



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
(INGENIERÍA CIVIL) – (HIDRÁULICA)

PREDICCIÓN MEDIANTE ANÁLISIS NO LINEAL EN HIDROLOGÍA

TESIS
QUE PARA OPTAR POR EL GRADO DE:
DOCTOR EN INGENIERÍA

PRESENTA:
LEONARDO DANIEL AMORES ROVELO

TUTOR PRINCIPAL
DR. CARLOS A. ESCALANTE SANDOVAL, FACULTAD DE INGENIERÍA
COMITÉ TUTOR
DR. JESÚS GRACIA SÁNCHEZ, INSTITUTO DE INGENIERÍA
DR. RODOLFO SILVA CASARÍN, INSTITUTO DE INGENIERÍA
DRA. LILIA REYES CHÁVEZ, FACULTAD DE INGENIERÍA
DR. EDGAR GERARDO MENDOZA BALDWIN, INSTITUTO DE INGENIERÍA

MÉXICO, D. F. MAYO 2015

JURADO ASIGNADO:

Presidente: DR. JESÚS GRACIA SÁNCHEZ

Secretario: RODOLFO SILVA CASARÍN

Vocal: DR. CARLOS A. ESCALANTE SANDOVAL

1 er. Suplente: DRA. LILIA REYES CHÁVEZ

2 do. Suplente: DR. EDGAR GERARDO MENDOZA BALDWIN

CIUDAD UNIVERSITARIA, D.F.

TUTOR DE TESIS:

DR. CARLOS A. ESCALANTE SANDOVAL

FIRMA

Esta tesis está dedicada a mi madre Aida y a mi padre Amparo que en paz descanse por su infinito amor aliento y apoyo.

AGRADECIMIENTOS

El camino ha sido difícil y he requerido de mucho esfuerzo y dedicación, sin embargo para mi buena fortuna me he encontrado con muchas personas que me han dado su compañerismo y tiempo, lo cual me ha ayudado a finalizar esta tesis.

En primer lugar agradezco a Dios, porque en todo momento siempre sostuvo mi mano y es esa fuerza que me anima a continuar diciéndome “*sigue adelante*”. Dios en su infinita misericordia y paciencia ha querido poner en mi camino a todas aquellas personas que en mucho o poco han contribuido a la culminación de mi investigación doctoral. A mi madre la santa virgen María que junto a su hijo mi eterno salvador el señor Jesucristo nunca me desampararon.

A mi madre Aida por su valioso apoyo constante, sin ella no hubiera podido seguir adelante en esta investigación. A mi padre y abuelo Amparo quien se adelantó en el camino, por sus valiosos consejos y sabiduría, tengo la esperanza que él está orgulloso de mí donde quiera que se encuentre.

Agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el patrocinio para realizar mis estudios de doctorado.

Una parte importante de esta investigación ha sido el Dr. Carlos Escalante quien ha sido mi tutor y amigo, con quien a través de esfuerzos, compartir ideas y mediante pruebas y ensayos se han desarrollado los modelos propuestos en esta tesis. No podría haber deseado tan minucioso interlocutor con tan profundo conocimiento en el campo de la hidrología.

A mis amigos los doctores Luis Fernando y Jorge Luis por las fructíferas horas llenas de ideas revolucionarias en el campo de la Hidráulica.

Leonardo Daniel Amores Rovelo

Hermano sol hermana luna

*Si quieres que se cumplan tus sueños
trabaja en ello lento pero seguro
el trabajo sincero crece con pureza.*

*Si quieres vivir en libertad
tomate tu tiempo
avanza lentamente
haz unas pocas cosas pero hazlas bien
las alegrías más sencillas son celestiales.*

*Día a día, piedra a piedra
construye lentamente tu secreto
día a día, piedra a piedra
sabrás que el cielo es la gloria.*

*Si quieres que se cumplan tus sueños
trabaja en ellos lento pero seguro
comienza con lo pequeño para un gran final
el trabajo sincero crece con pureza.*

*Si quieres vivir en libertad
tomate tu tiempo
avanza lentamente.*

San Francisco de Asís

Resumen

En esta investigación se presentan y desarrollan modelos basados en RNAs mediante un enfoque Regional. Esto es RNAs-Regional, Ensamblado-Regional y Modular-Regional. Dentro de los modelos tradicionales de predicción hidrológica se tienen a los modelos de RNAs, Ensamblados y Modulares. Estos modelos son evaluados en diferentes casos de estudio. El enfoque Regional propuesto en los diferentes modelos RNAs depende de la adecuada delimitación de regiones homogéneas o de la potencialidad de homogeneidad entre estaciones. En este caso se prueban homogeneidades mediante el algoritmo de Fisher y mediante *c*-Medias. Posteriormente se utilizan dichas regiones en los modelos predictores evaluando su desempeño. De acuerdo con la hipótesis regional, se espera que una regionalización adecuada brinde un pronóstico óptimo al tener una mayor cantidad de información, esto es mediante el refinamiento de las regiones homogéneas con ayuda de los Momentos-L. En el desarrollo de esta tesis inicialmente se tomaron miles de estaciones climatológicas e hidrométricas en cada sitio estudiado depurándose muchas de ellas debido al riguroso control de calidad. El cual consistió en inspeccionar los datos dudosos o anómalos para luego evaluar la homogeneidad de las series analizadas. Una vez llevado a cabo el control de calidad de las series de lluvia y escurrimientos, se tomaron aquellas series lo suficientemente largas y se cuidó que la cantidad de datos vacíos no fuera mayor del 10%. Para evaluar los modelos en la predicción de lluvia, se tomó como caso de estudio a Australia y la Costa Oeste de Estados Unidos y para la predicción de escurrimientos los casos de estudio fueron Canadá y la Cuenca del Amazonas. Finalmente los modelos regionales propuestos se aplicaron a un caso de estudio en la República Mexicana siendo este el estado de Zacatecas. Para Australia se tomaron 118 estaciones climatológicas de la NOAA (National Oceanic and Atmospheric Administration). Estas estaciones climatológicas tienen una longitud de registro de 106 años (1272 meses) con observaciones comprendidas entre el periodo 1907 y 2012. En el caso de la Costa Oeste de Estados Unidos se tomaron 56 estaciones climatológicas de la NOAA (National Oceanic and Atmospheric Administration) con una longitud de registro de 113 años, con un periodo comprendido del año 1900 al 2012. En la predicción de los escurrimientos, las áreas de estudio fueron Canadá, la Cuenca del Amazonas y como caso de estudio en México el estado de Zacatecas. En Canadá se estudiaron datos de escurrimientos mensuales de 90 estaciones hidrométricas en las regiones este y oeste de Canadá a partir del CD-ROM de Environment Canada (HYDAT, 2013). Los datos observados comprenden entre 99 y 114 años (1188 y 1368 meses) con observaciones comprendidas entre el periodo 1900 y 2008 para las series más largas y entre 1915 a 2008 para las series más cortas. En la cuenca del Amazonas se utilizaron series de escurrimientos de 18 estaciones hidrométricas con longitudes de registro variables comprendidas desde 1951 a 2012. Los registros fueron tomados del Observatory (ORE) HYBAM (Geodynamical, hydrological and biogeochemical control of erosion/alteration and material transport in the Amazon basin). Finalmente en Zacatecas se tomaron series de lluvias de 87 estaciones climatológicas de la Base de Datos Climatológica Nacional CLImate COMputing (CLICOM). Las series de lluvia acumulada mensual cuentan con una longitud de registro de 71 años en el periodo comprendido del año 1940 al año 2010. El enfoque Regional para modelos de predicción basados en modelos no lineales vía RNAs se ha desarrollado con el auxilio de Matlab R12 pensado para pronóstico lluvia y escurrimiento y de esta forma proporcionar otra alternativa en la toma de decisiones dentro de los

aprovechamientos hidráulicos. Este trabajo desarrolla modelos basados en RNAs desde un enfoque Regional sobre una base de datos disponibles de lluvia y de escurrimientos. Los resultados encontrados muestran múltiples ventajas no solo en términos de precisión de los modelos en general sino también en el pronóstico de las variables analizadas donde la información meteorológica juega un papel importante para la simulación de fenómenos extremos que ayudan en el aprovechamiento hidráulico. Los resultados de las evaluaciones indican que en general, los modelos Regionales presentan los mejores ajustes, sin embargo, se observó también que el modelo Ensamblado-Regional obtuvo los mejores ajustes, aunque con una diferencia mínima a los obtenidos con los modelos Modular-Regional y RNAs-Regional. De ahí que los tres modelos son una adecuada herramienta en el pronóstico hidrológico. A pesar de la mejora en la predicción mediante los modelos regionales planteados en esta tesis se espera que la principal contribución sea la de inspirar futuras investigaciones y uso de los modelos regionales en las dependencias encargadas en la planeación de los recursos hídricos del país.

CONTENIDO

INTRODUCCIÓN

I)	Motivación.	i
II)	Objetivos.	ii
III)	Estructura de la tesis.	iii
IV)	Referencias.	v
1	SERIES DE TIEMPO EN HIDROLOGÍA	1
1.1	Introducción.	1
1.2	Análisis de series de tiempo.	2
1.3	Estructura de series de tiempo.	2
1.4	Modelación de series de tiempo.	2
1.5	Características más destacadas de las series de tiempo hidrológicas	3
1.6	Clasificación de las series de tiempo.	5
1.7	Características estadísticas de las series de tiempo hidrológicas. ...	5
1.8	Técnicas de autocorrelación.	7
1.9	Aplicaciones hidrológicas.	8
1.10	Referencias.	9
2	REDES NEURONALES ARTIFICIALES (RNAs)	11
2.1	Introducción.	11
2.2	Historia de redes neuronales.	12
2.3	Aplicación de RNAs para predicción y modelado en hidrología. ...	14
2.4	Perceptron multicapa.	15
2.4.1	Estructura de la red.	15
2.5	Algoritmos de entrenamiento de MFNN.	16
2.5.1	Algoritmo de retropropagación.	17
2.5.2	Entrenamiento de MFNN mediante el algoritmo de Levenberg-Marquardt.	19
2.5.2.1	Algoritmo del descenso más rápido.	20
2.5.2.2	Algoritmo de Newton.	21
2.5.2.3	Algoritmo de Gauss-Newton.	24
2.5.2.4	Algoritmo de Levenberg-Marquardt.	26
2.6	Referencias.	26
3	MODELOS ENSAMBLADOS	29
3.1	Introducción.	29
3.2	Redes ensambladas.	31
3.3	Algoritmo de AdaBoost.M2.	31
3.4	Algoritmo LPBoost.	32
3.5	Algoritmo TotalBoost.	34
3.6	Algoritmo RUSBoost.	35

CONTENIDO

3.7	Referencias.....	37
4	MODELOS MODULARES	38
4.1	Introducción.....	39
4.2	Concepto de modularidad.....	39
4.3	Redes Neuronales Artificiales Modulares RNAM.....	41
4.4	Modelo matemático de una RNAM.....	41
4.5	Referencias.....	44
5	REGIONALIZACION DE CUENCAS	45
5.1	Introducción.....	45
5.2	Atributos aunados en la regionalización.....	47
5.3	Clasificación de algoritmo clúster.....	47
5.4	Pasos en la regionalización mediante el análisis clúster.....	48
5.5	Clasificación mediante el algoritmo Fisher.....	49
5.6	Teoría de Momentos-L.....	50
5.7	Pruebas de homogeneidad regional	51
	5.7.1 Prueba de homogeneidad usando la medida de heterogeneidad.....	52
5.8	Ajuste de regiones.....	53
5.9	Regionalización mediante el análisis Fuzzy C-Medias.....	53
5.10	Referencias.....	55
6	EXPERIMENTOS Y RESULTADOS	57
6.1	Introducción.....	58
6.2	Estado del arte de la última década en el pronóstico de lluvia.....	58
6.3	Estado del arte de la última década en el pronóstico de escurrimiento.....	59
6.4	Modelos de predicción en combinación con técnicas de agrupamiento.....	60
6.5	Medidas de eficiencia.....	64
6.6	Comparación de los modelos de predicción en lluvias.....	65
6.7	Comparación de los modelos RNAs y RNAs-Regional.....	71
6.8	Comparación de los modelos Ensamblado y Ensamblado Regional.....	75
6.9	Comparación de los modelos Modular y Modular Regional.....	79
6.10	Comparación de modelos de predicción en escurrimientos.....	83
6.11	Aplicaciones y resultados.....	86
6.12	Análisis de grupos mediante c-Medias.....	86
6.13	Selección de datos usando la medida de discordancia D_i	87
6.14	Heterogeneidad regional y formación de regiones homogéneas....	88
6.15	Comparación de los modelos RNAs y RNAs Regional.....	90

CONTENIDO

6.16	Comparación de los modelos Ensamblado y Ensamblado-Regional	94
6.17	Comparación de los modelos Modular y Modular Regional.	98
6.18	Comparación de escurrimientos en las escalas mensual trimestral y semestral.	102
6.19	Preparación de datos.	104
6.20	Metodología.	105
6.21	Delimitación de regiones homogéneas.	105
6.22	Análisis clúster.	107
6.23	Análisis de series mensuales.	111
6.24	Análisis de series trimestrales.	114
6.25	Análisis de series semestrales.	117
6.26	Comparación de lluvias en las escalas mensual trimestral y semestral.	120
6.27	Preparación de datos.	121
6.28	Metodología.	123
6.29	Delimitación de regiones homogéneas.	123
6.30	Resultados.	124
6.31	Análisis de series mensuales.	124
6.32	Análisis de series trimestrales.	131
6.33	Análisis de series semestrales.	137
6.34	Comparación de lluvias en las escalas mensual trimestral y semestral en Zacatecas.	143
6.35	Preparación de datos.	144
6.36	Metodología.	145
6.37	Delimitación de regiones homogéneas.	145
6.38	Resultados.	146
6.39	Referencias.	154
7	CONCLUSIONES Y RECOMENDACIONES	155
A	Redes neuronales artificiales RNAs.	157
B	Algoritmos de optimización.	160
	Lista de figuras.	165
	Lista de tablas.	169

INTRODUCCIÓN

INTRODUCCIÓN

I) Motivación

La noción básica de redes neuronales artificiales RNAs, probablemente fue formalizada por primera vez por McCulloch y Pitts (1943) en su modelo de neurona artificial. La investigación en este campo permaneció dormida en los primeros años, quizás esto fue porque aún no se descubría la capacidad de estos modelos y porque no había una indicación clara sobre su gran potencial. Sin embargo, el interés en estos modelos fue en un momento de forma dramática con los trabajos de Hopfield (1982) y Rumelhart et al (1986). Estos estudios no solo dieron lugar a las redes neuronales artificiales en una base matemática firme, también abrieron la puerta a aplicaciones potenciales teniendo un progreso rápido en diferentes frentes: desarrollo teórico de diferentes algoritmos de aprendizaje, capacidades computacionales y aplicaciones en diversas áreas de la neurofisiología al mercado.

Los estudios iniciales en redes neuronales fueron impulsados por el deseo de tener ordenadores que imitaran el aprendizaje humano. Las RNAs también se les han referido como neuroordenadores por gente que quiere preservar la analogía inicial. La siguiente definición de RNAs fue hecha por Haykin (1994) es una adaptación por otra temprana ofrecida por Aleksander y Morton (1990):

“Es un procesador masivamente distribuido en paralelo que tiene una condición natural para almacenar conocimiento experimental y ponerlo en disposición para su uso, lo que se asemeja al cerebro humano en dos aspectos:

- 1. El conocimiento se adquiere por la red a través de un proceso de aprendizaje*
- 2. Las fuerzas de conexión de interneuronas conocidos como pesos sinápticos se utilizan para almacenar la información.”*

Hasta la fecha hay un gran número de artículos que exaltan las virtudes de las redes neuronales como herramienta computacional y realza las bondades sobre las técnicas convencionales. Aunque también hay una serie de fracasos decepcionantes asociados a esta tecnología pero estos tienden a ser menores. Aparte de ser una novedad, RNAs también son una mezcla de éxitos y fracasos. Sin embargo si se conociera siempre el resultado de un problema se tendría poco interés. Investigaciones afirman que deberían utilizar RNAs por que poseen atributos deseables de aproximadores universales, habilidad de aprendizaje y de procesar grandes volúmenes de datos a altas velocidades (Hornik *et al.*, 1989, Hornik *et al.*, 1990, Cybenko, 1989).

