^* diferentes a los especificados anteriormente. Se llevan a cabo entonces los pasos indicados a partir de la sección 2.3.1, para la obtención de un nuevo modelo.

Es importante la preparación de una memoria de especificación, ya que pueden revisarse y revalidarse muchos de los supuestos del modelo cuando fue especificado originalmente, haciéndose vital un proceso de documentación ordenado y minucioso.

2.4 Propiedades del modelo

A continuación se mencionan algunas propiedades relevantes que se observaron en el sistema de inferencia borroso multidimensional.

1. *Optimización de regiones y consecuentes de manera simultánea:* El modelo construido está basado en las técnicas de particionamiento recursivo expuestas en [21]. Entre las bondades del uso de este particionamiento se encuentra la optimización simultánea de las regiones en que se divide el espacio de entrada y de los coeficientes de los consecuentes para el sistema de inferencia.
2. *Puntos de partida fijos:* El particionamiento recursivo mitiga la dependencia de puntos de partida aleatorios, que se observa en enfoques basados en algoritmos de particionamiento como *Reubicación de grupos*, *K-means*, *fuzzy C-means*, *Self-organizing maps* (véase [47]) y agrupamiento de *Gustafson-Kessel* (véase [16]).
3. *Capacidad de representación:* Al tener la estructura de un sistema de inferencia borrosa tipo Takagi-Sugeno, el modelo puede representar dinámicas no lineales (véase [62]).
4. *Comparación entre el particionamiento recursivo y los sistemas de inferencia borrosos con conjuntos multidimensionales:* En este punto se muestran algunos conceptos generales de ambos modelos y luego se los compara.

El particionamiento recursivo pretende explicar el comportamiento de una función desconocida en una región del dominio a través de una función de una complejidad menor:

$$\text{Si } (x \in R_m) \text{ entonces } \hat{f}(x) = g_m(x \mid \{a_j\}_1^p) \quad (2.21)$$

La aproximación final a la función modelada se obtiene utilizando las regiones y las funciones que explican el comportamiento de la función en dichas regiones:

$$\hat{f}(x) = \sum_{m=1}^M a_m B_m(x) \quad (2.22)$$

$$B_m(x) = H [x \in R_m] \quad (2.23)$$

En esta formulación, $a_m = g_m$ y es un valor constante para la región dada. B_m es una función base y H es una función que toma el valor de uno si su argumento es cierto o cero en caso contrario.

Por otro lado, en un sistema de inferencia borrosa multidimensional (cuando el sistema es del tipo Takagi-Sugeno), se obtiene una descripción local de f en la región definida por el conjunto borroso:

$$\text{Si } (x \text{ es } A_h) \text{ entonces } \hat{f}(x) = f_h(x) \quad (2.24)$$

A_h es el grado de pertenencia de un elemento x a un conjunto h . La salida del sistema con M reglas, cada una con un conjunto borroso multidimensional, está dada por:

$$\hat{f}(x) = \frac{\sum_{m=1}^M \mu_m(x) f_m(x)}{\sum_{m=1}^M \mu_m(x)} \quad (2.25)$$

En 2.25 usando la expresión $a_m = \frac{f_m(x)}{\sum_{m=1}^M \mu_m(x)}$, se obtiene:

$$\hat{f}(x) = \sum_{m=1}^M a_m(x) \mu_m(x) \quad (2.26)$$

Con base en los conceptos anteriores, las similitudes entre los modelos son las siguientes:

- Los modelos 2.22 y 2.26 tienen la misma estructura y la misma intención. La estructura consiste en la agregación de funciones que describen comportamientos locales de f en regiones del espacio de entrada. La intención es tomar ventaja del hecho de que en regiones bien acotadas del espacio de entrada, la función f puede aproximarse con funciones de poca complejidad. Aunque el comportamiento de f para la región completa del espacio de entrada sea complejo, en regiones pequeñas puede describirse de manera simple.

- En el modelo de particionamiento recursivo 2.22, B_m determina la pertenencia o no de un elemento x a una región m , según se plantea en [21]. La función H toma el valor de uno cuando su argumento es mayor o igual que cero. Y toma el valor de cero en caso contrario. Utilizando el producto de varias funciones H con varias variables, se acota una región. La función que resulta del producto de varias funciones H tendrá el valor de uno sólo para la región acotada y cero para puntos afuera de dicha región. En el modelo de inferencia borrosa con conjuntos multidimensionales 2.26, μ_m determina el grado de pertenencia (entre 0 y 1) de un elemento x a una región m . Por tanto, B_m (cuyos posibles valores son sólo 0 y 1) es un caso especial de μ_m .
- Las funciones g_m del modelo de particionamiento recursivo y las funciones f_m del modelo de inferencia borrosa pueden ser constantes, lineales o de una complejidad mayor. Sin embargo el modelo de particionamiento refiere a las funciones constantes $g_m(x) = a_m$ como las que mejor resultados han dado en problemas de regresión.

2.5 Conclusiones

Dentro de las diversas opciones de construcción de los sistemas de inferencia borrosa con conjuntos multidimensionales, se adoptó aquí un enfoque constructivo, orientado en cada etapa hacia la descripción adecuada de la dinámica que se está modelando. Se evita, de esta manera, el uso de configuraciones iniciales aleatorias adoptadas por otros autores. A partir del resultado obtenido aquí, puede experimentarse con algoritmos de agrupamiento existentes para encontrar posibles resultados favorables en un esquema más complejo.

Las similitudes estructurales encontradas entre el particionamiento recursivo de MARS y los sistemas de inferencia planteados pueden ser un indicativo a favor de la adopción de dichas técnicas de particionamiento.

Capítulo 3

Ejemplificación del uso de la metodología

3.1 Introducción

En este capítulo se ilustra el uso del modelo y las herramientas metodológicas presentados en el capítulo 2. Para tal propósito se utilizó la serie de datos *Airline* (una de las series *benchmark* que posteriormente se usarán para comparar la capacidad de generalización de los modelos formulados). Este conjunto de datos contiene la medida del número mensual de pasajeros en vuelos internacionales desde enero de 1949 hasta diciembre de 1960 (véase la Figura 3.1). Autores como [17] y [24] han usado este conjunto de datos para el entrenamiento de modelos de redes neuronales, dadas las características de no linealidad de la serie y la existencia de ciclos multiplicativos. Estas características la hacen atractiva para la experimentación con modelos no lineales.

Cada sección del capítulo se dedica a un paso general de las indicaciones metodológicas.

3.2 Identificación

3.2.1 Análisis exploratorio de los datos

De la serie (en la Figura 3.1) puede observarse que no es estacionaria y que posee una varianza cambiante. Esta disposición de la varianza sugiere la utilización de alguna transformación que permita tornarla más homogénea. Dado que el aumento de la varianza, a juzgar por el gráfico de la serie, aumenta de manera aproximadamente proporcional al nivel de la serie, puede probarse una transformación logarítmica, que es útil en estos casos.

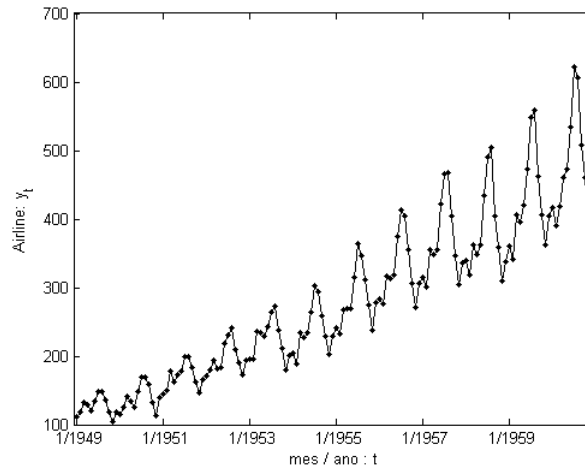


Figura 3.1: Serie Airline.

3.2.2 Algunos estadísticos descriptivos

De los estadísticos descriptivos y el histograma de \log Airline (Figura 3.2), puede verse que los datos no son fuertemente asimétricos y no parece haber una tendencia muy marcada a la existencia de valores extremos. En algunas ocasiones el histograma de los datos y el coeficiente de asimetría pueden dar indicios de posibles agrupaciones de datos que el modelo transformará en conjuntos borrosos. En este caso, sin embargo, es difícil encontrar tales indicios.

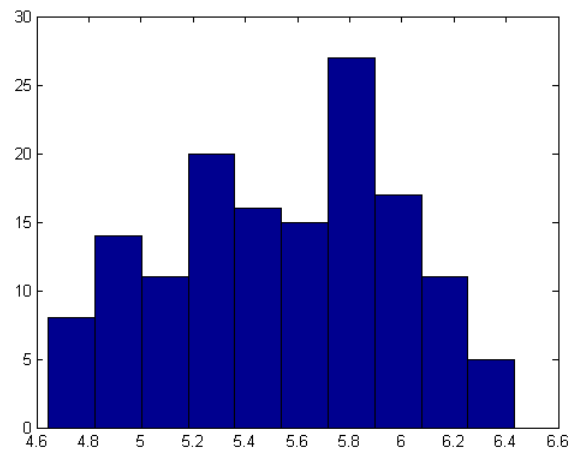
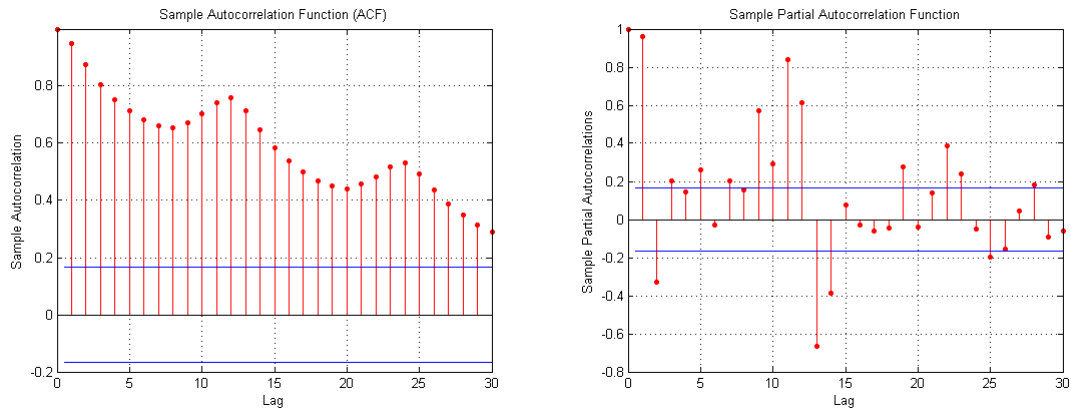


Figura 3.2: Histograma del logaritmo de los datos en la serie Airline.

Tabla 3.1: Estadísticos descriptivos.

Medida (sobre el log de la serie)	Valor
Tamaño de la muestra	144
Media	5.5422
Desviación estándar	0.4415
Curtosis	2.0279
Mínimo	4.6444
Máximo	6.4329
Mediana	5.5816
Coefficiente de asimetría	-0.1207



(a) Autocorrelograma.

(b) Autocorrelograma parcial.

Figura 3.3: Autocorrelogramas para la serie Airline.

3.2.3 Indicaciones sobre la selección de variables

El autocorrelograma y el autocorrelograma parcial pueden dar algunos indicios de rezagos candidatos para la formulación de las variables de entrada al modelo. Sin embargo, el uso de estas herramientas debe ser cuidadoso dado que proporcionan medidas de asociación lineal y pueden existir interacciones no lineales. Sin embargo, de los autocorrelogramas (mostrados en la Figura 3.3) podría extraerse los rezagos 1, 12, 13 para su inclusión en los modelos, dada su evidente importancia.

3.2.4 Estudio de linealidad en la serie

Con un nivel de significancia del 5%, la prueba propuesta en [67] conduce al rechazo de linealidad para los datos de la serie *Airline*. La prueba se realizó con modelos $AR(i)$, $i = 1, \dots, 12$. Dados estos indicios de comportamientos no lineales, es plausible

utilizar un modelo no lineal.

3.3 Especificación del sistema de inferencia y estimación de parámetros

Asumiendo que se carece de conocimiento empírico acerca del fenómeno que se estudia, es conveniente utilizar un modelo no paramétrico. La búsqueda de un modelo no paramétrico adecuado debe realizarse en una clase amplia de los mismos, que incluya redes neuronales, algoritmos evolutivos y sistemas de inferencia borrosa. Dada la variedad de modelos, se acota aquí el espacio de búsqueda a los modelos de inferencia borrosa multidimensionales propuestos en el Capítulo 2. Dichos modelos, descritos en la ecuación 2.5, tienen los siguientes parámetros:

- L : número de reglas del sistema de inferencia,
- w_i : factores de pertenencia normalizados de los datos a los conjuntos borrosos,
- c_i : centros usados en la construcción de los conjuntos borrosos.