En hidrología se ha tardado en adoptar a las RNAs, principalmente por la aceptación por parte de los profesionales de nuevas tecnologías que tienen que ser demostradas y probadas. Otra razón es porque los trabajos iniciales relacionados con RNAs han hecho uso de esta tecnología como una caja negra. Con algunas ventajas sobre los modelos de regresión las RNAs poseen una mejor capacidad de generalización y no requieren estrictamente de una forma funcional matemática, en este sentido las redes neuronales artificiales son llamadas modelos de libre aproximación.

INTRODUCCIÓN

Las RNAs es uno de los enfoques más amplios dentro de la hidroinformática que surgió en la década de 1990 como una vía para administrar la sobrecarga de información de una manera eficaz (Govindaraju y Rao, 2000). En el desafío de la hidrología se encuentra la gestión del agua en sus múltiples dimensiones y aplicaciones puesto que se necesitan de técnicas que requieren una gran variedad de componentes: la complejidad hidráulica, la calidad del agua, la planificación financiera y los programas sociales. Dentro de la hidrología las técnicas disponibles son muchas y variadas, cada una con sus ventajas y desventajas. La variabilidad en el clima, la singularidad de las cuencas, el caos del sistema y las complejidades de integración a gran escala junto con la dificultad del acceso de la información, hacen que la tarea de predicción sea difícil. Las RNAs no pretenden sustituir los modelos de predicción sino más bien una forma alternativa y complementaria de abordar los problemas de predicción.

Las inundaciones y sequías siguen ocurriendo en muchos países en el mundo causando enormes bajas y severos daños. Para mitigar los efectos de estos fenómenos adversos se pueden adoptar tanto medidas estructurales como no estructurales. Sin embargo la alerta temprana de estos fenómenos con suficiente tiempo tiene la ventaja de poder salvar vidas permitiendo tomar medidas a tiempo. Por tal motivo RNAs deben de ser incluidas en la modelación hidrológica. Sin embargo esta tecnología está aún en su etapa inicial dentro de la hidrología. Mientras que muchos investigadores tienen a las RNAs en diferentes contextos, solo una pequeña porción de ellos son conscientes de la capacidad que brinda esta tecnología en términos de aplicaciones hidrológicas. En este trabajo se plantean *Modelos Regionales* basados en RNAs, *Modelos Ensamblados* y *Modelos Modulares* enfocados exclusivamente a la hidrología. La naturaleza de la hidrología como una disciplina la hace única, requiriendo diferentes perspectivas e interpretaciones que se mejoran cada vez más por los investigadores de este campo. Este trabajo representa una aportación más en tal esfuerzo.

II) Objetivos

El objetivo principal de esta tesis es desarrollar modelos de predicción mediante un enfoque regional basado en modelos de Redes Neuronales Artificiales (RNAs), Ensamblados y Modulares que permita predecir las variables hidrológicas de lluvia y escurrimiento en las escalas de tiempo mensual, trimestral y semestral.

Los objetivos específicos consisten en determinar si los modelos propuestos son mejores predictores que aquellos modelos usados tradicionalmente para la predicción de lluvia y escurrimiento.

- 1) Explorar RNAs arquitecturas, funciones de transferencia, número de capas ocultas así como algoritmos de entrenamiento y desarrollar un marco óptimo del uso de RNAs.
- 2) Explorar y optimizar los modelos modulares basados en RNAs que permite construir modelos basados en datos.
- 3) Explorar los modelos Ensamblados y optimizarlos para mejorar la construcción de modelos basados en datos.
- 4) Aplicar los modelos de predicción basados en RNAs con un enfoque Regional en la predicción mensual, trimestral y semestral de datos de lluvia y escurrimientos.

INTRODUCCIÓN

III) Estructura de la tesis

La tesis está organizada de la siguiente manera, en la introducción se presenta un panorama general del tema, se plantea el problema, se proponen hipótesis, se dan a conocer los objetivos. La tesis está estructurada en siete capítulos.

- Capítulo 1. Este capítulo se revisan los conceptos básicos de series de tiempo hidrológicas así como sus características estadísticas.
- Capítulo 2. Se hace una revisión en la literatura sobre los modelos RNAs aplicados a la hidrología con el fin de proporcionar el contexto científico en esta investigación. Dicha revisión se presenta como un resumen histórico de las principales aplicaciones de las RNAs a la hidrología en la última década.
- Capítulo 3. Introduce a los modelos ensamblados ENN (Ensemble Neural Network). Se ha demostrado que la capacidad de generalización de RNAs se puede mejorar mediante la combinación de varias RNAs ensambladas.
- Capítulo 4. En este capítulo se exploran los modelos modulares MNN. Se especifica su funcionamiento así como la descripción general para su aplicación al pronóstico de lluvia y escurrimiento.
- Capítulo 5. En este capítulo se exploran diferentes métodos de análisis clúster para la formación de regiones mediante atributos de la misma dando un especial énfasis en Momentos-L.
- Capítulo 6. En este capítulo se exponen los resultados encontrados con el enfoque Regional propuesto mediante los modelos no lineales RNAs, Ensamblados y Modulares basados en RNAs.
- Capítulo 7. En el capítulo siete se presentan las conclusiones obtenidas a través del desarrollo de la tesis, así como algunas propuestas para futuras investigaciones afines al tema.

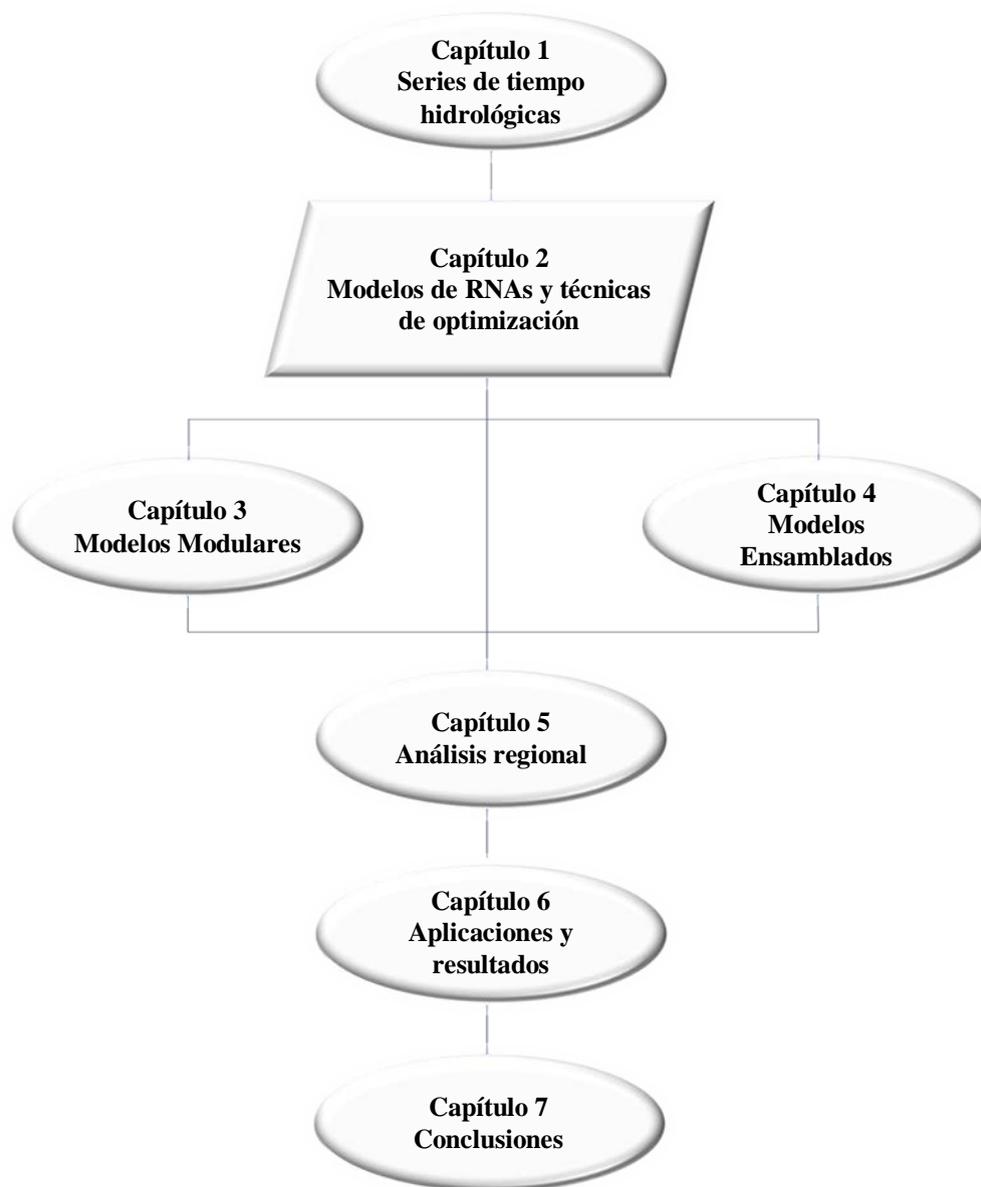


Figura i Diagrama de la tesis

INTRODUCCIÓN

IV) Referencias

- Aleksander, I. And H. Morton (1990). An introduction to Neural Computing, Chapman and Hall, London.
- Cybenko, G. (1989): "Approximation by Superpositions of a Sygmoidal Function", *Mathematics of Control, Signals and Systems*, 2, pp. 303-314.
- Govindaraju, R.S., Rao, A.R., (2000). *Artificial Neural Networks in Hydrology*, Kluwer, The Netherlands.
- Haykin, S. (1994) *Neural Networks: A Comprehensive Foundation* Macmillan College Publishing.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554.
- Hornik, K., Stinchcombe, M. y White, H. (1989): "Multilayer feedforward networks are universal approximators". *Neural Networks*, 2, pp. 359-366.
- Hornik, K., Stinchcombe, M. y White, H. (1990). Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3, 551-560.
- McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133.
- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representation by error propagation. In: Rumelhart DE, McClelland JL and the PDP Research Group (eds) *Parallel distributed processing*, vol 1, chap 8. MIT Press, Cambridge, Mass.

CAPÍTULO**UNO**

SERIES DE TIEMPO EN HIDROLOGÍA

En este capítulo se presenta elementos fundamentales de las series de tiempo en hidrología, las cuales son la materia prima en el desarrollo de esta investigación. Mismas que han servido para el desarrollo y calibración de los modelos analizados y propuestos siendo el principal objetivo de esta tesis introducir modelos capaces de predecir valores de series de tiempo hidrológicas. En este capítulo se presentan las fórmulas estadísticas que fueron necesarias en la aplicación de los modelos no lineales estudiados así como el proceso regional para tal fin es necesario conocer las características estadísticas de las series de tiempo hidrológicas para realizar un análisis adecuado.

1.1 Introducción

Al término de *serie de tiempo* se ha definido como “una secuencia de valores recogidos en el tiempo de una variable en particular” (Haan, 1977). Una serie de tiempo puede estar constituida solo por eventos deterministas, estocásticos o una combinación de ambos.

Una serie de tiempo hidrológica generalmente está integrada por una componente estocástica y una determinista (Haan, 1977). La componente determinística puede ser clasificada como, una tendencia, saltos, periodicidad o combinación de éstas. Los intervalos de tiempo en las series de tiempo hidrológicas generalmente son del orden de horas, días, semanas, meses, periodos o años.

En el análisis de las series de tiempo hidrológicas es de suma importancia el conocimiento de las características estadísticas pues existe dependencia estadística en las series de tiempo provenientes de procesos físicos. Se asumen dos tipos de dependencias estadísticas: *dependencia temporal*, que tiene relación con las características estadísticas de los valores de la serie en dos instantes de tiempo distintos y la *dependencia espacial* la cual guarda relación con la correlación de la serie de datos en distintos lugares. Por lo tanto es esencial conocer las características comunes de las series de tiempo hidrológicas.

1.2 Análisis de series de tiempo

El análisis de series de tiempo es la investigación de una secuencia de datos distribuidos en el tiempo o la síntesis de un modelo para la predicción en el que el tiempo es una variable independiente. Uno de los principales objetivos del análisis de series de tiempo es detectar y describir cuantitativamente cada uno de los procesos que subyacen a una determinada secuencia de observaciones. En el campo de la hidrología la razón principal es detectar una tendencia debido a otra variable hidrológica aleatoria y en segundo lugar la serie de tiempo puede ser analizada para desarrollar y calibrar un modelo que pueda describir las características dependientes del tiempo de una variable hidrológica.

Durante la etapa del análisis de la serie de tiempo, se examinan los componentes sistemáticos para identificar sus características, incluyendo la magnitud, la forma y la duración.

En la etapa de síntesis, se utiliza la información obtenida en el análisis para desarrollar un modelo de series de tiempo y evaluar tal modelo mediante pruebas de bondad de ajuste. Finalmente en la verificación, el modelo desarrollado es evaluado usando series de datos independientes.

1.3 Estructura de series de tiempo

Una serie de tiempo es comúnmente descrita como una función integrada por cuatro componentes: *tendencia*, *estacionalidad*, *componente estocástica dependiente* y *residual independiente*. En general, una serie de tiempo puede ser expresada matemáticamente como.

$$x_t = T_t + S_t + \varepsilon_t + \delta_t \quad (1.1)$$

Donde T_t es la componente de tendencia, S_t es la estacionalidad, ε_t componente estocástica dependiente y δ_t es la componente residual independiente. Dentro del análisis de las series de tiempo se acepta que los datos (observaciones) consisten en un patrón sistemático y un ruido aleatorio (error), donde este último hace que el patrón sea difícil de identificar. El patrón sistemático está representado por los dos primeros componentes de la ecuación (1.1) siendo de naturaleza determinista mientras que el componente estocástico representa el error aleatorio.

1.4 Modelación de series de tiempo

La modelación de una serie de tiempo puede ser simple o compleja, esto depende del tipo de serie que se modela, del tipo de modelo que se utiliza así como de las técnicas de modelación que se han escogido. Para series de tiempo en las que sus características estadísticas no varían con el tiempo, es posible modelarlas con modelos más simples que aquellas en las cuales sus características varían con el tiempo. En un modelo en particular existen varias técnicas para estimar sus parámetros así como para estimar el rendimiento del modelo. También en este caso existen técnicas más complejas que otras.

En general la modelación puede ser organizada en los siguientes pasos (Box and Jenkins, 1970).

- 1) Selección del tipo de modelo
- 2) Identificación de la forma del modelo
- 3) La estimación de los parámetros del modelo y
- 4) La comprobación del modelo

1.5 Características más destacadas de las series de tiempo hidrológicas

En la mayoría de los análisis estadísticos de series de tiempo hidrológicas en las diferentes escalas de tiempo usuales para el estudio de recursos hídricos se basan en un conjunto de supuestos fundamentales, los cuales son: la serie es homogénea, estacionaria, libre de tendencias, cambios, sin persistencias y no periódicas (Adeloye & Montaseri, 2002).

Cuando se habla de homogeneidad se considera que los datos pertenecen a una población y por lo tanto tienen una media invariante en el tiempo. La no homogeneidad puede darse debido al método de recolección de datos o al ambiente en el que se recogen (Fernando & Jayawardena, 1994). Por otra parte el concepto de estacionariedad implica que los parámetros estadísticos de las series calculadas a partir de diferentes muestras no cambian, excepto por las variaciones muestrales. Se dice que una serie de tiempo es exactamente estacionaria si sus propiedades estadísticas no varían sensiblemente con el tiempo. Un tipo de estacionariedad, llamada estacionariedad débil o estacionariedad de segundo orden es aquella en la que los momentos de primer y de segundo orden solo dependen de las diferencias de tiempo (Chen & Rao, 2002). Sin embargo por practicidad las series hidrológicas se dicen que son estacionarias.

Los cambios en las series hidrológicas pueden ser debidos a diferentes factores. Un tipo de cambio puede ser gradual esto es conocido como tendencia o de forma abrupta conocido como salto o puede adoptar formas más complejas (Shahin et al., 1993). La tendencia se define como un cambio unidireccional y gradual (descendente o ascendente) en el valor medio de una variable. Se dice que una serie de tiempo tiene tendencias, si existe una correlación significativa (positiva o negativa) entre los valores observados y el tiempo. Las tendencias y cambios en las series de tiempo hidrológicas generalmente son debidas a los cambios naturales o inducidos por el hombre (Haan, 1977; Salas, 1993).

Los cambios graduales o naturales en las series hidrológicas podrían también ser causados por un cambio climático de escala global o regional. Pero también estos cambios pueden ser debidos a la urbanización gradual sobre el sitio de monitoreo o por cambios en la estación de monitoreo esto es cambiar de lugar a la estación de medición. Los saltos en una serie de tiempo por lo general son resultado de eventos naturales catastróficos como terremotos, tsunamis, ciclones o grandes incendios forestales, que ocurren de forma repentina y alteran considerablemente el ciclo hidrológico de la región. Otra de las causas de los saltos en las series son por el hombre tales como el cierre de una presa, inicio o finalización de extracción en aguas subterráneas o de actividades de desarrollo en la región también influyen en los saltos (Haan, 1977).

La Periodicidad es otra de las características de las series temporales hidrológicas que generalmente es producida debido a los ciclos astronómicos, como la rotación de la Tierra alrededor del Sol (Haan, 1977; Kite, 1989). Los ciclos anuales suelen ser más evidentes en las series de lluvias, evapotranspiración, escurrimientos, niveles freáticos, humedad del suelo y de otros tipos de datos hidrológicos (Haan, 1977). Los ciclos semanales pueden estar presentes en los datos de uso del agua del sector doméstico, industrial o agrícola; muchas veces las series de tiempo de uso del agua contienen periodicidades anuales y semanales (Haan, 1977). Para identificar y cuantificar la periodicidad en una serie hidrológica, la escala de tiempo se debe considerar menor a un año (es decir, un mes o seis meses). La periodicidad no es perceptible en una serie de tiempo anual, y por lo tanto tampoco en la serie media anual o mensual. Por último el fenómeno de la "persistencia" denota la tendencia a que la magnitud de un evento para ser dependiente en la magnitud del evento anterior, es decir, tiene memoria. Por ejemplo, la tendencia de que los escurrimientos bajos sigan a los escurrimientos bajos y que los altos sigan a los altos. Por lo tanto, "persistencia" se puede considerar sinónimo de autocorrelación (O'Connel, 1977).

El análisis de series de tiempo tiene una gran gama de aplicaciones en la geología, la tecnología oceánica, la sismología, y también se ha aplicado a muchas situaciones hidrológicas y climatológicas. Por ejemplo, el análisis de series de tiempo se han llevado a cabo para el análisis de los datos de lluvia (por ejemplo, Mirza et al, 1998; Pugacheva et al, 2003;.. Astel et al, 2004), datos de escurrimiento (Radziejewski et al, 2000; Fanta et al, 2001; Adeloye y Montaseri, 2002), los datos de inundación (Westmacott & Burn, 1997; Robson et al, 1998; Cayan et al, 1999; Douglas et al, 2000; Zhang et al, 2001; Cunderlik & Burn, 2002), los datos de infiltración (Schwankl et al., 2000), y los datos de calidad del agua de superficie (Higashino et al., 1999), así como para la generación de datos de lluvia (Janos et al., 1988), determinar el consumo de agua patrones (Maidment & Parzen, 1984), la detección de tendencias en la evapotranspiración y velocidad del viento (Hameed et al, 1997; Raghuwanshi & Wallender, 1997), y para la detección de los cambios climáticos (Kite, 1989; Khan, 2001). El objetivo principal del análisis de series temporales es detectar y describir cuantitativamente todos los procesos que generan una secuencia dada de observaciones (Shahin et al., 1993). En hidrología el análisis de series de tiempo se utiliza para el desarrollo de modelos matemáticos para generar registros hidrológicos sintéticos, para pronosticar fenómenos hidrológicos, para detectar las tendencias y los cambios en los registros hidrológicos, y para rellenar los datos que faltan y ampliar los registros (Salas, 1993). Por ejemplo, la planificación y el diseño de estructuras de captación de agua de lluvia eficiente requiere registros pico de escurrimientos medidos, que pueden ser de menor longitud que el periodo de retorno de diseño. En esta situación, los registros máximos de escurrimientos se pueden ampliar mediante la generación de datos de escurrimientos sintéticos. Originalmente, la aplicación del análisis de series de tiempo se limitó al análisis de extremos (inundaciones y sequías). Sin embargo, el dominio de análisis de series de tiempo se ha ampliado para abarcar los problemas de las aguas subterráneas (Shahin et al., 1993). Con un dominio tan amplio, el análisis de series de tiempo se ha convertido en una poderosa herramienta para la planificación y gestión eficiente de los escasos recursos hídricos.

1.6 Clasificación de las series de tiempo

Las series de tiempo pueden ser clasificadas de muchas maneras de acuerdo diferentes criterios. De acuerdo con Salas (1993) la clasificación de las series hidrológicas es como sigue.

Series de tiempo continuas o discretas

Las series de tiempo pueden ser discretas o continuas. Son discretas si son registradas en diferentes instantes de tiempo o en diferentes puntos en el espacio (Haan, 1977; Shahin et al., 1993). O en otro caso si las observaciones son recolectadas continuamente en el tiempo o en el espacio, entonces se dice que es una serie de tiempo continua. Usualmente en hidrología las series de tiempo son del tipo discreta.

Duración completa o parcial de las series

Una serie de tiempo completa es la que contiene todas las observaciones registradas con el tiempo o en el espacio (Haan, 1977; Shahin et al., 1993). Como su nombre indica, una serie de duración parcial contiene las observaciones que son parte de una serie de tiempo completa. Una serie de duración parcial siempre contiene menos información que una serie de tiempo completa. Además, los puntos de observación en una serie de duración parcial pueden no ser equidistantes.

Univariada o multivariada

Si se observa sólo una variable en cada momento, la serie de tiempo se conoce como 'Series de tiempo univariadas'. Sin embargo, si se observan dos variables al mismo tiempo (simultáneamente), la serie es conocida como 'series de tiempo de dos variables'. Si más de dos variables se observan simultáneamente a la vez, la serie es conocida como "serie temporal multivariante".

1.7 Características estadísticas de las series de tiempo hidrológicas

En el análisis de series de tiempo hidrológicas, aquellas que satisfagan los supuestos de homogeneidad, no tendencia, no periodicidad, no persistencia y estacionalidad, pueden ser la excepción (Rao et al., 2003). Es necesario realizar análisis estadísticos preliminares en todos los estudios de recursos hídricos que implican el uso de los datos de series de tiempo hidrológicas (Adeloye & Montaseri, 2002). Sin embargo en la mayoría de los análisis de series de tiempo se utilizan métodos estándar después de la holgura en las condiciones de homogeneidad, bajo los supuestos de que los resultados no se afectan lo suficiente por estos supuestos (Rao et al., 2003). Dentro de los estudios de análisis de series hidrológicas no se consideran todos los aspectos de las series de tiempo. Gran parte de los trabajos hidrológicos se ocupan del análisis de tendencias lineales despreciando las otras características, es decir, la homogeneidad, la estacionalidad, periodicidad y persistencia se ignoran. De hecho, muy pocos estudios se refieren a un análisis completo y extenso de series de tiempo y componentes estocásticos en hidrología. Para corregir esta situación, el presente estudio intenta emplear todos los métodos posibles para diferentes pruebas de series de tiempo y para evaluar el rendimiento relativo de estas pruebas para la detección de características particulares. La aplicación de estas pruebas se demuestra a través de un

estudio de caso en el análisis de series de tiempo de las series de lluvias máximas diarias consecutivos anuales de Kharagpur, Bengala Occidental, India.

Las características estadísticas de mayor importancia comúnmente usadas en el análisis de series de tiempo hidrológicas son las siguientes.

La media es la primera característica más importante en las series de tiempo hidrológicas y se describe de la siguiente forma. Si se tiene una serie de tiempo x_t es continúa tal que $x_1, x_2, \dots, x_t \dots x_N$ sea de tamaño N . Entonces la media \bar{x} viene dada por

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t \quad (1.2)$$

Donde N es el tamaño de la muestra (longitud de la serie de tiempo hidrológica). La media muestral \bar{x} es la estimación de la media de la población (valor esperado). Esta es una medida de tendencia central de x_t .

La segunda característica estadística más importante de una serie de tiempo hidrológica es la varianza muestral s^2 y viene dada por

$$s^2 = \frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2 \quad (1.3)$$

Esta es un estimador sesgado de la varianza poblacional, el estimador insesgado poblacional viene dado mediante la siguiente expresión

$$s^2 = \frac{1}{N - 1} \sum_{t=1}^N (x_t - \bar{x})^2 \quad (1.4)$$

El estimador insesgado generalmente es el más usado en hidrología estadística. La raíz cuadrada de la varianza s^2 es llamada desviación estándar s . La relación de la media y la desviación estándar es el coeficiente de variación $\frac{s^2}{\bar{x}}$ para una población y para una muestra respectivamente. Un pequeño valor de s indica que los valores $x_1, x_2, \dots, x_t \dots x_N$ no difieren mucho de \bar{x} . En el caso contrario un valor grande de s indica una alta variación de los valores respecto de \bar{x} .

El coeficiente de asimetría muestral de una serie de tiempo puede ser determinado como

$$g = \frac{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^3}{s^3} \quad (1.5)$$

Donde s es la desviación estándar. Este es un estimador poblacional sesgado. El estimador insesgado viene dado mediante la siguiente expresión

$$g = \frac{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^3}{(N-1)(N-2)s^3} \quad (1.6)$$

Este estimador es generalmente el más usado en hidrología. El coeficiente de asimetría indica la asimetría de una serie de tiempo. Si $g = 0$ entonces la distribución de probabilidad de la serie x_t sería simétrica alrededor de \bar{x} . Si $g < 0$ entonces la función de distribución está sesgada a la derecha mientras que si $g > 0$ esta presenta un sesgo a la izquierda.

1.8 Técnicas de autocorrelación

La función de autocorrelación en esencia expresa el grado temporal de dependencia entre observaciones. En realidad es un proceso de auto-comparación que expresa la correlación lineal entre una serie igualmente espaciada y la misma serie en un lapso de tiempo o retraso determinado (Jenkins & Watts, 1968).

La función de autocovarianza es una medida que indica el grado de autodependencia lineal de una serie de tiempo. La autocovarianza poblacional γ_k entre x_t y x_{t+k} matemáticamente puede ser expresada de la siguiente manera (Box y Jenkins, 1976).

$$\gamma_k = E[(x_t - \mu)(x_{t+k} - \mu)] \quad (1.7)$$

Donde x_t es el valor de la variable en el tiempo t , k es el tiempo de retraso, μ es la media poblacional y E es el operador de esperanza. Por otra parte la función de autocorrelación poblacional (ρ_k) es definida como un ratio de la autocovarianza poblacional (γ_k) a la varianza poblacional [$Var(x_t)$]. Siendo esta.

$$\rho_k = \frac{\gamma_k}{Var(x_t)} \quad (1.8)$$

La autocovarianza muestral c_k es un estimador sesgado de la función de autocovarianza poblacional γ_k y esta puede ser determinada por

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (1.9)$$

Donde c_k es usualmente llamada autocovarianza de retraso k , donde k representa el tiempo de retraso o distancia entre los pares correlacionados (x_t, x_{t+k}) , \bar{x} es la media muestral y N es el tamaño de la muestra. Para los casos particulares en que $k = 0$ entonces c_0 se convierte en la varianza.

Una medida adimensional de la dependencia lineal es obtenida dividiendo c_k por c_0 dicha operación da

$$r_k = \frac{c_k}{c_0} = \frac{\frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\frac{1}{N} \sum_{t=1}^N (x_t - \bar{x})^2} \quad (1.10)$$

Donde r_k es llamado coeficiente de autocorrelación, el coeficiente de correlación serial o función de autocorrelación. El diagrama r_k versus k es generalmente llamado correlograma. El coeficiente de autocorrelación muestral r_k es más comúnmente usado como medida de dependencia de series de tiempo y usualmente el primer coeficiente de correlación r_1 .

Cabe mencionar que la autocorrelación de la población ρ_k puede ser estimada por la función de autocorrelación serial r_k usando los datos de la muestra mediante la siguiente expresión (Shahin et al., 1993; Haan, 2002):

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t \cdot x_{t+k}) - 1/(N-k) \cdot \sum_{t=1}^{N-k} x_t \sum_{t=1}^{N-k} x_{t+k}}{\left[\sum_{t=1}^{N-k} x_t^2 - \frac{1}{N-k} \cdot \left(\sum_{t=1}^{N-k} x_t \right)^2 \right]^{1/2} \cdot \left[\sum_{t=1}^{N-k} x_{t+k}^2 - \frac{1}{N-k} \cdot \left(\sum_{t=1}^{N-k} x_{t+k} \right)^2 \right]^{1/2}} \quad (1.11)$$

Donde N es el tamaño de la muestra, k es el tiempo de retraso y x_{t+k} es el valor de la variable en el tiempo $t + k$. Para $k = 0$ resulta $r_k = 1$. Así que si el retraso k incrementa entonces r_k decrementa es común en la práctica establecer el límite superior del retraso entre $0.1N$ a $0.25N$ dependiendo del tamaño de la serie (Matalas, 1967). La información detallada sobre la estructura interna de la serie de tiempo se puede conseguir examinando el autocorrelograma el cual toma valores entre 0 y 1 teniendo a k como abscisa. Los valores críticos superior e inferior de la función de autocorrelación pueden ser obtenidos mediante la prueba de Anderson (Anderson, 1942).

1.9 Aplicaciones hidrológicas

En las últimas décadas la aplicación de la estadística especialmente en hidrología se ha ampliado abarcando diferentes áreas, entre ellas la predicción. Aunado a ello también se tiene un rápido avance de las tecnologías de la información y de datos. Estos poderosos avances hacen énfasis en el análisis de las series de tiempo hidrológicas, convirtiéndola en una herramienta indispensable en la hidrología. Dentro del análisis de series de tiempo en hidrología están el desarrollo de modelos matemáticos para la generación sintética de datos hidrológicos, modelos para pronóstico, modelos de interpolación. El interés especial en los modelos de predicción reside en la incertidumbre ante el cambio climático y su principal interés por comprenderlo. Es importante destacar que la confiabilidad de un modelo hidrológico depende de los datos así como del tiempo en el desarrollo de este y la

complejidad que involucra. Sin embargo ningún modelo matemático puede describir la hidrología de algún sistema natural (Haan, 2002). Sin embargo si se puede disminuir el error modelando mejor la realidad hidrológica. Por eso dentro de este contexto este trabajo se enfoca en proponer modelos de predicción mediante el análisis no lineal mediante redes neuronales artificiales (RNAs) y desde un enfoque regional. Esto permitirá reducir la incertidumbre en la toma de decisiones en los problemas de predicción hidrológica.