La búsqueda de modelos se realiza variando estos parámetros para cada conjunto de rezagos escogidos. El límite L y la cantidad de rezagos determinan la cantidad máxima de parámetros que se permitirán para un modelo. Teniendo en cuenta que los autocorrelogramas sugieren la inclusión de y_{t-1} , y_{t-12} y y_{t-13} , se estableció en 4 el número máximo de rezagos incluidos, lo cual deja lugar para una variable más. Dada esta restricción, un límite L prudente sería 10, lo cual admitiría modelos de hasta 40 parámetros. Con estas restricciones, se realizó una exploración combinatoria del espacio de soluciones, comprendiendo grupos de 1,2,3 y 4 rezagos contenidos en $R = \{y_{t-1}, y_{t-2}, \dots, y_{t-k}\}$, con $k = 13$ (también sugerido por el autocorrelograma y autocorrelograma parcial). De los 1092 modelos obtenidos se muestran en la Tabla 3.2 los veinte primeros, seleccionados debido a que tuvieron un AIC (*Akaike information criteria*) menor. Se incluyen en dicha tabla algunas cifras útiles para la etapa de diagnóstico, como las pruebas de *Jarque-Bera* y *Lin-Mudholkar* para verificar la normalidad de los errores.

3.4 Diagnóstico

Algunos de los modelos no superaron las pruebas de normalidad en los residuales y otros tendían a ser inestables en sus parámetros, según las pruebas CUSUM y CUSUMQ (véase la columna de comentarios en la Tabla 3.2). Los modelos que superaron estas pruebas mostraron, no obstante, alguna estructura en sus residuales, reflejada en el autocorrelograma y en el gráfico de residuales. Por tanto, se probaron algunas variaciones de estos últimos para encontrar uno que explicara mejor

Tabla 3.2: Modelos con menor AIC para la serie de datos log Airline.

Modelo	Rezagos	Jarque-Bera (valor p)	Lin-Mudholkar (valor p)	Akaike IC	Comentarios Diag.
781	1,12,13	0.06790	0.01270	-3.78635	Falla cusumq.
164	1,5,12,13	0.04880	0.01510	-3.77556	Falla test norm.
185	1,6,12,13	0.06590	0.01100	-3.77354	Falla cusumq.
200	1,7,12,13	0.08170	0.01060	-3.77204	Falla cusumq.
136	1,4,12,13	0.03770	0.01360	-3.76813	Falla test norm.
220	1,11,12,13	0.08250	0.02230	-3.76010	Falla cusumq.
210	1,8,12,13	0.02490	0.01570	-3.75827	Falla test norm.
216	1,9,12,13	0.07350	0.01210	-3.75641	Falla cusumq.
219	1,10,12,13	0.10680	0.02100	-3.75379	Falla cusumq.
55	1,2,12,13	0.07630	0.01220	-3.75314	Falla cusumq.
100	1,3,12,13	0.21710	0.01990	-3.71062	Falla cusumq.
157	1,5,9,12	0.89300	0.06810	-3.47260	Ok diag. Mejorar.
160	1,5,10,12	0.78040	0.08020	-3.35235	Ok diag. Mejorar.
162	1,5,11,12	0.82040	0.09740	-3.34854	Ok diag. Mejorar.
481	3,7,10,12	0.01080	0.03950	-3.33511	Falla cusumq.
752	1,5,12	0.73000	0.10510	-3.33379	Ok diag. Mejorar.
153	1,5,8,12	0.85100	0.00790	-3.32234	Falla test norm.
114	1,4,6,12	0.97320	0.02550	-3.29597	Falla test norm.
132	1,4,10,12	0.91270	0.06850	-3.28617	Ok diag. Mejorar.
148	1,5,7,12	0.91100	0.00660	-3.28385	Falla test norm.

la dinámica de la serie. En la Tabla 3.3 se resumen los resultados de los modelos afinados (sólo se presentan aquellas variaciones que mostraron ser exitosas según las herramientas de diagnóstico).

Tabla 3.3: Variaciones de modelos seleccionados.

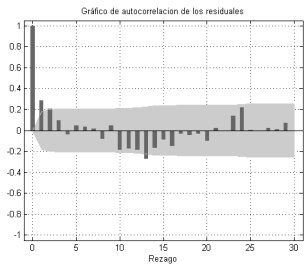
Modelo	Rezagos	Jarque-Bera (valor p)	Lin-Mudholkar (valor p)	Akaike IC
160-1	1, 2, 5,10,12	0.8479	0.1505	-3.29961
162-1	1,12,14	0.3651	0.156	-3.38958
162-2	1 ,5,11,12,14	0.1026	0.2117	-3.48957
752-1	1,5,12,14	0.7065	0.1032	-3.41507

Al observar los resultados obtenidos durante el ajuste de los modelos seleccionados en la Tabla 3.2 se encontró que la inclusión del rezago 13 llevaba en la mayoría de los casos a modelos inestables. Se decidió probar con el rezago 14 y se obtuvo mejores resultados. En la Figura 3.4 se presentan algunas herramientas de diagnóstico para el modelo 752-1, que mostró mejor comportamiento en cuanto a la carencia de estructura en el autocorrelograma de residuales.

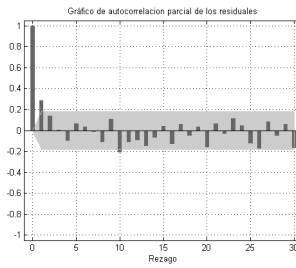
El autocorrelograma (Figura 3.4(a)) y diagrama de residuales (Figura 3.4(e)) muestran indicios de alguna estructura, pero no parece ser significativa dado que no hay correlaciones que superen las bandas de confianza. Los demás autocorrelogramas (Figuras 3.4(b), 3.4(c), 3.4(d)) no muestran evidencia de relaciones lineales remanentes. Por otro lado, el histograma de residuales y el gráfico de probabilidad normal (Figuras 3.4(f), 3.4(g)) no muestran desviaciones significativas con respecto a una secuencia de ruido blanco. Finalmente, en las Figuras 3.4(h) y 3.4(i) no se traspasa las bandas de confianza de 0.01, 0.05 y 0.10, lo cual verifica la estabilidad de los parámetros en el modelo. En la inspección de los resultados de diagnóstico para los demás modelos seleccionados no se halló diferencias significativas con respecto a los resultados mostrados para el modelo 752-1.

3.5 Metadiagnóstico

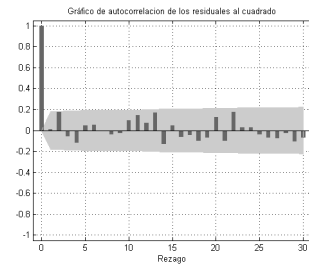
Los modelos seleccionados se muestran en la Tabla 3.4. De ellos, 162-1, 162-2 y 752-1 presentan lo menores errores en el período de validación, teniendo en cuenta las medidas MAD (*mean absolute deviation*) y MAPE (*mean absolute percentage error*), por lo cual son buenos candidatos para realizar pronóstico. El porcentaje de varianza explicada por los modelos (R^2) es alto en todos ellos, de manera que no se tomó como un criterio de selección.



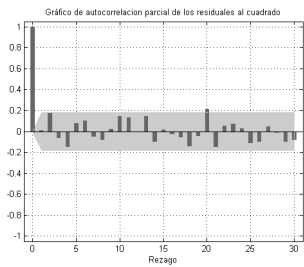
(a) Autocorrelograma.



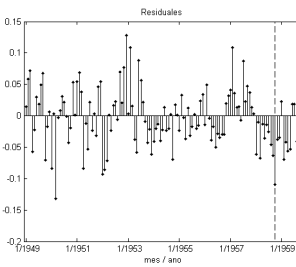
(b) Autocorrelograma parcial.



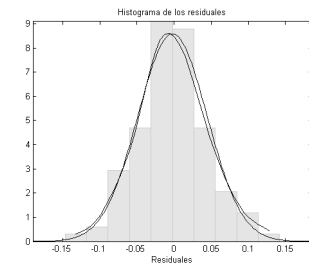
(c) Autocorrelograma res. al cuadrado.



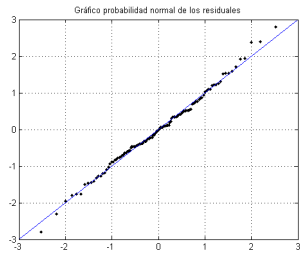
(d) Autocorrelograma parc. res. al cuadrado.



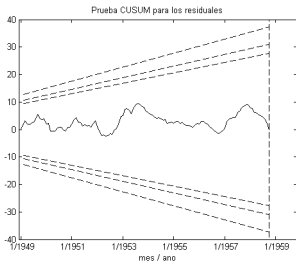
(e) Residuales.



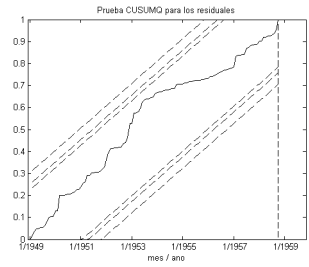
(f) Hist. residuales.



(g) Gráfico prob. normal.



(h) CUSUM.



(i) CUSUMQ.

Figura 3.4: Herramientas gráficas de diagnóstico para el modelo 752-1.

Tabla 3.4: Modelos seleccionados en la etapa de diagnóstico.

Modelo	Medidas de entrenamiento MAD (MAPE) (R^2)	Medidas de predicción MAD (MAPE) (R^2)
157	0.03411 (0.00623) (0.99238)	0.03674 (0.00597) (0.99949)
160	0.03756 (0.00684) (0.99086)	0.03930 (0.00641) (0.99940)
160-1	0.03664 (0.00668) (0.99098)	0.04282 (0.00698) (0.99927)
162	0.03710 (0.00675) (0.99084)	0.03880 (0.00633) (0.99940)
162-1	0.03861 (0.00704) (0.99014)	0.03361 (0.00543) (0.99945)
162-2	0.03686 (0.00672) (0.99051)	0.02887 (0.00469) (0.99965)
752	0.03852 (0.00703) (0.99041)	0.04422 (0.00719) (0.99927)
752-1	0.03625 (0.00659) (0.99146)	0.03529 (0.00572) (0.99951)
132	0.03795 (0.00693) (0.99037)	0.04249 (0.00694) (0.99925)

3.6 Construcción de pronósticos e intervalos de confianza

La Figura 3.5 muestra los valores reales de la serie (en color negro) y los valores pronosticados un paso adelante (en color azul) para el modelo 162-2 (incluido entre los modelos seleccionados). Se proporcionan también, con líneas continuas, las bandas de confianza del 95%. Según se aprecia, los intervalos de confianza son ajustados, lo cual indica una buena captura de la dinámica de la serie.

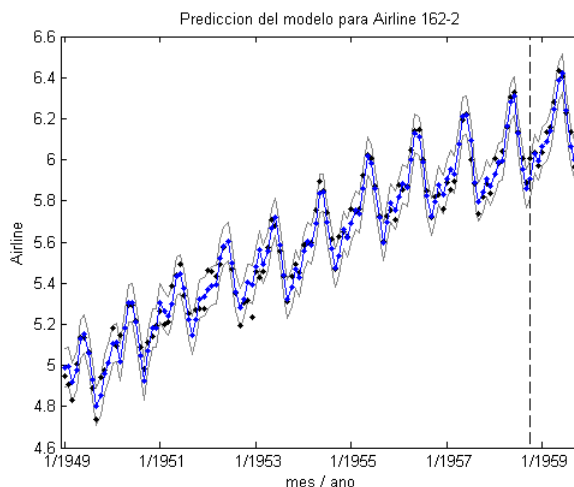


Figura 3.5: Predicción para el modelo 162-2.

3.7 Conclusiones

Las indicaciones metodológicas utilizadas en la construcción de un modelo para la serie *Airline* permitieron juzgar de manera objetiva las diversas opciones encontradas. Se destaca la utilidad de los criterios de selección de modelos en casos como estos en los que la cantidad de modelos alternativos es alta.

El proceso de construcción de modelos de series temporales no deja de tener componentes importantes de juicio subjetivo en algunas etapas. Sin embargo, es sumamente valioso el apoyo encontrado en herramientas metodológicas ya probadas, como las que se ejemplifican aquí, dado que permiten tomar decisiones más informadas en cuanto a los modelos que se construyen.

Capítulo 4

Pronóstico de series benchmark con sistemas de inferencia borrosa multidimensionales

4.1 Introducción

En capítulos anteriores se han mostrado una metodología y un modelo, basados en sistemas de inferencia borrosa y conjuntos borrosos multidimensionales, propuestos para el modelado y pronóstico de series temporales. Con el objetivo de establecer la efectividad de estos modelos se realizaron varios experimentos con series temporales comúnmente usadas en la evaluación de modelos no lineales. Al ser precisamente series cuya no linealidad se ha estudiado y que el propósito del capítulo es realizar comparativos entre los modelos, no se realizaron pruebas de no linealidad. Las medidas de desempeño utilizadas son SSE, MSE, MAD y MAPE, siendo las dos primeras las más comunes en los trabajos revisados.