1.10 Referencias

- Adeloye, A.J. and Montaseri, M., (2002). Preliminary streamflow data analyses prior to water resources planning study. *Hydrological Sciences Journal*, 47(5): 679-692.
- Anderson, R.L. (1942). Distribution of the serial correlation coefficient. *Annals of Mathematical Statistics*, 13: 1-13.
- Astel, A., Mazerski, J., Polkowska, Z. and Namiesnik, J. (2004). Application of PCA and time series analysis in studies of precipitation in Tricity (Poland). *Advances in Environmental Research*, 8(3-4): 337-349.
- Box, G. E. P. and Jenkins, G., (1970). *Time Series Analysis, Forecasting and Control*, San Francisco, Holden-Day.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, USA.
- Cayan, D. R., Redmond, K. T. & Riddle, L. G. (1999) ENSO and hydrologic extremes in the western United States. *J. Climate* 12, 2881-2893.
- Chen, H.-L. And Rao, A.R., (2002). Testing hydrologic time series for stationarity. *Journal of Hydrologic Engineering*, ASCE, 7(2): 129-136.
- Cunderlik, J. M. & Burn, D. H. (2002) Local and regional trends in monthly maximum flows in southern British Columbia. *Can. Water Resour. J.* 27 (2), 191-212.
- Douglas, E. M., Vogel, R. M. & Kroll, C.N. (2000) Trends in floods and low flows in the United States: impact of spatial correlation. *J. Hydrol.* 240 (1-2), 90-105.
- Fanta, B., Zaake, B. T. & Kachroo, R. K. (2001) A study of variability of annual river flow of the southern African region. *Hydrol. Sci. J.* 46 (4), 513-524.
- Fernando, D. A. K. & Jayawardena, A. W., (1994). Generation and forecasting of monsoon rainfall data. In: 20th WEDC Conference on Affordable Water Supply and Sanitation (Colombo, Sri Lanka), 310-313.
- Haan, C.T., (1977). *Statistical Methods in Hydrology*. The Iowa State University Press, Ames, Iowa, 378 pp.
- Haan, C.T. (2002). *Statistical Methods in Hydrology*. 2nd edition, Iowa State University Press, Iowa, 496 pp.
- Hameed, T., Marino, M.A., DeVries, J.J. and Tracy J.C. (1997). Method for trend detection in climatological variables. *Journal of Hydrologic Engineering*, ASCE, 2(4): 154-160.
- Higashino, M., Kanda, T. & Michioku, K. (1999) Time series analysis and transformation of water quality in an eutrophic reservoir. *Mizu Kankyo Gakkaishi* 22 (8), 668-676.
- Janos, B., Lucien, D. & Omar, H. R. (1988) Practical generation of synthetic rainfall event time series in a semi-arid climatic zone. *J. Hydrol.* 103, 357-373.
- Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and Its Applications*. Holden-Day, Oakland, CA, 525 pp.
- Khan, A. R. (2001) Analysis of hydro-meteorological time series in the upper Indus basin: searching evidence for climatic change. International Water Management Institute, Working Paper 23, Pakistan Country Series Number 7.
- Kite, G., (1989). Use of time series analyses to detect climatic change. *Journal of Hydrology*, 111: 259-279.
- Maidment, D. R. & Parzen, E. (1984) Time patterns of water use in six Texas cities. *J. Water Resour. Plan. Manage.* ASCE 110(1), 90-106.
- Hameed, T., Marino, M. A., DeVries, J. J. & Tracy J. C. (1997) Method for trend detection in climatological variables.
- Matalas, N.C. (1967). Time series analysis. *Water Resources Research*, 3(3): 817-829.
- Mirza, M. Q., Warrick, R. A., Ericksen, N. J. & Kenny, G. J., 1998. Trends and persistence in precipitation in the Ganges, Brahmaputra and Meghna river basins. *Hydrol. Sci. J.* 43 (6), 845-858.

- O'Connell, P.E., (1977). ARIMA models in synthetic hydrology. *In*: T.A. Ciriani, U. Maione and J.R. Wallis (editors), *Mathematical Models for Surface Water Hydrology*. John Wiley and Sons, Inc., New York.
- Pugacheva, G., Gusev, A., Martin, I., Schuch, N. & Pankov, V. (2003). 22-year periodicity in rainfalls in littoral Brazil. *Geophys. Res. Abstracts* 5 (06797). European Geophysical Society.
- Radziejewski, M., Bardossy, A. & Kundzewicz, Z. W. (2000) Detection of change in river flow using phase randomization. *Hydrol. Sci. J.* 45 (4), 547–558.
- Raghuwanshi, N. S. & Wallender, W. W. (1997) Field measured evapotranspiration as a stochastic process. *Agric. Water Manage.* 32, 111–129.
- Rao, A. R., Hamed, K. H. & Chen, H.-L. (2003) *Nonstationarities in Hydrologic and Environmental Time Series*. Water Science and Technology Library 45. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Robson, A. J., Jones, T. K., Reed, D. W. & Bayliss, A. C. (1998) A study of national trend and variation in UK floods. *Int. J. Climatol.* 18, 165–182.
- Salas, J.D., (1993). Analysis and modeling of hydrologic time series. *In*: D.R. Maidment (editor-in-chief), *Handbook of Hydrology*. McGraw-Hill, Inc., USA, pp. 19.1-19.72.
- Schwankl, L. J., Raghuwanshi, N. S. & Walender, W. W. (2000) Time series modeling for predicting spatially variable infiltration. *J. Irrig. Drain. Engng ASCE* 126 (5), 283–287.
- Shahin, M., Van Oorscot, H.J.L. and De Lange, S.J., (1993). *Statistical Analysis in Water Resources Engineering*. A.A. Balkema, Rotterdam, the Netherlands, 394 pp.
- Westmacott, J. R. & Burn, D. H. (1997) Climate change effects on the hydrologic regime within the Churchill-Nelson River Basin. *J. Hydrol.* 202 (1-4), 263–279.
- Zhang, X., Harvey, K. D., Hogg, W. D. & Yuzyk, T. R. (2001) Trends in Canadian streamflow. *Water Resour. Res.* 37 (4), 987–998.

CAPÍTULO**DOS**

REDES NEURONALES ARTIFICIALES (RNAs)

En este capítulo se exploran a RNAs a través de la historia así como las bases teóricas. Esta metodología computacional es usada ampliamente en diversos campos de investigación debido a su capacidad como aproximadores universales (Cybenko, 1989, Hornik et al., 1989). En recientes décadas han tenido una gran aceptación por los investigadores, además del éxito que han obtenido en la aplicación en diversas áreas; a pesar de eso RNAs han sido empleadas muy poco en el campo de la hidrología. Las RNAs son capaces de resolver problemas altamente no lineales, esto tiene una valiosa aplicación en la hidrología ya que la mayoría de los procesos que se llevan a cabo en la naturaleza son altamente no lineales. Finalmente también en este capítulo se estudia una de las clases de RNAs conocidas como perceptron multicapa o red multicapa con conexiones hacia adelante (multilayer feefforward networks) que sirve de base en la metodología propuesta en esta tesis.

2.1 Introducción

Las RNAs nacieron como un intento de imitar el funcionamiento del cerebro humano. RNAs consisten en un gran número de elementos de procesamiento simples llamados neuronas, donde cada nodo se conecta con otros por medio de enlaces directos. Cada enlace está asociado con un peso que representa la fuerza de la señal de salida. El procesamiento de cada nodo se lleva a cabo en dos etapas, es decir, la suma ponderada de las entradas multiplicadas por sus respectivos pesos y por la aplicación de la función de activación. RNAs se caracterizan por (1) patrones de conexiones entre las neuronas (arquitectura), (2) métodos de determinación de los pesos en las conexiones (entrenamiento o algoritmo de aprendizaje), (3) función de activación, y (4) su número de capas: una sola capa (Redes de Hopfield); bicapa (Carpintero / Grossberg redes de resonancia adaptativa); y multicapa (la mayoría de las redes de retropropagación). Si todas las señales fluyen solo en una dirección entonces la RNA es llamada Feed-Forward Network (FFN). De lo

contrario, se les llama Recurrents Networks. El modelo de RNAs más comúnmente utilizado en hidrología es Feed forward-Multilayer Perceptrons (FFMLPs), que por lo general contiene tres capas y emplean funciones de activación monótonas en todas las capas. Sus pesos se obtienen generalmente por entrenamiento supervisado mediante el algoritmo de retropropagación.

2.2 Historia de las redes neuronales

Las redes neuronales se originan a finales del siglo XIX cuando diversos científicos estudiaban la actividad cerebral y trataban de entender el funcionamiento de la mente, más tarde en el siglo XX en la década de los 40 es cuando se realizan las primeras investigaciones centradas en redes neuronales artificiales (RNAs) incrementando cada vez más sus aplicaciones así como el desarrollo de nuevos modelos de RNAs con diversas aplicaciones hasta la actualidad. Diversos investigadores han contribuido para lograr el desarrollo de las RNAs. Los principales aportadores al desarrollo de las RNAs son Warren McCulloch y Walter Pitts (McCulloch & Pitts, 1943) fueron quienes implementaron el primer modelo matemático de RNAs basado específicamente en impulsos binarios. Luego en 1949, Hebb (Hebb, 1949) propuso la regla Hebbiana que describe como el aprendizaje afecta la sinapsis entre dos neuronas. Para 1952 basándose en las propiedades físicas de las membranas celulares y las corrientes de iones a través de proteínas transmembranas Hodgkin y Huxley (Hodgkin & Huxley, 1952) incorporaron los fenómenos neuronales así como la descarga neuronal y propagación potencial de una serie de ecuaciones de evolución produciendo picos cuantitativamente precisos. Los modelos binarios de Warren McCulloch y Walter Pitts fueron utilizados por las RNAs Hopfield (Hopfield, 1982) y la memoria asociativa bidireccional discreta (Kosko, 1988). Posteriormente Donald Hebb desarrolló un procedimiento matemático de aprendizaje (Hebb, 1949). Los primeros resultados prácticos en RNAs se obtuvieron en 1951 mediante las aportaciones de Marvin Minsky (Minsky, 1954). Las primeras aplicaciones atmosféricas (Uttley, 1966) y reconocimiento de patrones (Uttley, 1975) fueron gracias a Albert Uttley (Uttley, 1956a; Uttley, 1956b) quien desarrolló nuevos paradigmas en RNAs. Ya en 1957 Frank Rosenblatt le añadió aprendizaje al pionero modelo introducido por McCulloch y Pitts (Rosenblatt, 1957; Rosenblatt, 1958) dando como resultado al modelo PERCEPTRON; sin embargo no encontró un modelo matemático para entrenar un modelo de tres capas. Más tarde Bernard Widrow (Widrow, 1959, Widrow, 1960) diseñó una red similar al perceptron llamada ADALINE así como su versión de múltiples capas llamada MADALINE estos modelos fueron utilizados con éxito en muchos problemas; sin embargo no pueden resolver problemas linealmente separables debido a sus función de activación de tipo lineal. En 1960 Winrow y Marcian Hoff (Winrow & Hoff, 1960) probaron matemáticamente que el error de las RNAs puede ser minimizado hasta límites muy pequeños. En 1961 en las investigaciones tempranas del desarrollo de métodos de codificación de información en RNAs uno de los primeros investigadores ha sido Steinbuch (Steinbuch, 1961) introduciendo las redes Steinbuch que se aplican al reconocimiento de escritura, mecanismos de diagnóstico de fallos de máquinas y control de múltiples procesos en producción (Steinbuch & Piske, 1963). Sin embargo en 1969, Minsky y Papert (Minsky & Papert, 1969) demostraron matemáticamente que el perceptron no se puede utilizar para funciones lógicas complejas. Esto trajo una considerable disminución en el interés de las

RNAs. El interés en RNAs fue recuperado gracias a las investigaciones de Grossberg (Grossberg, 1964), incluyendo estrictos análisis matemáticos permitiendo nuevos paradigmas en RNAs y el análisis de todas las facetas del procesamiento de información humana. Estas investigaciones cubren todo el espectro de RNAs. Con ello se abrieron paso nuevas investigaciones como las de Shun-Ichi Amari combinando RNAs con rigurosos modelos matemáticos (Amari, 1971; Amari, 1972; Amari, 1974; Amari, 1977 y Amari, 1983). Así como también las investigaciones del departamento de máquinas inteligentes de la Universidad de Edimburgo quienes en 1968 descubrieron las relaciones entre hologramas y memorias asociativas. Con dicha relación Longuet-Higgins creó un sistema de ecuaciones codificadas para almacenar y recuperar una ausencia de señales (Longuet-Higgins, 1968), más tarde con la ayuda de Willshaw y Buneman se introdujeron los principios holográficos (Willshaw et al., 1969, Willshaw & Longuet-Higgins, 1969). De este trabajo surge un modelo de RNAs temporal llamado HOLOPHONE (Willshaw & Longuet-Higgins, 1970). Ya en la década de los 70 y 80 se llevan a cabo trabajos pioneros en el aprendizaje competitivo y la autoorganización inspirada a partir de los patrones de conexión encontradas en la corteza visual. Fukushima propuso su cognitron (Fukushima, 1975) y modelos necognitron (Fukushima, 1980) bajo el paradigma de aprendizaje competitivo. También se tienen los trabajos de Kohonen con investigaciones en RNAs sobre paradigmas de conexiones aleatorias centrándose en memorias asociativas y matrices de correlación (Kohonen, 1972; Kohonen, 1974). En este mismo campo se tienen los trabajos de Anderson (Anderson, 1968; Anderson, 1970) así como Steinbush y Piske (Steinbush, 1961; Steinbush & Piske, 1963). En 1982, John Hopfield introdujo un método de análisis del estado estable de una red asociativa mediante una función de energía en sus estudios sobre sistemas de ecuaciones no lineales (Hopfield, 1982), demostrando que se puede construir una ecuación de energía que describa la actividad de una red monocapa en tiempo discreto y dicha ecuación puede disiparse y el sistema converger a un mínimo local. Más tarde Hopfield extendió su modelo para considerar tiempos continuos (Hopfield, 1984). El modelo de Hopfield introducido en 1982 marcó el comienzo de la era moderna de la investigación de RNAs. Este modelo trabajo con un nivel de sistema en lugar de un solo nivel de neurona. Este modelo se denomina red neuronal recurrente y trabaja con la regla Hebbiana, dicha red puede ser utilizada como memoria asociativa para el almacenamiento de información y para resolver problemas de optimización.

La máquina de Boltzmann (Ackley et al., 1985) se introdujo en 1985 como una extensión de la red de Hopfield mediante la incorporación de neuronas estocásticas. El aprendizaje Boltzmann se basa en un método llamado hibridación simulada (Kirkpatrick et al., 1983). Ya en 1987, Kosko propuso la memoria asociativa bidireccional adaptativa (BAM) (Kosko, 1987). En 1988, Chua y Yang (Chua & Yang, 1988) extendieron el modelo de Hopfield proponiendo el modelo red neuronal celular. La aportación más importante en la investigación de RNAs es el algoritmo de retropropagación (Backpropagation BP) de aprendizaje propuesto para el perceptrón multicapa (MLP) en 1986 por Rumelhart, Hinton, y Williams (Rumelhart et al., 1986). Sin embargo el algoritmo BP ya se había inventado en 1974 por Werbos (Werbos, 1974). Para 1988, Broomhead y Lowe propusieron el modelo de red de función de base radial (RBF) (Broomhead & Lowe, 1988). Tanto MLP y RBF son aproximadores universales. En 1985, Pearl introdujo el modelo de red bayesiana (Pearl, 1988). La red bayesiana es el modelo gráfico más conocido en IA. Posee la característica de ser tanto un estadístico y un formalismo conocimiento-representación.

Otro hito en el aprendizaje de máquinas y de las comunidades de redes neuronales es la máquina de soporte vectorial (SVM), propuesto por Vapnik a principios de 1990 (Vapnik, 1988). SVM se basa en la teoría del aprendizaje estadístico y es particularmente útil para la clasificación con muestras pequeñas. SVM se ha utilizado para la clasificación, regresión, y la agrupación. Gracias a su aplicación con éxito en la SVM, los métodos kernel han despertado un gran interés.

2.3 Aplicación de RNAs para predicción y modelado en hidrología

Los modelos hidrológicos son aproximaciones de la naturaleza, debido a que esta es compleja los modelos no pueden predecir perfectamente eventos en un sistema natural. Los modelos hidrológicos son herramientas valiosas en aplicaciones que van desde: predicción en tiempo real de eventos hidrológicos a políticas y estructuras para la mitigación de eventos hidrológicos extremos, como inundaciones y sequías (Gupta et al., 1999).

La aplicación de RNAs en la predicción y modelado hidrológico han sido diversas; en el análisis de aguas residuales para pronosticar volúmenes de lodos (Capodaglio et al, 1991). Zhang & Yamada (Zhang & Yamada, 1993.) utilizó RNAs para el pronóstico de la demanda diaria de agua y lo comparó con los métodos convencionales y encontraron que RNAs tienen una mayor ventaja para resolver problemas no lineales; Predicción del escurrimiento máximo y del tiempo de pico resultantes del patrón de lluvias (Smith & Eli, 1995).

Experimentos recientes han reportado que RNAs puede ofrecer una alternativa prometedora para la predicción de escurrimiento. Zhu y Fujita (Zhu y Fujita, 1993) utilizaron RNAs para pronosticar el escurrimiento de 1 a 3 horas, ellos encontraron que el modelo on-line funcionó muy bien, sin embargo el modelo off-líne no pudo predecir con exactitud el escurrimiento; Karunanithi (Karunanithi et al, 1994) usó un modelo MLP para predecir el escurrimiento. Ellos compararon los resultados obtenidos con el modelo MLP con la del modelo no lineal de capacidad analítica y encontraron que el modelo de RNAs presentó una mejor precisión que el segundo modelo. (Wei et al, 2002) propuso la utilización de RNAs para resolver los problemas poco estructurados de pronóstico de inundaciones. Ellos usaron una zona inundada de China durante el período de 1949 a 1994 como demostración; Filho y dos Santos (Filho & dos Santos, 2006) aplicaron un modelo a través de RNAs para modelado de escurrimientos en una cuenca densamente urbanizada. Estudiaron la cuenca del río Tamanduateí, un tributario de la cuenca del río Alto Tiete en el estado de Sao Paulo, Brasil, y encontraron que RNAs mejoran su rendimiento en un 40% con entradas conjuntas de lluvias y escurrimiento; Sahoo y Ray (Sahoo & Ray, 2006) describen la aplicación de un feed -forward back-propagation con base radial RN para pronosticar el escurrimiento en un arroyo en Hawaii propenso a inundaciones. El método tradicional de estimación de escurrimientos en este cuerpo de agua era mediante el uso de las curvas de calibración convencionales para flujos de Hawaii; Zealand y Burn (Zealand & Burn, 1999) investigaron la utilidad del modelo MLP para pronóstico de escurrimientos a corto plazo. Ellos exploraron las capacidades de MLP y compararon el rendimiento de esta herramienta con los enfoques convencionales utilizados para pronóstico de

escurrimientos. Llegaron a la conclusión de que RNAs superaron constantemente un modelo convencional durante la fase de verificación (prueba).

2.4 Perceptron multicapa

2.4.1 Estructura de la red

La arquitectura de RNAs más comúnmente utilizada para pronostico ha sido la RNAs perceptron de alimentación hacia adelante multicapa (Multilayer Feedforward Neural Network), por comodidad se adoptará la abreviación en inglés MFNN. Esta arquitectura MFNN es una generalización del perceptron simple y surgió como consecuencia de las limitaciones de dicha arquitectura en la separación no lineal. La aceptación de este tipo de arquitectura es debido a la capacidad teórica de aproximar complicadas funciones no lineales a precisiones arbitrarias (Hornik et al., 1989), es así que MFNNs son una nueva clase de funciones para aproximar o interpolar relaciones no lineales. La arquitectura de MFNN se caracteriza por que tiene muchas neuronas agrupadas en capas diferentes niveles. Cada una de las capas está formada por un conjunto de neuronas y se distinguen tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida como se observa en la Figura 2.1. En la práctica solo tres capas son necesarias, esta tesis se centra solo en MFNN de tres capas.

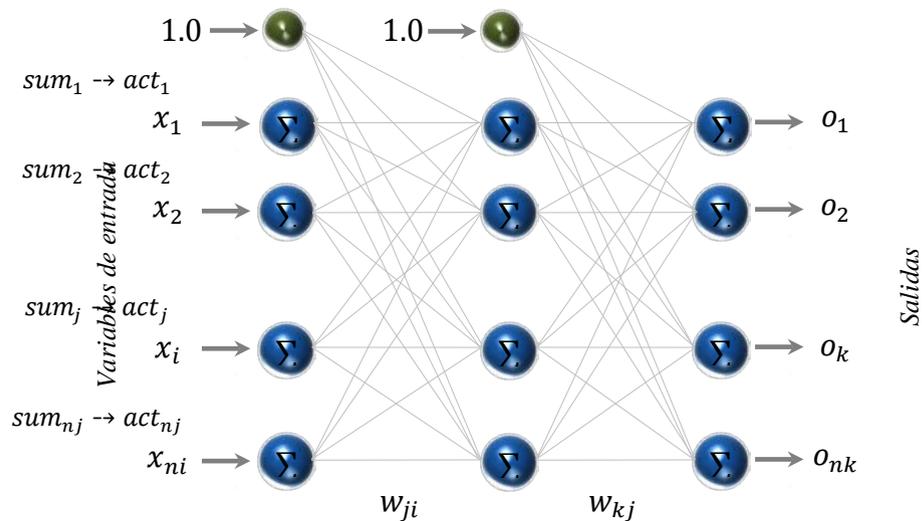


Figura 2.1. Arquitectura de MFNN típica con el algoritmo de aprendizaje de retropropagación.

El valor recibido por cada neurona de la capa de entrada es multiplicado por un peso asociado.

$$sum_j = \sum_{i=1}^n w_{ji} x_i + w_{jb} \quad (2.1)$$

$$act_j = f_j(sum_j) \quad (2.2)$$

La suma de cada entrada sum_j por su respectivo peso es evaluada mediante una función de activación act_j . Generalmente en MFNN se utilizan dos funciones de transferencia o activación usadas en el algoritmo de aprendizaje de retropropagación:

Función sigmoideal

$$f_j(sum_j) = \frac{1}{1 + e^{-sum_j}} \quad \forall sum_j \in [-\infty, \infty] \wedge f_j(sum_j) \in (0, 1) \quad (2.3)$$

Función tangente hiperbólica

$$f_j(sum_j) = \frac{e^{sum_j} - e^{-sum_j}}{e^{sum_j} + e^{-sum_j}} \quad \forall sum_j \in [-\infty, \infty] \wedge f_j(sum_j) \in (-1, 1) \quad (2.4)$$

Cada una de las neuronas son propagadas hacia adelante hasta llegar a la capa de salida.

$$sum_k = \sum_{j=1}^m w_{kj} act_j + w_{kb} \quad (2.5)$$

$$act_k = f_k(sum_k) \quad (2.6)$$

Los pesos en la red se inician de forma aleatoria, para entrenar la red se necesita comparar la salida calculada con la salida deseada, es así que se necesita una función de comparación. Hasta ahora el mejor es una modificación de error cuadrado medio.

$$e_p = \frac{1}{2} \sum_{k=1}^L (d_{pk} - o_{pk})^2 \quad (2.7)$$

El error e_p para un patrón p es calculado como la diferencia al cuadrado entre los valores de salidas dados por el modelo o_{pk} y los deseados o reales d_{pk} . Esto es por cada elemento en el vector de salida.

2.5 Algoritmos de entrenamiento de MFNN

En este estudio se utilizaron cinco algoritmos de entrenamiento, Quasi-Newton Backpropagation BFGS (QNBFGS), Scaled Conjugate Gradient Backpropagation (SCGB), One Step Secante Backpropagation (OSSB), Resilient Backpropagation (RB) y finalmente Levenberg Marquardt (LM) en el cual se pone mayor énfasis debido a que ha sido el mejor modelo de optimización encontrado en este trabajo. En los siguientes párrafos se explica el proceso de entrenamiento de las RNAs mediante el algoritmo de retropropagación, siendo este el enfoque original de RNAs después mediante el algoritmo de Levenberg-Marquardt los demás algoritmos se pueden encontrar en el Apéndice B.

El objetivo principal de estos algoritmos es reducir la función error E que para este caso se define en la ecuación 2.8, en la figura 2.2 se observa con el punto en el centro el óptimo global mientras que alrededor se tienen mínimos locales.

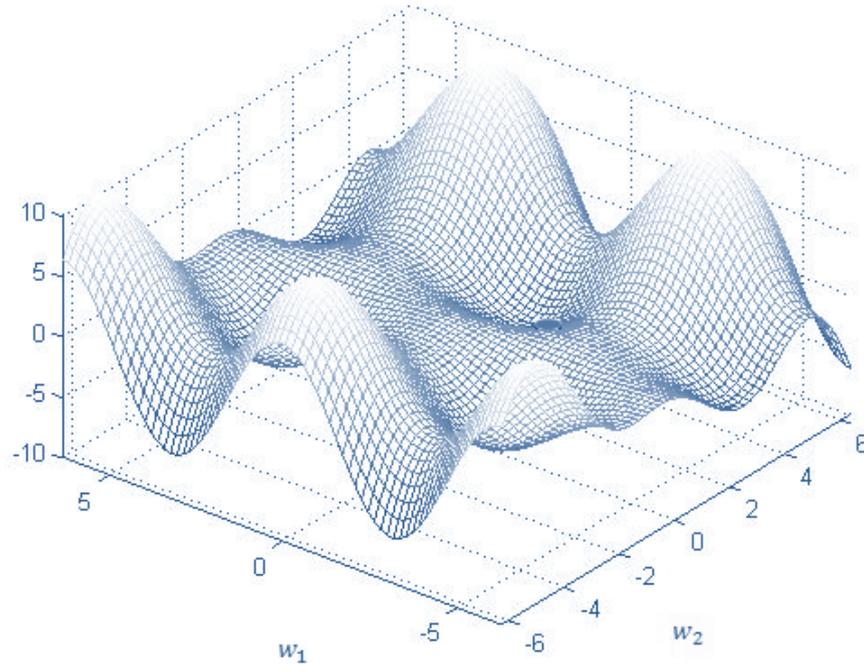


Figura 2.2. Superficie de una función error (E) en dos parámetros

$$E = \sum_{p=1}^P (d_p - o_p)^2 \quad (2.8)$$

Donde d_p es la salida deseada, o_p es la salida estimada, P es el número de datos de la serie analizada.

2.5.1 Algoritmo de retropropagación

El objetivo del entrenamiento de MFNN es minimizar el error de la red. Dentro de los algoritmos disponibles el algoritmo de retropropagación (backpropagation algorithm, BPA) (Rumelhart et al., 1986), también conocido como regla delta generalizada, podría decirse que es el más ampliamente usado. Ya que el objetivo del entrenamiento es minimizar el error con respecto a los pesos en las conexiones, se debe derivar el error e_p en la ecuación (2.9).

$$\frac{\partial e_p}{\partial w_{kj}} = \frac{\partial e_p}{\partial sum_{pk}} \frac{\partial sum_{pk}}{\partial w_{kj}} \quad (2.9)$$

Resolviendo para cada componente de forma individual en la ecuación (2.9) se tiene.

$$\frac{\partial sum_{pk}}{\partial w_{kj}} = \frac{\partial}{\partial w_{kj}} \sum_k w_{kj} act_{pj} = act_{pj} \quad (2.10)$$

Si se introduce una nueva variable $\delta_{pk} = \frac{\partial e_p}{\partial sum_{pk}}$ en 2.10 se tiene que.

$$\frac{\partial e_p}{\partial w_{kj}} = \delta_{pk} act_{pj} \Rightarrow \frac{\partial e_p}{\partial sum_{pk}} = \frac{\partial e_p}{\partial o_{pk}} \frac{\partial o_{pk}}{\partial sum_{pk}} \quad (2.11)$$

Donde cada componente de la ecuación (2.11) puede ser resuelto como sigue

$$\frac{\partial o_{pk}}{\partial sum_{pk}} = f'_k(sum_{pk}) \text{ y } \frac{\partial e_p}{\partial o_{pk}} = (d_{pk} - o_{pk}) \quad (2.12)$$

Sustituyendo las ecuaciones (2.12) en la ecuación (2.11) se tiene

$$\delta_{pk} = (d_{pk} - o_{pk}) f'_k(sum_{pk}) \quad (2.13)$$

Finalmente los cambios de los pesos entre la capa oculta y la capa de salida basadas en la minimización del error en la ecuación 2.8 son encontrados sustituyendo la ecuación 2.13 en

$$\delta_{pk} = \frac{\partial e_p}{\partial sum_{pk}}.$$

$$\Delta w_{kj} = \frac{\partial e_p}{\partial w_{kj}} = (d_{pk} - o_{pk}) f'_k(sum_{pk}) act_{pj} \quad (2.14)$$

En el caso de la capa oculta y la capa de entrada. Sin embargo se debe conocer el error atribuido a cada neurona en la capa oculta, entonces se necesita expresar la relación entre los valores calculados para cada neurona de salida y la activación de cada neurona en la capa oculta, esto es.

$$\frac{\partial e_p}{\partial w_{ji}} = \frac{1}{2} \sum_k \frac{\partial}{\partial w_{ji}} (d_{pk} - o_{pk})^2 = \sum_k (d_{pk} - o_{pk}) \frac{\partial o_{pk}}{\partial sum_{pk}} \frac{\partial sum_{pk}}{\partial act_{pj}} \frac{\partial act_{pj}}{\partial sum_{pj}} \frac{\partial sum_{pj}}{\partial w_{ji}} \quad (2.15)$$

Resolviendo cada componente y sustituyendo

$$\frac{\partial o_{pk}}{\partial sum_{pk}} = f_k(sum_{pk}), \quad \frac{\partial sum_{pk}}{\partial act_{pj}} = \frac{\partial}{\partial act_{pj}} \sum_k w_{kj} act_{pj} = w_{kj} \quad (2.16)$$

$$\frac{\partial act_{pj}}{\partial sum_{pj}} = f'_j(sum_{pj}), \quad \frac{\partial sum_{pj}}{\partial w_{ji}} = \frac{\partial \left(\sum w_{ji} x_{pi} \right)}{\partial w_{ji}} = x_{pi} \quad (2.17)$$

Sustituyendo las ecuaciones 2.16 en 2.17 dentro de la ecuación 2.15 resulta

$$\frac{\partial e_p}{\partial w_{ji}} = (d_{pk} - o_{pk}) f'_k(\text{sum}_{pk}) \text{act}_{pj} w_{kj} f'_j(\text{sum}_{pj}) x_{pi} \quad (2.18)$$

Simplificando la ecuación se obtiene

$$\frac{\partial e_p}{\partial w_{ji}} = \left(f'_j(\text{sum}_{pj}) \sum_k \delta_{pk} w_{kj} \right) x_{pi} \quad (2.19)$$

Simplificando la ecuación 2.19 con $\delta_{pj} = f'_j(\text{sum}_{pj}) \sum_k \delta_{pk} w_{kj}$ entonces el cambio de los pesos de las conexiones entre la capa oculta y la capa de entrada resulta.

$$\frac{\partial e_p}{\partial w_{ji}} = \Delta w_{ji} = \delta_{pj} x_{pi} \quad (2.20)$$

Ahora que ya se conocen los errores se necesita un método de aprendizaje para cambiar los pesos en cierta proporción (η) al error. Así los nuevos valores de la capa de salida y la capa oculta y los de la capa oculta y la capa de entrada son.

$$w_{kj}^{\text{nuevos}} = w_{kj}^{\text{anterior}} + \eta \delta_{pk} \text{act}_{pj} \quad \text{y} \quad w_{ji}^{\text{nuevos}} = w_{ji}^{\text{anterior}} + \eta \delta_{pj} x_{pi} \quad (2.21)$$

Después de cambiar los pesos se vuelven a aplicar las entradas y se vuelve a repetir el proceso. Cada ciclo que se repite constituye una iteración de la red. Alternativamente los valores del error se pueden resumir a lo largo de una época de entrenamiento. Una época es igual a la exposición de cada muestra entrenada, es decir una época es un paso por cada vez que la serie es entrenada. La red deja de aprender cuando el error se reduce hasta cierto umbral especificado o cuando se llega hasta cierto número de iteraciones previamente especificadas. Una vez que la red ha aprendido lo suficiente se prueba el rendimiento de la red mediante el proceso de validación, el cual consiste en presentar un nuevo conjunto de datos a la red previamente entrenada y calcular sus salidas. Este proceso se usa para probar que la red puede resolver el problema para el cual fue entrenada.