En la construcción de un modelo hay varias decisiones que deben tomarse. Si se trata de un modelo basado en redes neuronales, se tienen variables de decisión como el número de capas ocultas, el número de nodos en cada capa y el número de variables de entrada. En un modelo de inferencia borrosa puede requerirse establecer el número de reglas y el número de conjuntos borrosos en cada regla. En el modelo MFIS sólo es necesario establecer el número de variables de entrada, ya que la estructura de los modelos se determina a través del algoritmo de entrenamiento y el uso de conjuntos borrosos multidimensionales hace innecesario establecer una configuración de conjuntos borrosos unidimensionales en cada regla¹. La única variable de decisión que debe determinar el modelador (el número de variables de entrada), se determina con base en las distintas configuraciones de rezagos utilizadas en los trabajos tomados como referencia. Cabe anotar que algunos de los modelos obtenidos en este capítulo

¹Los modelos DAN2 en [24] cuentan con la propiedad similar de no requerir la determinación del número de capas ocultas de la red.

mostraron signos de inestabilidad en los parámetros o no superaron algún criterio de diagnóstico, según las indicaciones metodológicas sugeridas en el Capítulo 2. Sin embargo, un examen a tal nivel de detalle no se encontró en los trabajos de los cuales se tomaron los resultados para realizar comparativos y, además, el propósito del presente capítulo es la evaluación de la capacidad de generalización de los modelos.

En las secciones siguientes se presentan los resultados obtenidos para las series de datos *Airline*, *Sunspot* y *Pollution equipment*. Al final del capítulo se presentan algunas conclusiones con base en los resultados obtenidos con dichas series.

4.2 The Airline Passenger data set

Este conjunto de datos contiene la medida del número mensual de pasajeros en vuelos internacionales desde enero de 1949 hasta diciembre de 1960 (véase la Figura 4.1). Autores como [17] y [24] han usado este conjunto de datos para el entrenamiento de modelos de redes neuronales, dadas las características de no linealidad de la serie y la existencia de ciclos multiplicativos. Estas características la hacen atractiva para la experimentación con modelos no lineales.

En [17] se estudian varios modelos de redes neuronales con este conjunto de datos y se menciona el requerimiento de una alta componente de juicio por parte del experto para encontrar un modelo de redes neuronales adecuado. En [24] se presentan mejoras al proponer mecanismos automáticos para determinar la estructura de la red. Ambos trabajos contienen referencias a trabajos previos con este set de datos y por tanto se han tomado como referencia para la comparación de los resultados obtenidos con el modelo MFIS.

En los trabajos revisados se utiliza una transformación logarítmica para mitigar la variabilidad creciente en la serie y usualmente se utilizan los primeros 11 años (132 datos) para el ajuste del modelo, dejando las últimas 12 observaciones para pronosticar un paso adelante. Con esta configuración de datos se probaron varios modelos en los que se utilizan distintos rezagos. Se adoptó aquí la misma transformación y la misma división de los datos para varios modelos MFIS. También se utilizaron las configuraciones de rezagos probadas por los autores en dichos trabajos, con el ánimo de comparar los resultados publicados con los obtenidos aquí. En [17] se menciona el modelo ARIMA de orden $(0, 1, 1)(0, 1, 1)_{12}$ como el mejor de este tipo obtenido para la serie y se utiliza como punto de referencia para los resultados en dicho trabajo. En [24] se compara los resultados con respecto a los obtenidos con los modelos propuestos en [17]. En la Tabla 4.1 se muestran los resultados de los modelos obtenidos en [17] y en [24] (corresponden a los etiquetados como NN ó DAN2).

Cada modelo considerado puede alcanzar sus valores óptimos con base en su estructura interna y dado que esta difiere en los trabajos estudiados (incluyendo diferencias como el número de neuronas en la capa oculta de la red o el número de reglas del sistema de inferencia), no se consideran los parámetros de dicha estructura en las comparaciones, sino que es necesario utilizar medidas de desempeño para evaluar

la capacidad de generalización de los modelos. En los estudios considerados para la serie Airline se utiliza la suma de los errores al cuadrado (SSE) para entrenamiento y pronóstico. Se adopta aquí el mismo criterio de comparación para evaluar la capacidad de generalización del modelo MFIS. Con base en los rezagos utilizados en los estudios mencionados, se generaron 5 modelos MFIS, cuyos resultados se consignan al final de la Tabla 4.1.

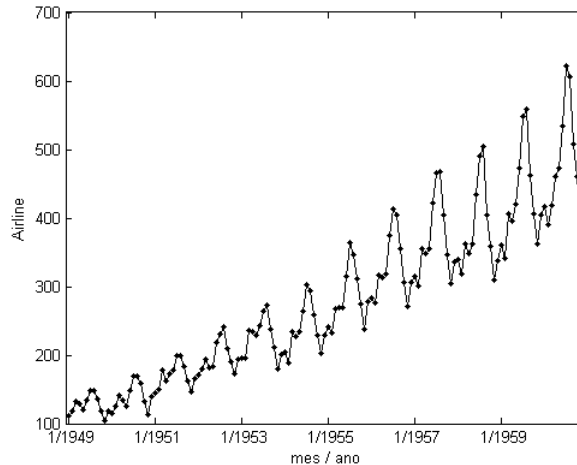


Figura 4.1: Serie Airline.

Al comparar los resultados, debe tenerse en cuenta que el modelo DAN2-1 equivale al MFIS-1, el DAN2-2 al MFIS-2 y así sucesivamente. A su vez, el modelo NN3 de la Tabla 4.1 equivale al modelo MFIS-1, el modelo NN4 equivale al modelo MFIS-2, y así sucesivamente. El único criterio de comparación es el conjunto de rezagos utilizados en la construcción del modelo. Con respecto a los modelos de redes neuronales y los modelos DAN2, se obtuvieron mejoras significativas tanto en el SSE de entrenamiento como en el SSE de pronóstico, lo cual sugiere un mejor ajuste. Los modelos NN10 y NN12 de la Tabla 4.1 se consideran como los de mejor desempeño en [17]. Con respecto al modelo NN10, MFIS-4 presenta un SSE de entrenamiento 77% menor y un SSE de pronóstico 95% menor. Con respecto al modelo NN12 se tiene disminuciones de 82% y 95% en el SSE de entrenamiento y pronóstico respectivamente. Los modelos DAN2-4 y DAN2-5 también ofrecen mejoras con respecto a los modelos NN10 y NN12. Con respecto a ellos se tiene las siguientes mejoras: el modelo MFIS-4 muestra una disminución del 37% en el SSE de entrenamiento con respecto al DAN2-4 y del 91% en el SSE de pronóstico. De los modelos MFIS presentados en la Tabla 4.1, el MFIS-3 muestra un menor SSE de pronóstico y un menor MAD (mean absolute deviation). En la Figura 4.2 se muestran los valores reales de la serie y los valores pronosticados con dicho modelo.

Tabla 4.1: Desempeño de modelos para la serie de datos Airline.

Modelo	Rezagos	Número de neuronas ocultas	Medidas de entrenamiento	Medidas de predicción
			SSE(MAD)(MAPE)	SSE(MAD)(MAPE)
NN1	1,2,3,4	2	7.74 (N/A) (N/A)	1.03 (N/A) (N/A)
NN2	1...13	2	0.73 (N/A) (N/A)	0.71 (N/A) (N/A)
NN3	1...13	4	0.26 (N/A) (N/A)	1.12 (N/A) (N/A)
NN4	1,12	2	2.3 (N/A) (N/A)	0.34 (N/A) (N/A)
NN5	1,12	4	2.16 (N/A) (N/A)	0.44 (N/A) (N/A)
NN6	1,12	10	1.77 (N/A) (N/A)	0.59 (N/A) (N/A)
NN7	1,2,12	2	2.17 (N/A) (N/A)	0.29 (N/A) (N/A)
NN8	1,2,12	4	1.91 (N/A) (N/A)	1.03 (N/A) (N/A)
NN9	1,2,12,13	2	0.99 (N/A) (N/A)	0.52 (N/A) (N/A)
NN10	1,2,12,13	4	0.81 (N/A) (N/A)	0.52 (N/A) (N/A)
NN11	1,12,13	1	1.18 (N/A) (N/A)	0.5 (N/A) (N/A)
NN12	1,12,13	2	1.03 (N/A) (N/A)	0.5 (N/A) (N/A)
NN13	1,12,13	4	0.84 (N/A) (N/A)	0.62 (N/A) (N/A)
DAN2-1	1...13		0.17 (0.009) (N/A)	0.23 (0.024) (N/A)
DAN2-2	1,12		0.85 (0.017) (N/A)	0.26 (0.024) (N/A)
DAN2-3	1,2,12		0.44 (0.014) (N/A)	0.19 (0.021) (N/A)
DAN2-4	1,2,12,13		0.3 (0.011) (N/A)	0.29 (0.03) (N/A)
DAN2-5	1,12,13		0.24 (0.012) (N/A)	0.22 (0.024) (N/A)
MFIS-1	1...13		0.17 (0.029) (0.005)	0.032 (0.04) (0.006)
MFIS-2	1,12		0.36 (0.044) (0.008)	0.017 (0.034) (0.005)
MFIS-3	1,2,12		0.36 (0.043) (0.008)	0.017 (0.032) (0.005)
MFIS-4	1,2,12,13		0.19 (0.031) (0.006)	0.027 (0.035) (0.006)
MFIS-5	1,12,13		0.19 (0.031) (0.006)	0.027 (0.035) (0.006)

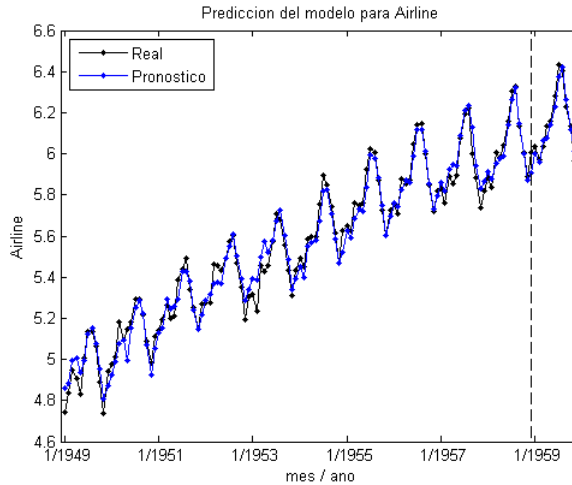


Figura 4.2: Predicción para la serie de datos Airline.

4.3 The Sunspot data set

El conjunto de datos contiene el número anual de manchas solares desde 1700 hasta 1987. La serie de datos se considera no lineal y no Gaussiana (véase [72]) y se ha utilizado con frecuencia en la evaluación de modelos no lineales. En la Figura 4.3 se aprecia la serie. Tal como se menciona en [72], se percibe un patrón cíclico de unos 11 años. Este coincide con los ciclos de 11 y 11.1 años encontrados por S.H. Schwabe y Johann Rudolf Wolf, respectivamente, en sus estudios (véase [8]). Los datos del período 1700-1920 (221 datos) se han utilizado para evaluar el desempeño de modelos lineales y no lineales, incluyendo modelos basados en redes neuronales: se toma el período mencionado (221 puntos) para entrenamiento y los datos entre 1921 y 1956 (35 puntos) para pronóstico. En [24] y Zhang [72] se han realizado pruebas con un horizonte de pronóstico adicional, más amplio, incluyendo los datos entre 1921 y 1987 (67 puntos) y en ambos trabajos se utilizan los datos sin realizar transformación alguna. La evaluación del modelo MFIS se realizó utilizando los dos horizontes de tiempo mencionados y los datos se dejaron sin transformar.

En la Tabla 4.2 se muestran los resultados de varios modelos construidos para la serie *Sunspot*. Las medidas de desempeño disponibles son el MSE y el MAD. Al final de dicha tabla se resumen los resultados obtenidos con el modelo MFIS. De nuevo, el criterio para comparar un modelo con otro es el conjunto de valores rezagados de la serie que el modelador utilizó en sus experimentos. Los modelos obtenidos muestran un alto grado de precisión y, al compararlos con los demás, se encuentran mejoras con respecto a los de redes neuronales, híbridos y ARIMA. En el caso de los modelos DAN2, se obtuvieron mejoras para dos de los tres modelos reportados en [24]. El modelo MFIS-4, por ejemplo, muestra una disminución del 39% en el

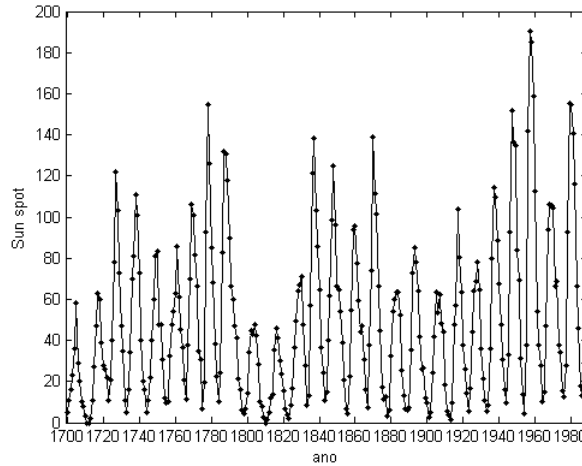


Figura 4.3: Serie Sonspot.