2.5.2 Entrenamiento de MFNN mediante el algoritmo de Levenberg-Marquardt

El algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) fue desarrollado independientemente por Kenneth Levenberg y Donald Marquardt, este algoritmo proporciona una solución numérica para optimización de funciones no lineales. Una de las ventajas de este algoritmo es que es rápido y tiene una convergencia estable. La lenta convergencia del método del descenso más rápido es mejorada en gran medida por el algoritmo de Gauss-Newton (Osborne, 1992). Usando las derivadas de segundo orden de la función error, el algoritmo de Gauss-Newton encuentra tamaño de pasos adecuados para cada dirección lo que le permite converger muy rápido, especialmente si la función error tiene una superficie cuadrática. Esto solo ocurre cuando la aproximación de la función

cuadrática es razonable. De lo contrario el algoritmo de Gauss-Newton podría ser más divergente.

El algoritmo de Levenberg-Marquardt combina el método de descenso más rápido y el algoritmo de Gauss-Newton. Con ellos se logra la ventaja de velocidad del algoritmo de Gauss-Newton y la estabilidad del método del descenso más rápido. En este algoritmo no importa lo compleja que es la función error. Aunque el algoritmo de Levenberg-Marquardt tiende a ser un poco más lento que el algoritmo de Gauss-Newton (en convergencia), converge mucho más rápido que el método descenso más rápido. La idea básica del algoritmo de Levenberg-Marquardt es realizar un proceso de entrenamiento combinado: alrededor de las áreas con curvaturas complejas, el algoritmo de Levenberg-Marquardt se mantiene en el algoritmo de descenso más rápido hasta que la curvatura local es adecuada para hacer una aproximación cuadrática; entonces se convierte aproximadamente en el algoritmo de Gauss-Newton, que puede acelerar la convergencia significativamente.

La derivación del algoritmo de Levenberg-Marquardt puede ser presentado en cuatro partes: 1) algoritmo del descenso rápido, 2) método de Newton, 3) algoritmo de Gauss-Newton y 4) algoritmo de Levenberg-Marquardt.

Los índices usados son:

p es el índice del patrón, que va desde 1 a P donde P es el número de patrones.

M es el índice de las salidas y va desde 1 a M donde M de las salidas.

i y j son los índices de los pesos, desde 1 a N donde N es el número total de pesos.

k es el índice de las iteraciones.

$$E(x, w) = \frac{1}{2} \sum_{p=1}^P \sum_{m=1}^M e_{p,m}^2 \quad (2.22)$$

donde

x es el vector de entrada

w es el vector de pesos

$e_{p,m}$ es el error de entrenamiento en la salida m aplicado al patrón p y es definido como

$$e_{p,m} = d_{p,m} - o_{p,m} \quad (2.23)$$

donde

d es el vector de salidas deseadas

o es el vector de salidas actual

2.5.2.1 Algoritmo del descenso más rápido

El algoritmo del descenso más rápido es un algoritmo de primer orden. Este usa las derivadas de primer orden de la función total de error para encontrar el mínimo error en el

espacio. Normalmente el gradiente \mathbf{g} es definido como la derivada de primer orden de la función error.

$$\mathbf{g} = \frac{\partial E(x, \mathbf{w})}{\partial \mathbf{w}} = \left[\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \frac{\partial E}{\partial w_3}, \dots, \frac{\partial E}{\partial w_{N-1}}, \frac{\partial E}{\partial w_N} \right]^T \quad (2.24)$$

La regla de actualización de pesos del algoritmo del descenso más rápido puede ser escrita como sigue.

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha \mathbf{g}_k \quad (2.25)$$

Donde α es la constante de aprendizaje, el proceso de entrenamiento del algoritmo del descenso más rápido es de convergencia asintótica. Todos los elementos del vector gradiente alrededor de la solución podrían ser muy pequeños y por lo tanto el cambio de pesos sería muy pequeño.

2.5.2.2 Algoritmo de Newton

El método de Newton asume que todos los componentes de los gradientes $g_1, g_2, g_3 \dots \dots g_N$ son funciones de pesos.

$$\begin{aligned} g_1 &= F_1(w_1, w_2, w_3 \dots \dots w_N) \\ g_2 &= F_2(w_1, w_2, w_3 \dots \dots w_N) \\ g_3 &= F_3(w_1, w_2, w_3 \dots \dots w_N) \\ &\dots \dots \dots \\ g_N &= F_N(w_1, w_2, w_3 \dots \dots w_N) \end{aligned} \quad (2.26)$$

Donde $F_1, F_2, F_3 \dots \dots F_N$ son relaciones no lineales entre pesos y componentes del gradiente. Al desarrollar cada $g_i (i = 1, 2, \dots, N)$ en las ecuaciones anteriores mediante series de Taylor y calcular la aproximación de primer orden se tiene:

$$\begin{aligned} g_1 &\approx g_{1,0} + \frac{\partial g_1}{\partial w_1} \Delta w_1 + \frac{\partial g_1}{\partial w_2} \Delta w_2 + \frac{\partial g_1}{\partial w_3} \Delta w_3 + \dots + \frac{\partial g_1}{\partial w_N} \Delta w_N \\ g_2 &\approx g_{2,0} + \frac{\partial g_2}{\partial w_1} \Delta w_1 + \frac{\partial g_2}{\partial w_2} \Delta w_2 + \frac{\partial g_2}{\partial w_3} \Delta w_3 + \dots + \frac{\partial g_2}{\partial w_N} \Delta w_N \\ &\dots \dots \dots \\ g_N &\approx g_{N,0} + \frac{\partial g_N}{\partial w_1} \Delta w_1 + \frac{\partial g_N}{\partial w_2} \Delta w_2 + \frac{\partial g_N}{\partial w_3} \Delta w_3 + \dots + \frac{\partial g_N}{\partial w_N} \Delta w_N \end{aligned} \quad (2.27)$$

Mediante la combinación del vector gradiente \mathbf{g} en 1 se puede determinar que

$$\frac{\partial g_i}{\partial w_j} = \frac{\partial \left(\frac{\partial E}{\partial w_j} \right)}{\partial w_j} = \frac{\partial^2 E}{\partial w_i \partial w_j} \quad (2.28)$$

Reemplazando la ecuación 2.28 en 2.27 se tiene:

$$\begin{aligned} g_1 &\approx g_{1,0} + \frac{\partial^2 E}{\partial w_1^2} \Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ g_2 &\approx g_{2,0} + \frac{\partial^2 E}{\partial w_2 \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_2^2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ &\dots \dots \dots \\ g_N &\approx g_{N,0} + \frac{\partial^2 E}{\partial w_N \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_N \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_N^2} \Delta w_N \end{aligned} \quad (2.29)$$

Comparando con el método del descenso más rápido, las derivadas de segundo orden de la función error necesitan ser calculadas para cada componente del vector gradiente. Para obtener el mínimo de la función E , entonces cada elemento del vector gradiente debe de ser cero, entonces el lado izquierdo de la ecuación 2.29 se vuelve cero.

$$\begin{aligned} 0 &\approx g_{1,0} + \frac{\partial^2 E}{\partial w_1^2} \Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ 0 &\approx g_{2,0} + \frac{\partial^2 E}{\partial w_2 \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_2^2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ &\dots \dots \dots \\ 0 &\approx g_{N,0} + \frac{\partial^2 E}{\partial w_N \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_N \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_N^2} \Delta w_N \end{aligned} \quad (2.30)$$

Mediante la combinación de la ecuación 2.30 y 2.26 se tiene.

$$\begin{aligned}
-\frac{\partial E}{\partial w_1} = g_{1,0} &\approx \frac{\partial^2 E}{\partial w_1^2} \Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\
-\frac{\partial E}{\partial w_2} = g_{2,0} &\approx \frac{\partial^2 E}{\partial w_2 \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_2^2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\
&\dots \dots \dots \\
-\frac{\partial E}{\partial w_N} = -g_{N,0} &\approx \frac{\partial^2 E}{\partial w_N \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_N \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_N^2} \Delta w_N
\end{aligned} \tag{2.31}$$

Entonces se tienen N ecuaciones con N incógnitas con lo cual los Δw_n pueden ser calculados mediante un proceso iterativo.

La ecuación 2.30 puede ser escrita en forma de matriz.

$$\begin{bmatrix} -g_1 \\ -g_2 \\ \dots \dots \dots \\ -g_N \end{bmatrix} = \begin{bmatrix} -\frac{\partial E}{\partial w_1} \\ -\frac{\partial E}{\partial w_2} \\ \dots \dots \dots \\ -\frac{\partial E}{\partial w_N} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} \Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_2^2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ \dots \dots \dots \\ \frac{\partial^2 E}{\partial w_N \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_N \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_N^2} \Delta w_N \end{bmatrix} \begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \dots \dots \dots \\ \Delta w_N \end{bmatrix} \tag{2.32}$$

De la ecuación anterior se puede observar que se tiene la matriz Hessiana.

$$H = \begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} \Delta w_1 + \frac{\partial^2 E}{\partial w_1 \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ \frac{\partial^2 E}{\partial w_2 \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_2^2} \Delta w_2 + \frac{\partial^2 E}{\partial w_2 \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_2 \partial w_N} \Delta w_N \\ \dots \dots \dots \\ \frac{\partial^2 E}{\partial w_N \partial w_1} \Delta w_1 + \frac{\partial^2 E}{\partial w_N \partial w_2} \Delta w_2 + \frac{\partial^2 E}{\partial w_N \partial w_3} \Delta w_3 + \dots + \frac{\partial^2 E}{\partial w_N^2} \Delta w_N \end{bmatrix} \tag{2.33}$$

La ecuación 2.24 puede ser escrita en forma vectorial como sigue

$$-g = H \Delta w \Rightarrow \Delta w = -H^{-1} g \tag{2.34}$$

La actualización de los pesos por el método de Newton es

$$\mathbf{w}_{k+1} = \mathbf{w}_k - H_k^{-1} \mathbf{g}_k \tag{2.35}$$

2.5.2.3 Algoritmo de Gauss-Newton

Si se aplica el método de Newton para la actualización de los pesos, a fin de obtener la matriz Hessiana. Se calculan las derivadas de segundo orden de la función error, sin embargo esto puede ser muy complicado. Para simplificar este proceso se introduce la matriz Jacobiana (J) la cual está definida como sigue:

$$J = \begin{bmatrix} \frac{\partial e_{1,1}}{\partial w_1} & \frac{\partial e_{1,1}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{1,1}}{\partial w_N} \\ \frac{\partial e_{1,2}}{\partial w_1} & \frac{\partial e_{1,2}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{1,2}}{\partial w_N} \\ \frac{\partial e_{1,3}}{\partial w_1} & \frac{\partial e_{1,3}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{1,3}}{\partial w_N} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial e_{p-2,1}}{\partial w_1} & \frac{\partial e_{p-2,1}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{p-2,1}}{\partial w_N} \\ \frac{\partial e_{p-1,2}}{\partial w_1} & \frac{\partial e_{p-1,2}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{p-1,2}}{\partial w_N} \\ \frac{\partial e_{p,M}}{\partial w_1} & \frac{\partial e_{p,M}}{\partial w_2} & \dots & \dots & \dots & \frac{\partial e_{p,M}}{\partial w_N} \end{bmatrix} \quad (2.36)$$

Integrando las ecuaciones 2.22 y 2.24, los elementos del vector gradiente pueden ser calculados como.

$$g_i = \frac{\partial E}{\partial w_i} = \frac{\frac{1}{2} \sum_{p=1}^P \sum_{m=1}^M e_{p,m}^2}{w_i} = \sum_{p=1}^P \sum_{m=1}^M \left(\frac{\partial e_{p,m}}{\partial w_i} e_{p,m} \right) \quad (2.37)$$

Combinando las ecuaciones 2.36 y 2.37, la relación entre la matriz jacobiana J y el vector gradiente g es.

$$g = J e \quad (2.38)$$

Donde el vector error e tiene la siguiente forma.

$$e = \begin{bmatrix} e_{1,1} \\ e_{1,2} \\ \dots \\ e_{1,M} \\ \dots \\ e_{p,1} \\ e_{p,2} \\ \dots \\ e_{P,M} \end{bmatrix} \quad (2.39)$$

Sustituyendo la ecuación 2.22 en 2.33, los elementos en la i th fila y j th columna de la matriz Hessiana puede ser escrita de la siguiente forma:

$$h_{i,j} = \frac{\partial^2 E}{\partial w_i \partial w_j} = \frac{\partial^2 \left(\frac{1}{2} \sum_{p=1}^P \sum_{m=1}^M e_{p,m}^2 \right)}{\partial w_i \partial w_j} = \sum_{p=1}^P \sum_{m=1}^M \left(\frac{\partial e_{p,m}}{\partial w_i} \frac{\partial e_{p,m}}{\partial w_j} + S_{i,j} \right) \quad (2.40)$$

Donde $S_{i,j}$ es igual a.

$$S_{i,j} = \sum_{p=1}^P \sum_{m=1}^M \frac{\partial^2 e_{p,m}}{\partial w_i \partial w_j} e_{p,m} \quad (2.41)$$

Debido a que el supuesto del método de Newton es que si $S_{i,j}$ esta cercano a cero (Hagan & Menhaj, 1994), la relación entre la matriz Hessiana (H) y la matriz Jacobiana (J) se puede reescribir como.

$$H \approx J^T J \quad (2.42)$$

Combinando la ecuación 2.35, 2.38 y 2.42 la regla de actualización del algoritmo Gauss-Newton es presentada de la siguiente forma.

$$w_{k+1} = w_k - (J_k^T J_k)^{-1} J_k e_k \quad (2.43)$$

La ventaja del algoritmo de Gauss-Newton sobre el método de Newton, es que no requiere el cálculo de las derivadas de segundo orden del error total. Mediante la introducción de la matriz Jacobiana (J). Sin embargo el algoritmo de Gauss-Newton aún enfrenta el mismo problema de convergencia como el algoritmo de Newton para la optimización en espacios de error complejos. Matemáticamente el problema es que no se puede invertir la matriz $J^T J$.