MSE de pronóstico para el horizonte de tiempo de 35 puntos, con respecto al modelo híbrido, y una disminución del 21% en el MSE para el pronóstico en el horizonte de 67 puntos hacia adelante. Al comparar los resultados con respecto a los modelos DAN2, teniendo en cuenta la correspondencia de los rezagos usados, se observa un aumento en el MSE de entrenamiento, acompañado de mejoras apreciables en los MSE de pronóstico en el corto y largo plazo con respecto a los modelos DAN2-2 y DAN2-3. Con respecto DAN2-2, por ejemplo, se obtuvo una disminución del MSE de pronóstico en el corto plazo del 23% y del 8% en el largo plazo. Y con respecto al DAN2-3 se tienen disminuciones del 32% y 17% en el MSE de pronóstico para el corto y largo plazo, respectivamente, lo cual indica una alta capacidad de ajuste y generalización para la serie.

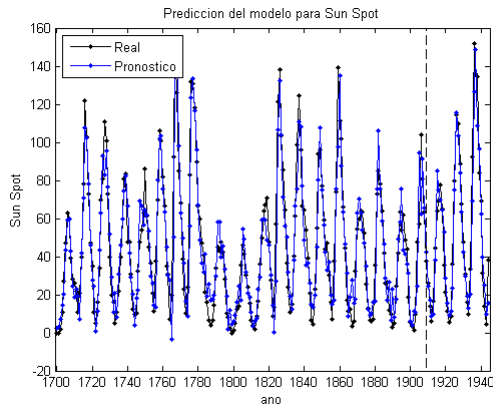
En las Figuras 4.4(a) y 4.4(b) se muestran los valores reales de la serie y los valores pronosticados con el modelo MFIS-4 para los horizontes de pronóstico de 35 y 67 puntos hacia adelante, respectivamente.

4.4 The Pollution Equipment data set

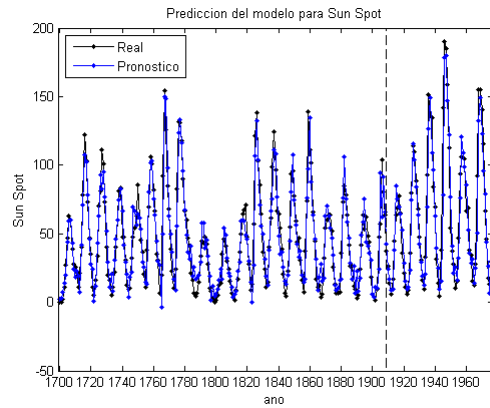
Este conjunto de datos contiene las cifras de envíos mensuales de equipos de polución desde enero de 1986 hasta octubre de 1996. Son 130 observaciones, medidas en miles de francos franceses, presentadas en [48]. En la Figura 4.5, donde se muestran los datos, pude observarse que la serie no es estacionaria en varianza. Debido a esto, en [24] y [48] se usó una transformación logarítmica antes de construir los modelos. En la Tabla 4.3 se muestran los resultados obtenidos con varios modelos, según se reporta en [24]. Uno de ellos es un modelo ARIMA $(2, 1, 0)(1, 0, 0)_{12}$, mientras que otros son de redes neuronales. En todos ellos se consideró dos horizontes de pronóstico:

Tabla 4.2: Desempeño de modelos para la serie de datos Sunspot.

Modelo	Rezagos	Medidas de Entrenamiento	Medidas de Predicción	
		MSE (MAD)(MAPE)	35 pts hacia adelante MSE (MAD)(MAPE)	67 pts hacia adelante MSE (MAD)(MAPE)
ARIMA	N/A	N/A (N/A) (N/A)	217 (11.3) (N/A)	306 (13) (N/A)
ANN	N/A	N/A (N/A) (N/A)	205 (10.2) (N/A)	351 (13.5) (N/A)
Hybrid	N/A	N/A (N/A) (N/A)	187 (10.8) (N/A)	280 (12.8) (N/A)
DAN2-1	1,3,4,9,10,11	78 (7) (N/A)	145 (9.7) (N/A)	258 (12) (N/A)
DAN2-2	1,2,9,11	95 (7.4) (N/A)	146 (9.6) (N/A)	266 (12.3) (N/A)
DAN2-3	1,2,3,9,10,11	120 (8.4) (N/A)	186 (9.9) (N/A)	267 (12.4) (N/A)
MFIS-2	1,3,4,9,10,11	158 (9.60) (0.55)	152 (9.3) (0.29)	336 (12.9) (0.3)
MFIS-3	1,2,9,11	119 (8.50) (0.52)	113 (7.6) (0.19)	244 (11.5) (0.25)
MFIS-4	1,2,3,9,10,11	142 (9.00) (0.5)	127 (8.7) (0.27)	221 (10.7) (0.26)



(a) 35 puntos hacia adelante.



(b) 67 puntos hacia adelante.

Figura 4.4: Predicción para la serie de datos Sunspot.

uno de un año y otro de dos años.

En la construcción de los modelos MFIS se utilizaron los mismos horizontes de pronóstico y se realizó la transformación de los datos utilizada en los trabajos mencionados. En los resultados obtenidos (véase los modelos finales de la Tabla 4.3) se observan mejoras con respecto al MSE de entrenamiento y de pronóstico para ambos horizontes de tiempo, con respecto a los modelos ARIMA y el de redes neuronales. En el caso del modelo ARIMA, el MSE de entrenamiento es comparable con el obtenido con el modelo MFIS. El MSE de pronóstico para un año se redujo en un 12% y el MSE de pronóstico para dos años se redujo en un 82%. En cuanto al modelo de redes neuronales, se obtuvo una reducción del MSE de entrenamiento del 13%; del 88% en el MSE de pronóstico para un año y del 69% en el MSE de pronóstico a dos años. Con respecto a los modelos DAN2 no se obtuvieron mejoras. El modelo MFIS-3 se considera el de mejor desempeño entre los modelos MFIS obtenidos, dado que el MSE de entrenamiento es más bajo que para los demás modelos MFIS, el MSE de pronóstico en el corto plazo es comparable al de los demás modelos y el MSE de pronóstico en el horizonte de tiempo largo es menor. En la Figura 4.6(a) se muestran los valores reales y los valores pronosticados con dicho modelo, para el horizonte de tiempo de un año. En la Figura 4.6(b) se muestran los resultados para el horizonte de tiempo de dos años. En ambas figuras puede observarse que, aunque hay un aprendizaje de la dinámica de la serie (de manera general), el ajuste no es tan bueno como en el caso de los modelos construidos para las series *Airline* y *Sunspot*. Sin embargo, los resultados del modelo MFIS con respecto a los modelos de referencia, muestran un desempeño medio.

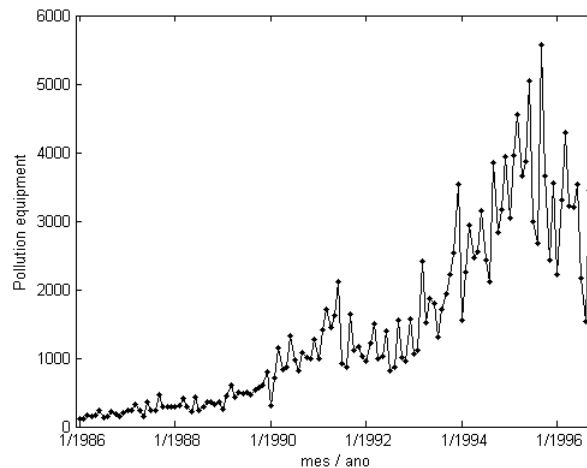
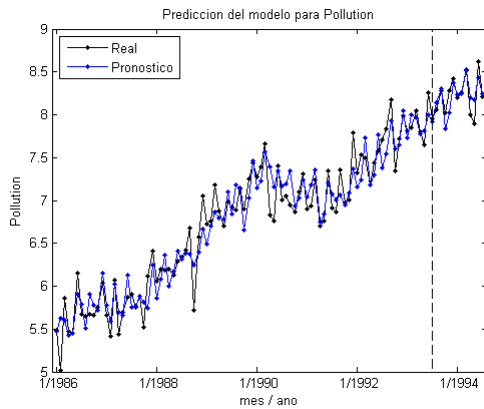


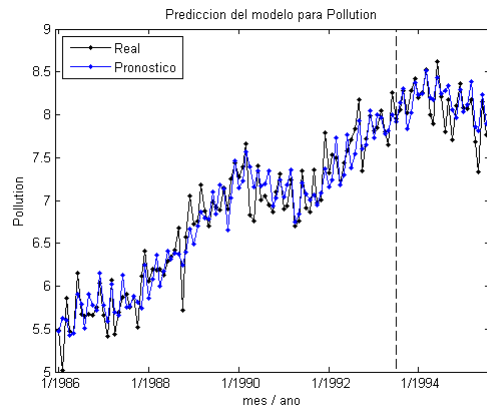
Figura 4.5: Serie Pollution equipment.

Tabla 4.3: Desempeño de modelos para la serie de datos Pollution equipment.

Modelo	Rezagos	Medidas de Entrenamiento		Medidas de Pronóstico	
		MSE (MAD)(MAPE)	MSE (MAD)(MAPE)	1 año MSE (MAD)(MAPE)	2 años MSE (MAD)(MAPE)
ARIMA	1,2,3,12,13,14,15	0.052 (0.181)	0.025 (0.135)	0.268 (0.395)	
ANN	1...12	0.054 (0.188)	0.193 (0.394)	0.146 (0.334)	
DAN2-1	1,2,3,12,13,14,15	0.02 (0.101)	0.015 (0.049)	0.025 (0.113)	
DAN2-2	1...12	0.019 (0.108)	0.018 (0.106)	0.02 (0.115)	
DAN2-3	1...15	0.013 (0.085)	0.01 (0.079)	0.023 (0.119)	
MFIS-1	1,2,3,12,13,14,15	0.051 (0.178) (0.027)	0.022 (0.118) (0.014)	0.047 (0.164) (0.02)	
MFIS-2	1...12	0.047 (0.177) (0.027)	0.023 (0.109) (0.013)	0.045 (0.16) (0.02)	
MFIS-3	1...15	0.044 (0.166) (0.025)	0.023 (0.115) (0.014)	0.042 (0.157) (0.02)	



(a) 1 año hacia adelante.



(b) 2 años hacia adelante.

Figura 4.6: Predicción para la serie de datos Pollution equipment.

4.5 Conclusiones

En este capítulo se han realizaron pruebas con diferentes series *benchmark* con el fin de establecer la efectividad de los modelos MFIS en la modelación de series temporales. Los datos utilizados corresponden a series usadas comúnmente en la evaluación de modelos de predicción y exhiben características de no linealidad y no estacionalidad adecuadas para probar la capacidad de generalización de los modelos. En la mayoría de los casos examinados se obtuvieron desempeños superiores con los modelos MFIS, evidenciados en errores de pronóstico menores, lo cual indica una alta capacidad de generalización. Con una de las series, sin embargo, se observó un desempeño menor a uno de los modelos competidores (DAN2), aunque mostrando mejoras con respecto a los demás modelos para la misma serie.

La efectividad se estableció usando medidas como MSE, MAD y MAPE. Aunque, en general, se encontraron resultados superiores con el modelo MFIS, se requiere de una mayor experimentación para establecer sus bondades o falencias en el estudio de fenómenos que se hayan modelado a través de series temporales.

Capítulo 5

Modelado de la demanda mensual de electricidad en Colombia usando sistemas de inferencia borrosa multidimensionales

5.1 Introducción

Después de evaluar en el capítulo previo los modelos MFIS propuestos con algunas series *benchmark*, en este capítulo se realiza una aplicación a un caso práctico: la demanda mensual de electricidad en Colombia. Se pretende estudiar la capacidad del modelo para capturar la dinámica de la serie y las posibilidades de avance en el entendimiento del problema que pueda llegar a tenerse usando modelos MFIS. A continuación se realiza una introducción breve al problema de la demanda de electricidad.

La importancia que posee la energía eléctrica en el desarrollo de las sociedades es innegable y como consecuencia de esta importancia se han querido buscar mejores servicios por parte de los agentes que intervienen en su producción y distribución. Surgen entonces requerimientos en cuanto a factores que determinan la calidad del servicio, como la capacidad de suministrar la energía con un número mínimo de interrupciones y con los menores costos posibles (véase [6]). La capacidad de determinar en el largo y corto plazo cuál será la demanda de energía eléctrica es un componente de gran importancia en la consecución de estos requerimientos. También es de suma importancia la capacidad que tengan los agentes del mercado eléctrico de entender el comportamiento de la demanda.

Varios autores han señalado características de los mercados de energía eléctrica que muestran las complejidades y particularidades en el modelamiento de este problema frente al comportamiento de otros mercados. Se tienen por ejemplo los procesos de desregularización y liberalización tendientes a conseguir una atención oportuna de la

demanda (véase [61]), la existencia de avances tecnológicos para optimizar el uso de la energía, los esfuerzos gubernamentales por conseguir también un uso eficiente de la energía, la existencia de patrones climáticos con efectos importantes sobre la demanda en algunos mercados y la presencia de complejidades particulares para diferentes niveles de agregación temporal, entre otras características (véase [20]).

Para modelar y pronosticar la demanda de electricidad con sus complejas características se han utilizado modelos estadísticos (como en [12]) y de inteligencia computacional (véase por ejemplo [1] y [46]), con diferentes niveles de agregación temporal. En algunos enfoques se usan variables explicativas y en otros se utilizan técnicas univariadas. En la Tabla 5.1 se reproducen los resultados de una revisión de técnicas realizada en [20]. La aplicación que se presenta aquí se suma a las realizadas en área de inteligencia computacional. Se explora la capacidad de los modelos MFIS para describir el comportamiento de la variable bajo estudio con modelos simples de regresión en regiones particulares del espacio de entrada.

El propósito de este capítulo es modelar la demanda mensual de energía eléctrica en Colombia, con modelos MFIS, observar las propiedades de los modelos, determinar su ayuda en la comprensión de la serie y comparar los resultados con los arrojados por otros modelos. Se utiliza la serie de datos sin variables explicativas, siguiendo el enfoque en [20], dadas las ventajas que supone esta estrategia frente a las dificultades en la selección y correcto uso de variables explicativas. Las medidas de desempeño observadas son MSE, MAD y MAPE. Los rezagos de la serie se establecieron con ayuda de las indicaciones metodológicas mostradas en el Capítulo 2. Aunque es posible que otros conjuntos de regresores sean más apropiados para este problema, se trabajó con los sugeridos más adelante con el propósito de evaluar la utilidad de los modelos MFIS en la tarea de modelamiento y, también, de evaluar la calidad de los resultados con respecto a otros modelos. En la revisión de las propiedades de uno de los modelos construidos se incluye el análisis de los grupos (clusters) que resultan en su construcción, con respecto a la serie. A pesar de que el modelo MFIS es de tipo *caja negra*, dado que no se tiene un significado explícito de los conjuntos borrosos encontrados, es interesante observar la relación que existe entre las características de la serie y los grupos de datos encontrados durante la construcción de los modelos.

Para establecer la capacidad de generalización de los modelos MFIS en este problema práctico, se comparó los resultados con los obtenidos a través de otros modelos, como redes neuronales autoregresivas y perceptrones multicapa. Dichos modelos se usaron según las directrices presentadas en [69] y se seleccionaron aquellos con menores errores de predicción.

En las siguientes secciones se exponen los modelos MFIS construidos, se analizan los resultados y se presentan las conclusiones sobre la aplicación de los modelos a la demanda de electricidad.

Tabla 5.1: Diferentes modelos para la demanda de energía eléctrica (tomado de [20]).

Región	Técnica/Modelo	Variables	Información
Arabia Saudita	Redes abductivas	Clima, demografía, indicadores económicos	Mensual
Delhi	Redes neuronales	Población, clima	1984-1993
Chile	Panel de datos		Mensual, 1995-2001
Israel	Regresión dinámica Cointegración		Trimestral
Turquía	Análisis de ciclos		Anual, 1988-1997
N Cyprus	Regresión múltiple	Consumidores, precio, turistas	Anual, 1988-1997
España	ARIMA		Mensual, 1980-1996
Estados Unidos	Funciones de transferencia	Clima, precio, ingresos	Mensual, 1969-1990
Grecia	Climáticas		Diario y mensual
Nueva Zelanda	Regresión lineal múltiple	GDP, ingreso, precio	Anual, 1965-1999
Líbano	Cointegración	PIB, total importaciones	1993-1997
Líbano	AR, ARIMA, híbrido		Mensual
Grecia	Funciones de transferencia	Ingresos, temperatura, precio	Mensual, trimestral
Colombia	Red neurodifusa Perceptrón multicapa Componentes principales	PIB, consumo ACPM, temperatura superficial del océano, tarifa de la energía, consumo de gas natural	Mensual, 1:1999-12:2004
Región Sur-Occidente de Colombia	Funciones Spline ARIMA Un modelo para cada hora del día	Hora, día, mes	Horario, 1:2001-11:2004
Demanda de una empresa Colombiana	ARIMA Un modelo para cada hora del día		Horario, 1:2001-12:2001
Colombia	ARIMA con intervenciones		Hora 12:00 1:1996-8:2002
Colombia	No especificados	PIB, tarifas de la energía, crecimiento de la población	Anual

5.2 Modelo MFIS construido a partir de los datos de demanda de electricidad

Para la construcción de los modelos se utilizaron los datos de demanda mensual de electricidad (en miles de GWh-mes) desde agosto de 1995 hasta enero de 2007 (138 observaciones). Los primeros 126 datos se utilizaron para entrenamiento y los últimos 12, correspondientes al período entre febrero de 2006 y enero 2007, se utilizaron para pronosticar un paso adelante. No se realizó transformación alguna sobre los datos antes de usar los modelos. En la Figura 5.1 se muestra la serie. Puede observarse un período de crecimiento inicial seguido de una caída y luego una tendencia de crecimiento sostenido, a partir de 2000.

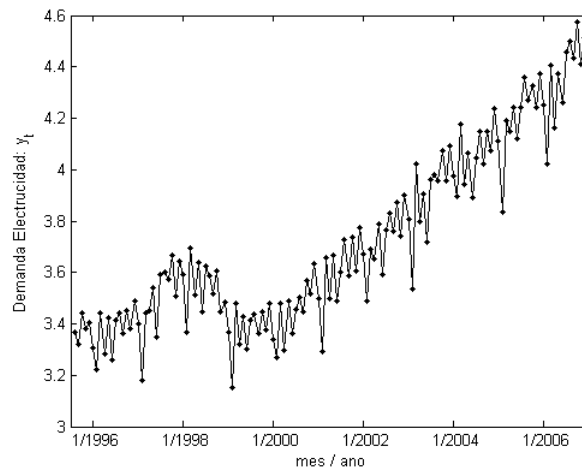


Figura 5.1: Demanda mensual de electricidad (miles de GWh-mes).

Los modelos obtenidos para esta serie de datos se construyeron según se indica en el Capítulo 2. Se utilizaron varias especificaciones, que se reportan en la Tabla 5.2. Para cada conjunto de rezagos tenidos en cuenta se entrenó el modelo con un número inicial de grupos $h_{min} = 1$ y también se realizaron experimentos con un número fijo de grupos para examinar posteriormente las regiones con base en las cuales el modelo construyó dichos grupos. La columna donde se muestra la configuración del número de grupos indica, para cada modelo, si este parámetro de configuración se dejó fijo o variable. Como se observa en la tabla, las medidas de desempeño usadas indican una alta capacidad de generalización. El MSE de pronóstico, en la mayoría de los casos, es próximo al MSE de entrenamiento lo cual puede ser un indicativo de la ausencia de *sobre ajuste* y de *falta de ajuste*.

De los diferentes modelos obtenidos, MFIS-7 y MFIS-4-2 superan las pruebas de

diagnóstico de manera más satisfactoria que los demás. Y de ellos, MFIS-7 es un buen candidato para modelamiento y pronóstico, dado que es superior en cuanto a las medidas de desempeño MAD, MAPE y R^2 tanto en la etapa de entrenamiento como en la etapa de validación. En la Figura 5.2 se muestra los valores reales de la serie (en color negro) y los valores pronosticados (en color azul). Se proporcionan también, con líneas continuas, las bandas de confianza del 95%. La Tabla 5.3 resume los parámetros del modelo, cuya estructura es la de la ecuación 2.5. Los centros se utilizan en el cálculo de los grados de pertenencia $\mu(x)$ a través de una función de distancia, D (véase la sección 2.2).

Tabla 5.2: Desempeño de modelos MFIS para la serie de demanda de electricidad.

Modelo	Rezagos	Config. Num. Grupos	Num. Grupos	Medidas de Entrenamiento	Medidas de Pronóstico 1 año adelante
MFIS-1	1,2	Variable	1	0.01401 (0.08969)(0.02474)	0.01799 (0.09677)(0.02237)
MFIS-1-2	1,2	Fijo	2	0.01387 (0.09135)(0.02498)	0.04633 (0.19873)(0.04499)
MFIS-2	1,2,3	Variable	1	0.01241 (0.08537)(0.02357)	0.01601 (0.0968)(0.02242)
MFIS-2-2	1,2,3	Fijo	2	0.01234 (0.086)(0.02361)	0.03756 (0,17751)(0,04025)
MFIS-3	1,2,6,7	Variable	1	0.0106 (0.07429)(0.02056)	0.01162 (0.07583)(0.01769)
MFIS-3-2	1,2,6,7	Fijo	2	0.01126 (0.07903)(0.02149)	0.06637 (0.23032)(0.05203)
MFIS-4	1,2,6,7,12,13	Variable	3	0.00337 (0.0431)(0.01179)	0.0038 (0.04519)(0.01037)
MFIS-4-2	1,2,6,7,12,13	Fijo	2	0.00421 (0.04834)(0.01317)	0.00685 (0.07457)(0.01696)
MFIS-5	1,6,7,12,13	Variable	1	0.00504 (0.05275)(0.01443)	0.00552 (0.05672)(0.01301)
MFIS-5-2	1,6,7,12,13	Fijo	2	0.00439 (0.04936)(0.0135)	0.00586 (0.06562)(0.01495)
MFIS-6	1,2,6,12,13	Variable	3	0.0035 (0.04326)(0.0118)	0.00369 (0.04608)(0.0105)
MFIS-6-2	1,2,6,12,13	Fijo	2	0.00427 (0.0476)(0.01299)	0.00536 (0.06409)(0.01461)
MFIS-7	1,2,6,12,13,24	Variable	2	0.003 (0.04379)(0.01182)	0.00515 (0.05928)(0.01374)

Tabla 5.3: Especificación del modelo MFIS-7.

Parámetro (en la regla)	Valor	Parámetro (en el centro)	Valor
Regla 1			
ϕ_0	-2.03462		
ϕ_1	-0.24203	h_1	3.74639
ϕ_2	-0.10001	h_2	3.75947
ϕ_3	0.34677	h_3	3.70257
ϕ_4	-0.25296	h_4	3.64062
ϕ_5	0.19378	h_5	3.62758
ϕ_6	1.74936	h_6	3.53875
Regla 2			
ϕ_0	2.46570		
ϕ_1	1.30619	h_1	3.45372
ϕ_2	0.21444	h_2	3.35598
ϕ_3	-0.46938	h_3	3.42273
ϕ_4	1.09627	h_4	3.48458
ϕ_5	-1.17358	h_5	3.49829
ϕ_6	-0.73779	h_6	3.49874

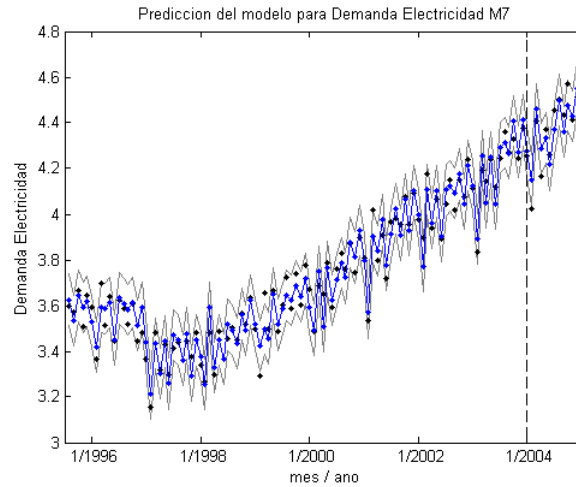
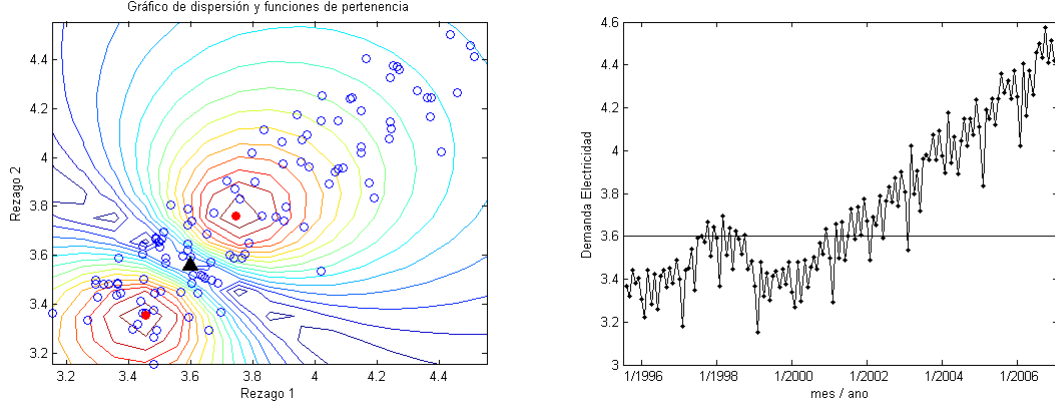


Figura 5.2: Predicción para la serie de datos de la demanda eléctrica mensual en Colombia. Modelo MFIS-7.