2.5.2.4 Algoritmo de Levenberg-Marquardt

Con el fin de asegurar la inversión de la matriz Hessiana (H), el algoritmo de Levenberg-Marquardt introduce otra aproximación de la matriz Hessiana (H).

$$H \approx J^T J + \mu I \quad (2.44)$$

Donde μ es siempre positivo, llamado coeficiente de combinación. I Es la matriz identidad.

De la ecuación 2.44 se puede observar que los elementos en la diagonal principal de la aproximación de la matriz Hessiana (H) siempre serán mayores que cero. Por lo tanto con esta aproximación (ecuación 2.44) se puede asegurar que la matriz H es siempre invertible. Mediante la combinación de las ecuaciones 2.43 y 2.44 la regla de actualización de pesos del algoritmo de Levenberg-Marquardt se puede escribir como sigue.

$$w_{k+1} = w_k - (J^T J + \mu I)^{-1} J_k e_k \quad (2.45)$$

Debido a que este algoritmo nace de la combinación del algoritmo del descenso más rápido y del algoritmo de Gauss-Newton, el algoritmo de Levenberg-Marquardt cambia entre los dos algoritmos durante el proceso de iteración. Cuando el coeficiente de combinación μ es muy pequeño (casi cero), la ecuación 2.45 se está acercando a la Ecuación 2.43 y se utiliza el algoritmo de Gauss-Newton. Cuando coeficiente de combinación μ es muy grande, la ecuación 2.45 se aproxima a la Ecuación 2.35 y se utiliza el método del descenso más rápido.

2.6 Referencias

- Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science*, 9(2), 147–169.
- Amari, S. (1983). Field theory of self-organizing neural nets. *IEEE Transaction on Systems, Man and Cybernetics*, SMC-13 (9 & 10), 741–748.
- Amari, S. (1974). A method of statistical neurodynamics. *Kybernetik* 14, 201—215.
- Amari, S. 1972. Characteristics of random nets of analog neuron-like elements. *IEEE Trans. on Systems, Man, and Cybernetics* SMC-2, 643—657.
- Amari, S. (1971). Characteristics of randomly connected threshold element networks and network systems. *Proc. IEEE* 59, 35~47.
- Amari, S. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. To be published.
- Anderson, J. A. (1968). A memory storage model utilizing spatial correlation functions. *Kybernetik* 5, 113-119.
- Anderson, J.A. (1970). Two models for memory organization using interacting traces. *Mathematical Biosciences*, 8,137-160.
- Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321–355.
- Capodaglio, A. G., Jones, H. V., Novotny, V., & Feng, X. (1991). Sludge Bulking analysis and forecasting: Application of system identification and artificial neural computing technology *Water Research*, 25(10), 1217-1224.
- Chua, L. O., & Yang, L. (1988). Cellular neural network: I. Theory; II. Applications. *IEEE Transactions on Circuits and Systems*, 35(10), 1257–1290.

- Cybenko, G. (1989). "Approximation by Superpositions of a Sigmoidal Function," *Math. Control Signals Systems*, 2, 303-314.
- Filho, A. J. P., & dos Santos, C. C. (2006). Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data. *Journal of Hydrology*, 317(1-2), 31-48.
- Fukushima, K. (1975). Cognition: A self-organizing multilayered neural network. *Biological Cybernetics*, 20, 121-136.
- Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4), 193-202.
- Grossberg, S. (1964). *The theory of embedding fields with applications to psychology and neurophysiology*. Rockefeller Institute for Medical Research.
- Gupta, H.V., Sorooshian, S., Yapo, P.O. (1999). Status of automatic calibration for hydrologic models: comparison with multilevel expert calibration. *J. Hydrol. Eng.* 4 (2), 135-143.
- Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of ion currents and its applications to conductance and excitation in nerve membranes. *Journal of Physics*, 117, 500-544.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent computational abilities. *Proceedings of the National Academy of Sciences*, 79:2554.
- Hornik, K., Stinchcombe, M., and White, H. (1989). "Multilayer Feedforward Networks are Universal Approximators," *Neural Networks*, 2(5), 359-366.
- Karunanithi N, Grenney WJ, Whitley D, Bovee K. (1994). Neural networks for river flow prediction. *Journal Computing in Civil Engineering ASCE* 8(2): 201-220.
- Kirkpatrick, S., Gelatt, C. D, Jr, & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.
- Kohonen, T. (1972). Correlation associative memories, *IEEE Transactions on Computers*, C-21, 4, 353-359.
- Kohonen, T. (1974). An adaptive associative memory principle. *IEEE Transactions on Computers*, C-23:444-445.
- Kosko, B. (1987). Adaptive bidirectional associative memories. *Applied Optics*, 26, 4947-4960.
- Kosko, B. (1988). Bidirectional associative memories. *IEEE Transactions on Systems, Man and Cybernetics*, 18, 49-60.
- Levenberg, K. (1944). A method for the solution of certain problems in least squares, *Quarterly of Applied Mathematics*, 5, 164-168, 1944.
- Lonquet-Higgins, H.C. (1968). Holographic Model of Temporal recall. *Nature (London)* 217, 104.
- M. T. Hagan and M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, 5(6), 989-993, 1994.
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters, *SIAM Journal on Applied Mathematics*, 11(2), 431-441, June 1963.
- McCulloch, W. S. and Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:115-133.
- Minsky, M. (1954). *Theory of Neural-Analog Reinforcement System and its Applications to the Brain-Model Problem* (Ph. D. Thesis). University of Princeton, Princeton, NJ.
- Minsky, M. L., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Osborne, M. R. (1992). Fisher's method of scoring, *International Statistical Review*, 86, 271-286.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
- Rosenblatt F. (1957) *The perceptron: a perceiving and recognizing automaton* Technical Report 85-460-1 Cornell Aeronautical Laboratory
- Rosenblatt F. (1958). *The perceptron: A theory of statistical separability in cognitive systems*. Technical Report VG-1196-G-1, Cornell Aeronautical Laboratory.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*, 1: Foundation (pp. 318-362). Cambridge: MIT Press.
- Sahoo GB, Ray C. (2006). Flow forecasting for a Hawaii stream using rating curves and neural networks. *Journal of Hydrology* 317:63-80.
- Smith J, Eli RN. (1995). Neural-network models of rainfall-runoff process. *Journal Water Resource Planning and Management, ASCE* 121(6): 499-508.

- Steinbuch, K. (1961). Die lernmatrix. *Kybernetic*, 1:36-45
- Steinbuch, K. and U.A.W. Piske (1963), "Learning Matrices and Their Applications", *IEEE Transactions on Electronic Computers*, EC-12:846-862.
- Uttley A. M. (1966). The transmission of information and the effect of local feedback in theoretical and neural networks. *Brain Research* 102 (1966), 23–35
- Uttley, A.M. (1956a). Conditional probability machines and conditioned reflexes. In C.E. Shannon and J. McCarthy (Eds.), *Automata Studies*. Princeton, N.J.: Princeton Univer. Press.
- Uttley, A.M. (1956b). Temporal and spatial patterns in a conditional probability machine. In C.E. Shannon and J. McCarthy (Eds.), *Automata Studies*. Princeton, N.J.: Princeton Univer. Press.
- Uttley, A.M. (1975). The Informon in classical conditioning. *J.Theoret. Biol.*49, 355-376.
- Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.
- Wei, Y., Xu, W., Fan, Y., & Tasi, H.-T. (2002). Artificial neural network based predictive method for flood disaster. *Computers & Industrial Engineering*, 42(204), 383-390.
- Werbos, P. J. (1974). *Beyond regressions: newtools for prediction and analysis in the behavioral sciences*. PhD Thesis, Harvard University, Cambridge, MA.
- Widrow B. and Hoff M. (1960): Adaptive switching circuits, *IRE Western Electric Show and Convention Record*, Part 4, pp.96-104.
- Widrow B. (1959). "Adaptive Sampled-Data Systems --- A Statistical Theory of Adaptation," *IRE WESCON Convention Record*, 4:74-85, 1959.
- Widrow B. (1960). "An Adaptive 'Adaline' Neuron Using Chemical 'Memistors'," *Stanford Electronics Laboratories Technical Report 1553-2*, October 1960.
- Willshaw, D J, Buneman, O P and Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature*, 222: 960-962.
- Willshaw, D., & Longuet-Higgins, H. (1970). Associative memory models. *Machine Intelligence*, 351.
- Willshaw, D.J., Longuet-Higgins, H.C. (1969). The holophone recent developments. In: Metzler, B., Michie, D. (Eds.): *Machine intelligence*, Vol. 4, pp. 349--357. Edinburgh: University Press.
- Zealand CM, Burn DH, Simonovic SP. (1999). Short term streamflow forecasting using artificial neural networks *Journal of Hydrology* 214:32–48.
- Zhang, S. P., Watanabe, H., & Yamada, R. (1993). Prediction of Daily Water Demands by Neural Networks. Paper presented at the *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, University of Waterloo, Waterloo, Ontario, Canada.
- Zhu, M., & Fujita, M. (1993). Application of Neural Networks to Runoff Forecast. Paper presented at the *Stochastic and Statistical Methods in Hydrology and Environmental Engineering*, University of Waterloo, Waterloo, Ontario, Canada.

CAPÍTULO**TRES**

MODELOS ENSAMBLADOS

Recientemente los modelos ensamblados ENN (Ensemble Neural Network) se han convertido en un tema de investigación activa en todo el mundo por diversos investigadores ya que se ha demostrado que la capacidad de generalización de RNAs se puede mejorar mediante la combinación de varias RNAs en ensamblajes. Los modelos ensamblados mediante RNAs se pueden definir como un conjunto de modelos entrenados de forma independiente como independientes RNAs cuyas salidas o predicciones se combinan mediante un modelo para obtener una única estimación de la salida deseada. Esto permite una mayor estabilidad y generalización del modelo. Otra ventaja es también que se puede utilizar con mayor eficacia los datos de entrenamiento para este tipo de modelos que los de interrupción o temprano detenimiento. En este capítulo se aplican los métodos más populares de ensamblaje. Se utilizan siete enfoques diferentes para la creación de modelos de ensamblaje para mejorar el pronóstico mediante un proceso de regionalización.

3.1 Introducción

Uno de los modelos de predicción del cual se ha obtenido una fructífera aplicación en recientes décadas han sido RNAs, sin embargo una de las preocupaciones en RNAs es su capacidad de generalización la cual es reflejada en la precisión de la predicción de los datos que se utilizan como validación. Las RNAs son un enfoque relativamente nuevo para el modelado de problemas que involucran una alta no linealidad consiguiendo con ello numerosas aplicaciones en muchos campos científicos. RNAs también han sido ampliamente utilizados para resolver diversos problemas hidrológicos, como por ejemplo el modelado de lluvia-escorrentamiento, la predicción de escurrimientos, modelación de la calidad del agua, modelación de aguas subterráneas, la modelización hidrológica de series de tiempo, y la operación de los embalses (ASCE, 2000). Govindaraju y Rao (2000)

demuestran las diversas aplicaciones en hidrología y documentan una gran gama de arquitecturas y los diversos algoritmos de entrenamiento usados en RNAs.

En estudios recientes los múltiples sistemas clasificadores también llamados sistemas de ensamblaje o ensamble han llamado la atención dentro de la comunidad de inteligencia y aprendizaje vía máquina computacional. Este crecimiento de atención en dichos modelos ensamblados ha sido debido a que han demostrado ser muy eficientes además de versátiles en una gran variedad de problemas en diferentes áreas científicas. En RNAs se ha demostrado que la capacidad de generalización mejora mediante los modelos de ensamblaje donde cada componente del modelo es utilizado para dar la misma solución a pesar que se podría obtener esta solución por diferentes métodos (Sharkey, 1999).

El enfoque de modelo de ensamblaje en RNAs es un número finito de RNAs individuales que se entrenan con un mismo propósito cuyas predicciones se combinan para generar una salida única. Los modelos ensamblados mediante RNAs ofrecen un número de ventajas sobre los modelos con una RNAs. Una ventaja muy importante es que tienen el potencial para mejorar la generalización y aumento de la estabilidad (Sharkey, 1999). Los modelos ensamblados han sido aplicados con éxito en diversos campos, tales como la predicción de series temporales, la robótica, y el diagnóstico médico, entre otros. Los modelos ensamblados originalmente fueron desarrollados para disminuir la varianza y de esta forma mejorar la predicción. Los modelos ensamblados ya se han utilizado con éxito en una gran variedad de problemas de aprendizaje de máquinas, tales como la selección de características, la estimación de la confianza, completar características faltantes, aprendizaje gradual, corrección de errores, clasificación de datos, aprendizaje de distribuciones no estacionarias, entre otros. Los modelos ensamblados imitan nuestra naturaleza en la búsqueda de varias opiniones para obtener una decisión crucial. A menudo se piensa en opiniones individuales y se combinan para llegar a decisión final (Polikar, 2006).

En general los *modelos ensamblados* se refieren a componentes basados en un mismo modelo básico. Sin embargo, en esta tesis se utiliza también modelos híbridos que no son de la misma familia. Este último también se conoce en la literatura como "múltiples sistemas clasificadores". La aplicación exitosa de los modelos de ensamble se puede encontrar en muchos campos, tales como: finanzas (Leigh et al., 2002), bioinformática (Tan et al., 2003), medicina (Mangiameli et al., 2004), quimioinformática (Merkwirth et al., 2004), fabricación (Maimon & Rokach 2004), geografía (Bruzzone et al., 2004), y Recuperación de Imágenes (Lin et al., 2006). La idea de construir un modelo predictivo que integra múltiples modelos se ha investigado durante mucho tiempo. La historia de los modelos ensamblados se remonta a 1977 con Tukey Twicing (Tukey, 1977): un ensamble de dos modelos de regresión lineal. Tukey sugirió un ajuste del primer modelo a los datos originales y el segundo modelo a los residuales. Dos años más tarde, Dasarthy y Sheela (1979) sugirieron particionar el espacio de entrada con dos o más clasificadores. Los principales avances en este campo se lograron durante los años noventa. Hansen y Salamon (1990) sugirieron un ensamble de RNAs configurados de forma similar para mejorar el rendimiento predictivo en vez de una sola RNAs. En los últimos años, los estudios

experimentales realizados por el aprendizaje de máquinas muestran que la combinación de las salidas de múltiples clasificadores reduce el error de generalización (Domingos, 1996; Quinlan, 1996; Bauer & Kohavi, 1999; Opitz & Maclin, 1999) de los clasificadores individuales. Los modelos de ensamble son muy eficaces debido principalmente al fenómeno que se presenta en los diversos tipos de clasificadores, ya que tienen diferentes "sesgos inductivos" (Mitchell, 1997). Los modelos de ensamble aprovechan dicha diversidad para reducir la varianza-del error (Tumer & Ghosh, 1996; Ali & Pazzani, 1996) sin aumentar el sesgo de error.

3.2 Redes ensambladas

En una red ensamblada, los módulos estarán formados por las redes neuronales individuales. Entonces un modelo de red ensamblado es entrenado en cada uno de sus miembros individuales y son combinados usando varios métodos. Sin embargo independientemente del método de combinación que se utilice, la regla de combinación es independiente de los datos de entrada. Una red ensamblada tiene una arquitectura muy flexible puesto que cada miembro clasificador puede ser desarrollado adoptando la forma que quiera sin embargo el resultado final de todos los clasificadores es afectado por el algoritmo de combinación. En este trabajo se ha utilizado el perceptron multicapa como miembro clasificador y los algoritmos de combinación AdaBoost.M2, LPBoost, TotalBoost y RUSBoost.