La gráfica de predicción y las medidas usadas en el metadiagnóstico (MAD,



(a) Gráfico de dispersión y funciones de pertenencia (restringidas a 2D). (b) Demanda de electricidad con indicación de partición (3.6 GWh-mes).

Figura 5.3: Agrupamiento de datos realizado por el modelo MFIS-7.

MAPE) muestran en general que el modelo sigue de cerca el comportamiento de la serie. En [20] se estudió esta serie de tiempo a través del análisis de componentes no observables y se encontró una tendencia subyacente que indica una caída en 1998 seguida de una fase de transición en el primer semestre de 1999 y finalmente un crecimiento sostenido desde el año 2000. El modelo MFIS capturó adecuadamente esta dinámica, según se observa en la Figura 5.2.

El modelo, como se expone en el Capítulo 2, está formado por varias reglas borrosas, cada una de las cuales se deriva de un conjunto borroso. En la construcción de dichos conjuntos se establece unas regiones importantes para el ajuste del modelo a los datos. En el caso del modelo MFIS-7 las regiones están definidas en un espacio de 6 dimensiones, correspondientes a los rezagos 1, 2, 6, 12, 13 y 24. Al examinar el modelo, se encuentra que se detectaron dos regiones. La Figura 5.3(a) ilustra, a través de curvas de nivel, las funciones de pertenencia de los conjuntos borrosos asociados a las regiones (se restringe la visualización a dos dimensiones para facilitar la representación gráfica). Se incluyen los datos, con círculos en color azul, y los centros de los conjuntos, en color rojo. Las dos superficies que representan las funciones de pertenencia se interceptan en la recta $r_2 = -0.7254r_1 + 6.1692$. El triángulo en color negro indica el punto de dicha intersección que está más cercano a los centros.

El nivel de demanda correspondiente al punto de intersección representado con el triángulo en la Figura 5.3(a) se adicionó a la serie de datos en la Figura 5.3(b). Allí puede observarse que dicho nivel de demanda indica que en el lapso de tiempo entre 1995 y el segundo semestre de 2001 existe una dinámica con diferencias significativas a la del período que abarca desde 2002 en adelante. Al revisar el análisis de las tendencias en [20], se encuentra que es bastante específico, mientras que los modelos locales usados en MFIS realizan una distinción más amplia, al parecer entre un período de oscilación y un período de tendencia.

Es importante recalcar que las regiones encontradas durante el entrenamiento del modelo MFIS son puntos de partida para la construcción de conjuntos borrosos. Una vez el modelo dispone de estos conjuntos, los límites entre ellos son difusos y existen cambios de nivel suaves entre uno y otro. Las regiones, y los conjuntos basados en ellas, se determinan por el algoritmo con base en el mejor ajuste del modelo a los datos y no en todos los casos es factible realizar una interpretación de dichas particiones en términos del fenómeno que se está modelando.

5.3 Contraste de resultados con otros modelos no lineales

La comparación de los resultados obtenidos con los modelos MFIS en la predicción de la demanda de electricidad, con respecto a otros modelos, dará una imagen de su capacidad descriptiva y de generalización. Para tal propósito se utilizaron los modelos MLP (*multilayer perceptron*) y ARNN (*autoregressive neural network*) reportados en [69]. De los diversos modelos MLP y ARNN obtenidos se seleccionaron aquellos con mejor desempeño teniendo en cuenta sólo el MSE obtenido en el pronóstico de la demanda un año hacia adelante. En la Tabla 5.4 se resumen los resultados, incluyendo los obtenidos con los modelos MFIS.

Al comparar los resultados de la Tabla 5.4 se encuentra que el MSE de entrenamiento de los modelos MFIS-1, MFIS-2, MFIS-3 y MFIS-5 tuvo desmejoras con respecto a los otros modelos, al ser más alto entre un 21% y un 37%. Los modelos MFIS-4 y MFIS-6 y MFIS-7 tuvieron mejoras al presentar disminuciones del MSE entre un 16% y un 32%. En cuanto a la medida de desempeño para el pronóstico 1 año hacia adelante, las diferencias son más amplias. MFIS-4 tuvo una ligera mejora del 2% sobre el modelo ARNN-4 mientras que desmejoró un 2% con respecto al MLP-4. Los modelos MLP-3 y ARNN-3 se mostraron superiores al presentar disminuciones del 56% y 98%, respectivamente, con respecto al modelo MFIS-3 en la configuración de rezagos 1,2,6,7. A su vez, los modelos MLP-6 y ARNN-6 presentaron disminuciones del 105% y 50% con respecto a MFIS-6 en la configuración con rezagos 1,2,6,12,13. Para las demás configuraciones de rezagos, MFIS tuvo también desempeños inferiores, pero con diferencias ligeras. Es importante notar que los modelos ARNN y MLP muestran, en la mayoría de los casos estructuras complejas, con 5 capas ocultas o más, mientras que los modelos MFIS no sobrepasaron las tres reglas borrosas, que refleja una estructura simple.

La capacidad de generalización (que también puede observarse en la Figura 5.2) se muestra cercana en muchos casos a la mostrada por los modelos MLP y ARNN, aunque hay algunos casos de desempeño muy superior en estos últimos modelos. Posiblemente las estructuras más complejas de los mismos permitan un mejor ajuste a la dinámica de la serie.

Tabla 5.4: Desempeño de varios modelos para la serie de demanda de electricidad.

Modelo	Rezagos	Medidas de Entrenamiento			Medidas de Pronóstico		
		MSE	(MAD)	(MAPE)	1 año adelante	MSE	(MAD)
MLP-1(H=9)	1,2	0.01155	(0.08301)	(0.02278)	0.01521	(0.09792)	(0.02267)
ARNN-1(H=3)	1,2	0.01187	(0.08403)	(0.02309)	0.01593	(0.09142)	(0.02116)
MFIS-1	1,2	0.01401	(0.08969)	(0.02474)	0.01799	(0.09677)	(0.02237)
MLP-2(H=9)	1,2,3	0.01041	(0.08025)	(0.02206)	0.01412	(0.08675)	(0.02016)
ARNN-2(H=5)	1,2,3	0.01110	(0.08131)	(0.02243)	0.01445	(0.08527)	(0.01981)
MFIS-2	1,2,3	0.01241	(0.08537)	(0.02357)	0.01601	(0.0968)	(0.02242)
MLP-3(H=7)	1,2,6,7	0.00790	(0.06390)	(0.01758)	0.00743	(0.07246)	(0.01680)
ARNN-3(H=5)	1,2,6,7	0.00773	(0.06424)	(0.01762)	0.00586	(0.05919)	(0.01382)
MFIS-3	1,2,6,7	0.0106	(0.07429)	(0.02056)	0.01162	(0.07583)	(0.01769)
MLP-4(H=5)	1,2,6,7,12,13	0.00483	(0.04993)	(0.01365)	0.00372	(0.05114)	(0.01176)
ARNN-4(H=1)	1,2,6,7,12,13	0.00482	(0.04972)	(0.01360)	0.00388	(0.05197)	(0.01195)
MFIS-4	1,2,6,7,12,13	0.00337	(0.0431)	(0.01179)	0.0038	(0.04519)	(0.01037)
MLP-5(H=10)	1,6,7,12,13	0.00407	(0.04675)	(0.01277)	0.00302	(0.03957)	(0.00913)
ARNN-5(H=8)	1,6,7,12,13	0.00465	(0.05022)	(0.01379)	0.00411	(0.04657)	(0.01074)
MFIS-5	1,6,7,12,13	0.00504	(0.05275)	(0.01443)	0.00552	(0.05672)	(0.01301)
MLP-6(H=10)	1,2,6,12,13	0.00415	(0.04568)	(0.01252)	0.00180	(0.03793)	(0.00870)
ARNN-6(H=5)	1,2,6,12,13	0.00439	(0.04799)	(0.01318)	0.00246	(0.03983)	(0.00909)
MFIS-6	1,2,6,12,13	0.0035	(0.04326)	(0.0118)	0.00369	(0.04608)	(0.0105)
MLP-7(H=6)	1,2,6,12,13,24	0.00441	(0.04722)	(0.01273)	0.00374	(0.05234)	(0.01208)
ARNN-7(H=2)	1,2,6,12,13,24	0.00428	(0.04813)	(0.01298)	0.00369	(0.05130)	(0.01186)
MFIS-7	1,2,6,12,13,24	0.003	(0.04379)	(0.01182)	0.00515	(0.05928)	(0.01374)

5.4 Conclusiones

En este capítulo se ajustaron varios modelos de inferencia borrosa multidimensional (MFIS) a la serie de demanda eléctrica en Colombia. Se estudiaron las propiedades de uno de ellos y se estableció algunas indicaciones de interpretabilidad del particionamiento conseguido por el modelo, en términos del problema tratado. También se comparó los resultados con modelos MLP y ARNN.

Se encontró que es factible realizar interpretaciones de las regiones del espacio de entrada que se construyen en la especificación del modelo MFIS, en términos del fenómeno bajo estudio. Ellas podrían dar un indicativo de dinámicas marcadamente distintas en la serie y apoyar un juicio experto en el análisis del fenómeno. Sin embargo, no siempre es posible contar con una interpretación plausible de dichas regiones ni visualizarlas adecuadamente.

Las interacciones entre las variables involucradas en la descripción de un fenómeno como el de la demanda eléctrica pueden exhibir comportamientos definidos en regiones que difícilmente pueden visualizarse cuando la cantidad de variables es alta. En estos casos las regiones n-dimensionales construidas le dan al modelo la capacidad para detectar estos comportamientos.

Al comparar el desempeño de los modelos MFIS con los modelos MLP y ARNN se encuentra un desempeño similar, aunque en la mayoría de los casos con un MSE de pronóstico mayor, lo cual indica la superioridad de los otros modelos en la capacidad de generalización. No obstante, la magnitud de las medidas es un indicativo de un muy buen ajuste a la dinámica de la serie tanto en el período de entrenamiento como en el período de validación y de una promisorio capacidad de generalización para los modelos MFIS.

Capítulo 6

Conclusiones y trabajo futuro

En esta tesis se formularon varios problemas concernientes a los sistemas de inferencia borrosa multidimensionales, mencionados a continuación:

1. La utilización del agrupamiento se ha visto como un paso de las metodologías, cuyo resultado se afina una vez se han construido las reglas del sistema. No se ha explorado la construcción de los grupos de manera que la inclusión de uno u otro este dictada por su aporte en la descripción de la dinámica de la serie. Más aún, se desconoce cuál metodología de agrupamiento sería más adecuada para las series temporales.
2. En los algoritmos de agrupamiento que se usan comúnmente hay dependencia de los puntos iniciales escogidos. Esto ocasiona que haya diversas estructuras posibles del modelo para diversos puntos de partida.
3. Los criterios para determinar el número apropiado de grupos (y de reglas asociadas) utilizan procedimientos de validación de grupos o procedimientos de particionamiento sobre los datos. No se ha observado que el número adecuado de grupos se asocie a la capacidad del sistema para describir adecuadamente la dinámica de la serie.
4. Las metodologías no están vinculadas al conocimiento propio de la Econometría y el modelamiento estadístico de series temporales.
5. El uso de dichos sistemas para el modelado y la predicción de series temporales no lineales es muy escaso y existe poca experiencia con series temporales de tipo económico.

Para avanzar en la solución de estos problemas se plantearon los siguientes objetivos:

1. Proponer un mecanismo de particionamiento o agrupamiento que incluya o descarte grupos con base en el aporte del modelo en la descripción de la dinámica de la serie.

2. Proponer un mecanismo de particionamiento o agrupamiento que mitigue la dependencia de puntos iniciales.
3. Proponer un criterio para establecer la complejidad del modelo con base en el desempeño global del sistema y que sea independiente de criterios de validez del algoritmo de agrupamiento.
4. Formular una metodología para el modelamiento y predicción de series temporales no lineales, que haga uso de los sistemas de inferencia borrosa en cuestión. Dicha metodología deberá estar basada en el modelamiento estadístico de series temporales.
5. Utilizar el modelo y la metodología obtenidos para analizar series benchmark, que permitan comparar el desempeño del modelo.
6. Utilizar el modelo y la metodología obtenidos para analizar una o más series reales, de tipo económico.

A continuación se menciona la forma en que se contribuyó al cumplimiento de los objetivos trazados.

1. Propuesta de un mecanismo de particionamiento que incluya o descarte grupos con base en el aporte del modelo en la descripción de la dinámica de la serie.

El propósito del agrupamiento es, generalmente, encontrar unos conjuntos borrosos con los cuales se crean prototipos de las reglas borrosas y el sistema se afina en una etapa posterior. Si el origen de los grupos es un algoritmo de agrupamiento, pueden obtenerse diversas configuraciones si los puntos de partida del algoritmo se varían. Por tanto, se buscó un origen de los grupos diferente, que permitiera encontrar configuraciones de grupos con base en el aporte a la descripción de la dinámica de la serie. Dado que los algoritmos de particionamiento en los que se basa MARS (*Multivariate Adaptive Regression Splines*) tienen la propiedad de encontrar regiones del espacio de entrada a las que se asocian modelos de regresión simples con base en la descripción de la función modelada y que dicha partición no depende de puntos aleatorios de inicio, se investigó estas técnicas. En dicho proceso de investigación se encontró que los modelos de particionamiento recursivo de MARS y los sistemas borrosos multidimensionales (con un grupo por cada regla) tienen similitudes interesantes en su estructura, lo cual brinda posibilidades de exploración conjunta de ambos tópicos: agrupamiento y particionamiento. Se adaptó entonces dicho esquema de particionamiento para la construcción de los sistemas MFIS de tal manera que las regiones del espacio de entrada que el algoritmo produce sirvieran como base para la construcción de los conjuntos borrosos del sistema de inferencia. Todo esto se integró de tal manera que una configuración de regiones con una partición adicional (y los grupos borrosos asociados) es admitida

sólo si la dinámica de la serie es descrita de una manera más precisa que con la configuración de regiones previa. A su vez las variables escogidas para realizar las particiones se establecen teniendo en cuenta la capacidad de descripción de la dinámica de la serie. No obstante los resultados alentadores obtenidos, puede experimentarse con algoritmos de agrupamiento existentes para encontrar posibles resultados favorables en un esquema más complejo, basado en los modelos MFIS ya obtenidos. Igualmente es necesario realizar experimentos en los cuales se hagan optimizaciones combinatorias con las variables incluidas en los consecuentes de las reglas, dado que el enfoque incremental simple que se adoptó en el presente trabajo no arrojó resultados favorables.

2. Propuesta de un mecanismo de particionamiento o agrupamiento que mitigue la dependencia de puntos iniciales.

Como ya se mencionó, una de las propiedades del mecanismo de particionamiento adoptado es que no se requiere partir de puntos aleatorios, como en otros algoritmos usados para la obtención de los grupos. De esta manera las particiones, y los grupos borrosos derivados de ellas, tienen una manera reglada de generación y, para una configuración dada, se obtendrá el mismo sistema borroso en ejecuciones sucesivas. Esta situación no siempre se presenta cuando en la construcción de los sistemas borrosos se empieza con un algoritmo de agrupamiento basado en puntos de partida aleatorios, o dictados por el modelador.

3. Propuesta de un criterio para establecer la complejidad del modelo con base en el desempeño global del sistema y que sea independiente de criterios de validez del algoritmo de agrupamiento.

Para lograr este objetivo se incluyó en la metodología y en los algoritmos de entrenamiento un criterio de parada que fue propuesto en [23]. En dicho trabajo este criterio ayuda a determinar el número adecuado de capas en la red neuronal, mientras que en el modelo MFIS propuesto aquí, ayuda a determinar el número adecuado de regiones en que debe particionarse el espacio de las variables de entrada, que será igual al número de grupos borrosos para el sistema de inferencia. Al no estar basado en el resultado de un algoritmo de agrupamiento, este método evita el uso de criterios de validación de grupos u otros criterios basados en la correctitud del agrupamiento.

4. Formulación de una metodología para el modelamiento y predicción de series temporales no lineales, que haga uso de los sistemas de inferencia borrosa en cuestión.

Para el alcance de este objetivo se tomó los aportes metodológicos de [64], [66] y [68] y se los adaptó al caso de los modelos MFIS propuestos. La metodología formulada adicionó fundamentos de la estadística al proceso de construcción de los modelos y permitió guiar de manera más formal la construcción de los mismos. Con las herramientas desarrolladas, el modelador podrá concentrarse en

la especificación de las entradas del sistema y en el ajuste de los parámetros de tolerancia al error, dado que el número de reglas se establece automáticamente con un criterio que tiene en cuenta un equilibrio entre el sobre-entrenamiento y la falta de entrenamiento. Dado que los modelos construidos tienen una estructura que puede diferir de otras propuestas, la metodología deberá adaptarse si se desea aplicar con otros esquemas de sistema de inferencia. Además de la parte estructural del modelo, la parte dinámica (algorítmica) también influye en aplicabilidad de la metodología. Por ejemplo, si los algoritmos usados para identificar un sistema de inferencia no incluyeran un criterio de parada, que equilibrara el sobre-ajuste y la falta de ajuste, deberá modificarse la metodología para indicar al modelador que debe examinar la complejidad del modelo.

5. Ajuste del modelo con series *benchmark*, que permitan comparar el desempeño del mismo.

Se obtuvo resultados alentadores con series *benchmark* como *Airline Passenger*, *Sunspot* y *Pollution Equipment*, utilizadas con frecuencia para evaluar el desempeño de modelos no lineales de series temporales. Se muestra con dichos resultados una muy buena capacidad de generalización en los modelos construidos, sobrepasando las medidas de desempeño de la mayoría de modelos encontrados en los trabajos mencionados (que incluyen modelos no lineales como DAN2). Después de contar con evidencia de resultados comparables o superiores a los de otros modelos no lineales actuales, puede explorarse en campos prácticos las condiciones y tipos de problemas para los cuales resultan más acertados los modelos propuestos en esta tesis.

6. Utilización del modelo y la metodología obtenidos para analizar una o más series reales, de tipo económico.

Como caso real de aplicación se escogió el problema de modelado de la demanda eléctrica en Colombia. Durante el modelamiento de esta serie se observó la oportunidad de interpretar, aunque con limitaciones, la configuración de particiones encontradas por los algoritmos implementados y apoyar un juicio experto en el análisis del fenómeno bajo estudio. La interpretabilidad es un tópico aún por explorar, aunque se observa que es más factible realizarla en condiciones donde las particiones se llevan a cabo en pocas dimensiones de los vectores de entrada. Por otro lado, se observa un desempeño similar de los modelos MFIS, con respecto a modelos como los perceptrones multicapa (MLP) y las redes neuronales autorregresivas (ARNN), aunque en la mayoría de los casos con un MSE de pronóstico mayor, lo cual indica la superioridad de los otros modelos en la capacidad de generalización. No obstante, la magnitud de las medidas de desempeño es un indicativo de un muy buen ajuste a la dinámica de la serie tanto en el período de entrenamiento como en el período de validación y de una promisorio capacidad de generalización para los modelos MFIS.

Al igual que con las series *benchmark*, es conveniente realizar más experimentación con los modelos propuestos aquí, de tal manera que se los pueda usar con mejor provecho después de ganar conocimiento en las áreas y condiciones con que resultan ser más productivos.

Bibliografía

- [1] T. Al-Saba y I. El-Amin. Artificial neural networks as applied to long term demand forecasting. *Artificial Intelligence in Engineering*, 13:189–197, 1999.
- [2] H. Alvarez y M. Peña. Modelamiento de sistemas de inferencia borrosa. *Avances en Sistemas e Informática*, 1:1–11, 2004.
- [3] U. Anders y O. Korn. Model selection in neural networks. *Neural Networks*, 12:309–323, 1999.
- [4] P. Angelov y R. Buswell. Automatic generation of fuzzy rule-based models from data by genetic algorithms. *Information Sciences*, 150:17–31, 2003.
- [5] J. Armstrong. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer, 2001.
- [6] A. F. Barrientos, J. Olaya, y V. González. Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 30:187–202, 2007.
- [7] G. E. P. Box y G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day Inc., 1970.
- [8] Britannica. Sunspot. *Electronic Encyclopædia*, 2008.
- [9] R. Brown, J. Durbin, y J. Evans. Techniques for testing the consistency of regression relationships over time. *Journal of the Royal Statistical Society, Series B*, 37:149–192, 1975.
- [10] D. Bunn y G. Wright. Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science*, 37, 1991.
- [11] J. Casillas, O. Cordon, M. D. Jesus, y F. Herrera. Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Trans. Fuzzy Systems*, 13:13–29, 2005.

- [12] E. Castaño. Reconstrucción de datos de series de tiempo: una aplicación a la demanda horaria de la electricidad. *Revista Colombiana de Estadística*, 30:247–263, 2007.
- [13] O. Castillo, A. Alanis, M. Garcia, y H. Arias. An intuitionistic fuzzy system for time series analysis in plant monitoring and diagnosis. *Applied Soft Computing*, 7:1227–1233, 2007.
- [14] O. Castillo y P. Melin. Simulation and forecasting complex financial time series using neural networks and fuzzy logic. *IEEE International Conference on Systems, Man, and Cybernetics*, 4:2664–2669, 2001.
- [15] P. Dash, A. Liew, S. Rahman, y S. Dash. Computing models for electric load forecasting engineering. *Applications of Artificial Intelligence*, 8:423–433, 1995.
- [16] M. Delgado, A. Gómez-Skarmeta, y F. Martín. A methodology to model fuzzy systems using fuzzy clustering in a rapid-prototyping approach. *Fuzzy Sets and Systems*, 97:287–301, 1998.
- [17] J. Faraway y C. Chatfield. Time series forecasting with neural networks: A comparative study using the airline data. *Applied Statistics*, 47(231–250), 1998.
- [18] A. Fiordaliso. A constrained takagi-sugeno fuzzy system that allows for better interpretation and analysis. *Fuzzy Sets and Systems*, 118:307–318, 2001.
- [19] G. S. Fishman. *Monte Carlo Concepts, Algorithms and Applications*. Springer Series in Operational Research. Springer, 1996.
- [20] C. J. Franco, J. D. Velásquez, y Y. Olaya. Caracterización de la demanda mensual de electricidad en Colombia usando un modelo de componentes no observables. *Cuad. Adm. Bogotá (Colombia)*, 21:221–235, 2008.
- [21] J. H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19:1–67, 1991.
- [22] Y. Gao y M. Er. NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches. *Fuzzy Sets and Systems*, 150:331–350, 2005.
- [23] M. Ghiassi y H. Saidane. A dynamic architecture for artificial neural networks. *Neurocomputing*, 63:397–413, 2005.
- [24] M. Ghiassi, H. Saidane, y D. K. Zimbra. A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting*, 21:341–362, 2005.

- [25] A. F. Gómez-Skarmeta, M. Delgado, y F. Martín. A fuzzy clustering-based rapid prototyping for fuzzy rule-based modeling. *IEEE Transactions on Fuzzy Systems*, 5:223–233, 1997.
- [26] A. F. Gómez-Skarmeta, M. Delgado, y M. A. Vila. About the use of fuzzy clustering techniques for fuzzy model identification. *Fuzzy sets and systems*, 106:179–188, 1999.
- [27] C. Granger y T. Teräsvirta. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press, 1993.
- [28] D. Graves y W. Pedrycz. Fuzzy prediction architecture using recurrent neural networks. *Neurocomputing*, 2008. Artículo en imprenta.
- [29] A. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press., 1989.
- [30] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall, Inc., 1999.
- [31] L. Herrera, H. Pomares, I. Rojas, A. Guillén, J. González, M. Awad, y A. Herrera. Multigrid-based fuzzy systems for time series prediction: CATS competition. *Neurocomputing*, 70(13-15):2410–2425, 2007.
- [32] R. Hogarth. *Judgement and Choice*. John Wiley and Sons, second edition, 1994.
- [33] A. Homaifar y E. McCormick. Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Transactions on Fuzzy Systems*, 3(2):129–139, 1995.
- [34] J. S. Jang. Adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–685, 1993.
- [35] J. S. Jang y E. Mizutani. *Neuro-Fuzzy and Soft Computing*. Prentice Hall, New York, 1997.
- [36] C. Jarque y A. Bera. A test for normality of observations and regression residuals. *International Statistical Review*, 55:163–172, 1987.
- [37] H. Jorquera, R. Pérez, A. Cipriano, A. Espejo, M. V. Letelier, y G. Acuña. Forecasting ozone daily maximum levels at Santiago, Chile. *Atmospheric Environment*, 32(20):3415–3424, 1998.
- [38] C.-F. Juang, I.-F. Chung, y C.-H. Hsu. Automatic construction of feedforward/recurrent fuzzy systems by clustering-aided simplex particle swarm optimization. *Fuzzy Sets and Systems*, 158(18):1979–1996, 2007.

- [39] N. N. Karnik y J. M. Mendel. Applications of type-2 fuzzy logic systems to forecasting of time-series. *Information Sciences*, 120(1-4):89–111, 1999.
- [40] N. Kasabov. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. The MIT Press, 1996.
- [41] D. Kim y C. Kim. Forecasting time series with genetic fuzzy predictor ensemble. *IEEE Transactions on Fuzzy Systems*, 5(4):523–535, Nov 1997.
- [42] J. Kim y N. Kasabov. Hyfis: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. *Neural Networks*, 12(9):1301–1319, 1999.
- [43] G. Koop, M. H. Pesaran, y S. M. Potter. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147, 1996.
- [44] A. Kroll. Identification of functional fuzzy models using multidimensional reference fuzzy sets. *Fuzzy Sets and Systems*, 80(2):149–158, 1996.
- [45] R. J. Kuo y K. C. Xue. An intelligent sales forecasting system through integration of artificial neural network and fuzzy neural network. *Computers in Industry*, 37(1):1–15, 1998.
- [46] R. Lamedica, A. Prudenzi, M. Sforza, M. Caciotta, y V. Cencelli. A neural network based technique for short-term forecasting of anomalous load periods. *IEEE Transactions on Power Systems*, 11(4):1749–1756, Nov 1996.
- [47] T. W. Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857–1874, 2005.
- [48] S. G. Makridakis, S. C. Wheelwright, y R. J. Hyndman. *Forecasting: Methods and applications*. John Wiley and Sons, New York, 3rd edition, 1998.
- [49] T. Masters. *Neural, Novel and Hybrid Algorithms for Time Series Prediction*. John Wiley and Sons, Inc., 1995.
- [50] R. McCleary y R. Hay. *Applied time series analysis for the social sciences*. SAGE Publications, Inc., 1980.
- [51] T. C. Mills. *The econometric modeling of financial time series*. Cambridge University Press, 1993.
- [52] P. C. Nayak, K. P. Sudheer, D. M. Rangan, y K. S. Ramasastri. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, 291(1-2):52–66, 2004.

- [53] J. Nie. Nonlinear time-series forecasting: A fuzzy-neural approach. *Neurocomputing*, 16(1):63–76, 1997.
- [54] S.-K. Oh, W. Pedrycz, y T.-C. Ahn. Self-organizing neural networks with fuzzy polynomial neurons. *Applied Soft Computing*, 2(1):1–10, 2002.
- [55] E. Page. Continuous inspection schemes. *Biometrika*, 41:100–114, 1954.
- [56] P.-F. Pai. Hybrid ellipsoidal fuzzy systems in forecasting regional electricity loads. *Energy Conversion and Management*, 47(15-16):2283–2289, 2006.
- [57] C. Potter y M. Negnevitsky. ANFIS application to competition on artificial time (CATS). In *Proceedings of 2004 IEEE International Conference on Fuzzy Systems*, pages 469–474, 2004.
- [58] H.-J. Rong, N. Sundararajan, G.-B. Huang, y P. Saratchandran. Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157(9):1260–1275, 2006. Fuzzy Concepts Applied to Food Control Quality Control.
- [59] M. Russo. Genetic fuzzy learning. *Evolutionary Computation, IEEE Transactions on*, 4(3):259–273, Sep 2000.
- [60] A. Sfetsos y A. H. Coonick. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2):169–178, 2000.
- [61] F. Steiner. Regulation, industry structure and performance in electricity supply industry. *OECD Economic Studies*, 32, 2000.
- [62] M. Sugeno y T. Yasukawa. A fuzzy-logic-based approach to qualitative modeling. *IEEE Transactions on Fuzzy Systems*, 1(1):7–31, Feb 1993.
- [63] Y. Tang y Y. Xu. Application of fuzzy naive Bayes and a real-valued genetic algorithm in identification of fuzzy model. *Information Sciences*, 169(3-4):205–226, 2005.
- [64] T. Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218, 1994.
- [65] H. Theil. *Applied Economic Forecasting*. Rand McNally, 1966.
- [66] H. Tong. *Non-linear Time Series, a dynamical system approach*. Oxford Statistical Science Series. Clarendon Press Oxford, 1990.
- [67] R. S. Tsay. *Analysis of Financial Time Series*. John Wiley and Sons, 2002.

- [68] J. D. Velásquez, I. Dyner, y R. C. Souza. Modelación de series temporales usando ANFIS. *Revista Iberoamericana de Inteligencia Artificial*, 20(34–52), 2004.
- [69] J. D. Velásquez, C. J. Franco, y H. A. García. Un modelo no lineal para la predicción de la demanda mensual de electricidad en Colombia. No publicado, 2008.
- [70] L.-X. Wang y J. Mendel. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *Neural Networks, IEEE Transactions on*, 3(5):807–814, Sep 1992.
- [71] H. Yin, S. C. Wong, J. Xu, y C. K. Wong. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies*, 10(2):85 – 98, 2002.
- [72] G. P. Zhang. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50:159 – 175, 2003.
- [73] M. Zounemat-Kermani y M. Teshnehlab. Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Applied Soft Computing*, 8(2):928 – 936, 2008.

Anexos

A. Resumen ejecutivo

Modelado de series temporales con sistemas de inferencia borrosa multidimensionales

Por: Juan Fernando Rendón Sánchez

Director de tesis: Juan David Velásquez Henao, M.Sc., Ph.D.

Tesis enviada a la Escuela de Sistemas como requerimiento parcial para el grado de Magister en Ingeniería - Ingeniería de Sistemas.

Facultad de Minas. Universidad Nacional de Colombia, Sede Medellín.

Junio 2009

Palabras clave: *series temporales, agrupamiento, multidimensional, inferencia borrosa.*

La predicción de series temporales no lineales ha sido un problema vigente durante las últimas dos décadas. Existen diversos modelos para su análisis, que se clasifican en lineales y no lineales. Dentro de los no lineales se tienen algunos no paramétricos como las redes neuronales y los sistemas de inferencia borrosa, que han obtenido una atención considerable dada su capacidad para reproducir dinámicas no lineales.

En el campo específico de los sistemas de inferencia borrosa hay muchas ramificaciones en cuanto a su uso para el análisis de series temporales. Entre los enfoques existentes se encuentra el uso de conjuntos borrosos multidimensionales para la identificación del modelo. Esta estrategia mitiga algunos problemas de desempeño presentados por los sistemas de inferencia borrosa que realizan particiones en cada dimensión del espacio de entrada. Sin embargo, la utilidad de este enfoque particular no se ha explorado con profundidad en el campo de las series temporales. A su vez, existen aspectos tanto metodológicos como de diseño del sistema que dan lugar a interrogantes por resolver.

De la problemática encontrada con respecto a estos sistemas, se pretende avanzar en esta investigación con respecto a algunos tópicos, como son: la tendencia a considerar la construcción de los grupos borrosos como una etapa inicial de la construcción del modelo y no como un paso ligado desde el comienzo a la explicación de la dinámica de la serie; la dependencia de puntos aleatorios de partida en los algoritmos de agrupamiento usados; la carencia de prácticas metodológicas que guíen un proceso de construcción de modelos de una manera más objetiva y basada en herramientas estadísticas y, finalmente, la carencia de experiencia en el uso de estos modelos con series de tipo económico.

La investigación permitió encontrar un modelo y un procedimiento de identificación que mitigan las falencias señaladas. El modelo está basado en características de la herramienta de regresión MARS, de la cual se tomaron aspectos de suma importancia como la capacidad de optimizar simultáneamente los parámetros y la estructura

del modelo y la construcción de particiones óptimas en el espacio de entrada. Se adoptaron herramientas metodológicas basadas en estadística que mejoraron el proceso de construcción y evaluación de modelos.

La capacidad de generalización de los modelos MFIS (*multidimensional fuzzy inference system*) construidos se comparó con la de otros modelos, principalmente DAN2 (*dynamic architecture for artificial neural networks*). Se encontraron resultados superiores con respecto a la mayoría de los modelos revisados, usando medidas de desempeño como SSE (*error sum of squares*), MSE (*mean square due to error*) y MAD (*mean absolute deviation*). Por otro lado, la aplicabilidad en el campo de las series de tipo económico se probó con el modelamiento de la demanda mensual de electricidad en Colombia.

Los hallazgos del trabajo incluyen un indicio claro de que estos sistemas de inferencia son promisorios en el modelado y predicción de series temporales, cuando se potencian con herramientas metodológicas que guíen su uso. Adicionalmente, se eliminó la dependencia de puntos aleatorios de partida y se formuló un mecanismo de construcción de grupos asociado a la descripción de la dinámica de la serie. Por otro lado, durante la exploración de MARS como modelo de regresión se encontraron similitudes estructurales y dinámicas con respecto a los modelos MFIS, lo cual brinda posibilidades de exploración conjunta de los tópicos de particionamiento y agrupamiento. Con la aplicación del modelo MFIS a la serie de datos de demanda de electricidad se encontró que existe la posibilidad realizar interpretaciones de los conjuntos o de las regiones con base en las cuales se construyen los mismos. Dicha interpretación puede ayudar a contrastar o apoyar un juicio experto acerca del fenómeno bajo estudio. Sin embargo este es aún un tópico por explorar.

B. Executive summary

Time series modelling with multidimensional fuzzy inference systems

By: Juan Fernando Rendón Sánchez

Thesis director: Juan David Velásquez Henao, M.Sc., Ph.D.

A thesis submitted to *Escuela de Sistemas* in partial fulfillment of the requirements for the degree of Master of Science.

Facultad de Minas. Universidad Nacional de Colombia, Sede Medellín.

June 2009

Key words: *time series, clustering, multidimensional, fuzzy inference.*

Non-linear time series forecasting has been an active research area during the last two decades. A variety of time series models exists, that can be classified into lineal and non-linear models. Among the non-linear ones there exist non-parametric options such as neural networks and fuzzy inference systems, which are highly valued due to their powerful capacity to reproduce non-linear dynamics.

Research in the area of fuzzy inference systems has lead to the existence of numerous ramifications regarding their use for time series analysis. One of those approaches focuses on multidimensional fuzzy sets to aid the model identification process. The main advantage of such approach is the improvement of performance with respect to fuzzy inference systems that use partitions in every input variable domain. However, the extent to which those inference systems can be used for time series analysis has not been explored in detail. In addition, there are topics regarding methodology and design that point to unanswered questions.

The topics addressed in this work include: the tendency observed to consider data clustering as a separate first step for model identification and not as a step strongly tied to the correct reproduction of the series dynamics, the dependence found in clustering algorithms on random starting points, the lack of methodological guidelines for model building and evaluation and, finally, the absence of applications of these models to real economic time series.

The research conducted produced a model and an identification procedure, that tend to solve the problems mentioned. The model is based on some characteristics of MARS (a regression tool). MARS provided important bases, given its recognized capability for simultaneous structure and parameter optimization. Its capacity to find optimal input space partitioning was also of high importance. A set of methodological guidelines, based on statistics, were adopted in order to improve the model building and evaluation process.

The generalization capabilities of the proposed MFIS (multidimensional fuzzy inference system) models were assessed comparing their performance against the results

reported in known works with models such as DAN2 (dynamic architecture for artificial neural networks). MFIS models showed to be superior in most cases, according to measures like SSE (error sum of squares), MSE (mean square due to error) y MAD (mean absolute deviation). On the other hand, the applicability of MFIS to economic time series was assessed using data of monthly electricity demand in Colombia.

The findings of this research include the evidence that support the applicability of multidimensional fuzzy inference systems for time series modelling and forecasting, when they are used with appropriate methodological guidelines. In addition, the dependence on random starting points was eliminated and a model building routine was proposed that constructs clusters taking into account the description of the series dynamics. On the other hand, while MARS algorithms were explored, some similarities, both dynamical and structural, were detected between them and MFIS models. Such similarities give rise to the possibility to explore partitioning and clustering in conjunction. The application of MFIS models to monthly electricity demand in Colombia showed that a possibility exists of interpreting the resulting clusters, or the regions used to produce them. That interpretation could be used to support or contrast expert knowledge related to the phenomenon under study. However, interpretation is a topic that must be further explored.