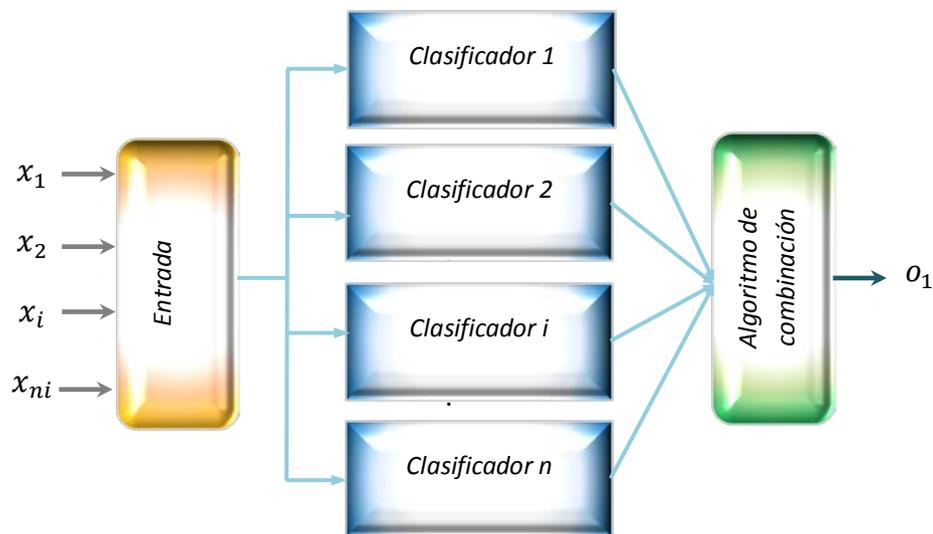


Figura 3.1. Estructura de la red ensamblada

3.2 Algoritmo AdaBoost.M2

El algoritmo *Adaptive Boosting* fue presentado para transformar una regla de aprendizaje débil en una regla de aprendizaje fuerte (Schapire, 1990). Este es como sigue; dada una serie de pares de datos para entrenamiento (x_i, y_i) donde $x_i \in X$ denota los patrones de

entrada y $y_i \in Y = \{1, \dots, k\}$ los valores correctos de una serie de k variables. En la clasificación multiclase $k > 2$, el algoritmo AdaBoost.M2 (Freund & Shapire, 1997) es la versión más comúnmente usado este es aplicable cuando el clasificador define un mapeo $h_t: X \times Y \rightarrow [0,1]$, que puede considerarse por ejemplo, como proporcionar una puntuación de confianza para cada variable. El algoritmo básico se describe como sigue.

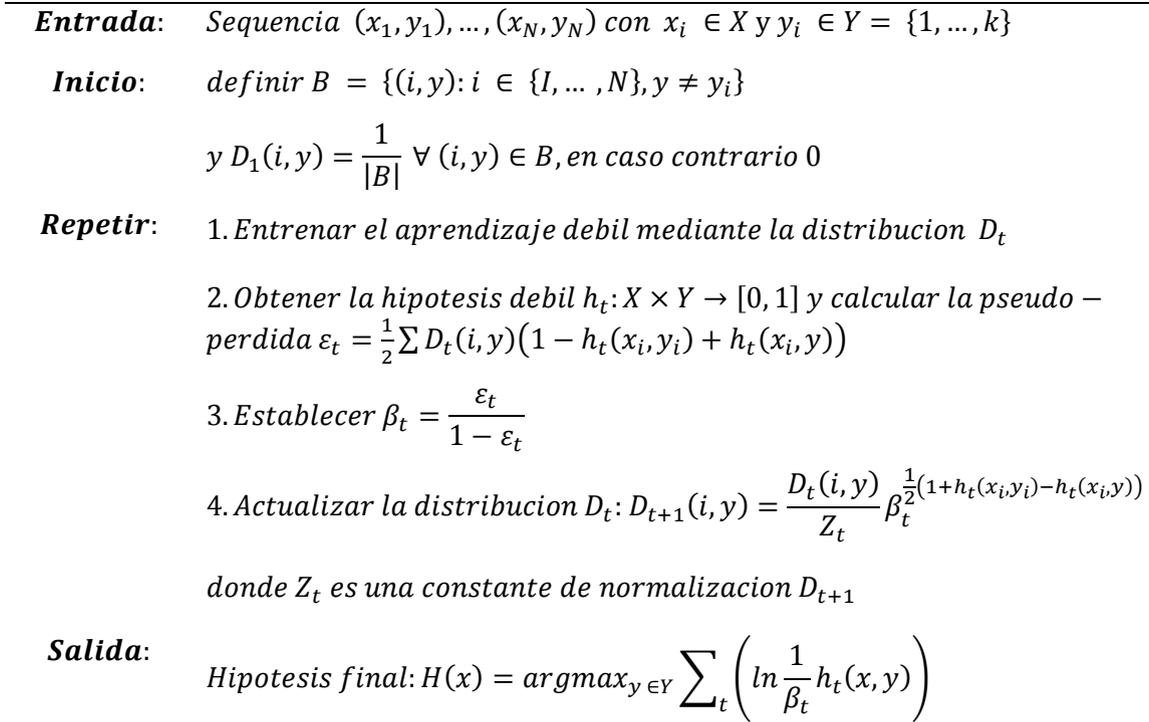


Figura 3.2. Esquema del algoritmo AdaBoost.M2.

El algoritmo AdaBoost mantiene una distribución $D_t(i, y)$ sobre los pares de variables (x_i, y) (siendo 0 para la correcta variable y_i ; $D_t(x_i, y_i) = 0$). Esta distribución es actualizada de acuerdo a las salidas del clasificador actual en la serie de entrenamiento: los pesos para las variables que pertenecen a los ejemplos clasificados con confianza se reducen con respecto a los de los ejemplos mal clasificados La regla de actualización está diseñada para garantizar cotas superiores en el entrenamiento y la generalización del paso de aprendizaje (Freund & Schapire, 1997). En cada iteración t , un nuevo clasificador es entrenado con respecto a la distribución D_t . En reconocimiento, las puntuaciones de los clasificadores individuales son combinados linealmente, con pesos siendo inversamente relacionados al error de entrenamiento de los clasificadores, para dar la salida final.

3.3 Algoritmo LPBoost

El algoritmo *Linear Programming Boosting* maximiza el margen entre muestras de entrenamiento de diferentes clases; de esta manera, este algoritmo pertenece a la clase de algoritmos de supervisado margen de maximización. El objetivo es la construcción de una función de aprendizaje en el espacio de las variables que minimiza el error en la clasificación y maximiza el margen. Este algoritmo es formulado como un programa lineal

que puede ser resuelto de manera eficiente usando técnicas de generación de columna que han sido desarrollados para problemas de optimización de gran escala. A diferencia de los algoritmos gradiente boosting que pueden converger en sólo en el límite el algoritmo LPBoost converge en un número finito de iteraciones en una sola solución global, esto lo hace computacionalmente competitivo con el algoritmo AdaBoost. Los resultados empíricos muestran que LPBoost converge rápidamente y a menudo es más rápido que otras formulaciones.

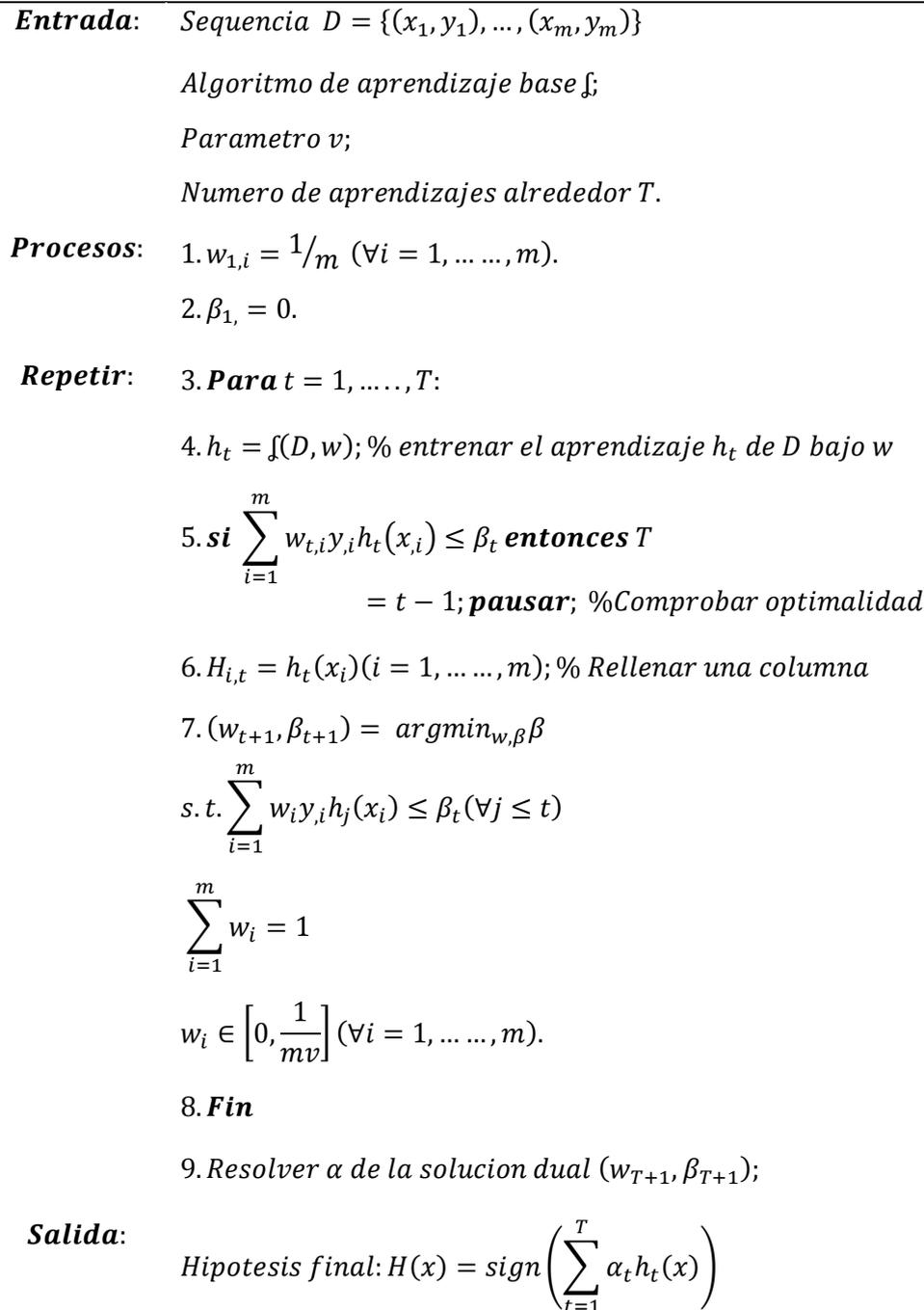


Figura 3.3. Esquema del algoritmo LPBoost.

3.4 Algoritmo TotalBoost

El algoritmo *Total Boosting* es un algoritmo correctivo. TotalBoost es un algoritmo de margen duro (Warmuth et al., 2006). La actualización de los pesos en este algoritmo están motivados por el principio de mínima entropía relativa de Jaynes: entre las soluciones que satisfacen algunas restricciones lineales se escoge la que minimiza una entropía relativa en relación con la distribución inicial d^1 . Además entre todas las distribuciones d que son soluciones para $d^t \in \operatorname{argmin} \max \mathbf{u}^t \cdot \mathbf{d}$ con $d \in S^N$ $t=1, \dots, T$ donde el vector u^t es definido como $y^n h^t(x)$. Esto es para TotalBoost, el algoritmo escoge la distribución que minimiza $\Delta(d, d^1)$. Esto asegura que los problemas de optimización de TotalBoost tienen soluciones únicas. TotalBoost resuelve los siguientes problemas de optimización. TotalBoost resuelve el siguiente problema de optimización.

$$d^{t+1} = \min_d \Delta(d, d^1)$$

$$s. t. \mathbf{d} \cdot \mathbf{u}^m \leq \hat{y}^t - \varepsilon, \text{ para } 1 \leq m \leq t; \mathbf{d} \in P^N$$

$$\text{donde } \hat{y}^t = \min_{m=1, \dots, t} \mathbf{d}^m \cdot \mathbf{u}^m$$

Para resolver el problema de optimización anterior se usa programación cuadrática secuencial, esto se describe en TotalBoost (Warmuth et al., 2006). Eliminando la entropía relativa del objetivo se llega al problema de optimización de LPBoost.

Entrada: Secuencia $(x_1, y_1), \dots, (x_N, y_N)$ con parametros de precisión $\varepsilon < 0$.

Inicio: d^1 a la distribución uniforme y \hat{y}_1 a 1

Repetir: Para $t = 1, \dots$

Enviar d^t al aprendizaje debil y obtener la hipotesis debil h_t

Poner $\mathbf{u}_n^t = h_t(x_n) y_n$.

Poner $\hat{y}_{t+1} = \min(\hat{y}_t \mathbf{d}^t \cdot \mathbf{u}^t)$.

Actualizar la distribución a

$$d^{t+1} = \operatorname{argmin}_d \Delta(d, d^1) s. t. \mathbf{d} \cdot \mathbf{u}^m \leq \hat{y}^t - \varepsilon, \text{ para } 1 \leq m \leq t; \mathbf{d} \in P^N$$

Si el problema arriba es factible o d^{t+1} contiene un cero entonces $T = t$ y se pausa.

Salida: Hipotesis final: $f_\alpha(x)$

$$= \sum_{t=1}^T \alpha_t h_t(x), \text{ donde los coeficientes } \alpha_t \text{ maximizan el margen}$$

duro bajo la hipotesis usando el problema LP (linear programming).

Figura 3.4. Esquema del algoritmo TotalBoost.

3.5 Algoritmo RUSBoost

El algoritmo *Random Undersampling Boosting* está compuesto por RUS (random undersampling) y Boost se basa en el algoritmo SMOTEBoost (Chawla et al., 2003) y este algoritmo a su vez es basado en el algoritmo AdaBoost.M2 (Freund & Schapire, 1996). El algoritmo SMOTEBoost tiene una mejora sobre el algoritmo AdaBoost mediante la introducción de una técnica de sobremuestreo inteligente (SMOTE(Chawla et al., 2002)) que ayuda a equilibrar la distribución de clases mientras que el algoritmo AdaBoost mejora el rendimiento clasificador utilizando estos datos balanceados. El algoritmo RUSBoost logra el mismo objetivo pero usando RUS en lugar de SMOTE. El resultado de ello es un algoritmo simple con tiempos de entrenamiento del modelo rápido y un desempeño favorable. Dejando a x_i ser un punto en el espacio de características X y y_i sea una variable clase en una serie de variables clases Y . Cada uno de los m ejemplos en la serie de datos S puede ser representado por las duplas (x_i, y_i) . Permitiendo a t ser una iteración entre uno y el máximo número de iteraciones T (número de clasificadores en el ensemble), h_t es la hipótesis débil entrenada en la iteración t , y $h_t(x_i)$ sea la salida de la hipótesis h_t para el ejemplo x_i . Sea $D_t(i)$ los pesos del ejemplo i en la iteración t .

En el primer paso los pesos de cada ejemplo son inicializados a $1/m$ donde m es el número de ejemplos en la serie de datos entrenados. En el segundo paso las hipótesis débiles T son entrenadas iterativamente, esto se muestra en los pasos a a g . En el paso 2a RUS es aplicado para eliminar los ejemplos de clase mayoritario hasta $N\%$ de la nueva serie de datos entrenados S'_t que pertenece a la clase minoritaria. Por ejemplo si la relación de clase deseada es 50:50 entonces los ejemplos de la clase mayoritaria son removidos aleatoriamente hasta que los números de clases de la mayoría y la minoría son iguales. Como resultado S'_t tendrá una nueva distribución de pesos D'_t . En el paso 2b, S'_t y D'_t son pasados al aprendizaje base *WeakLearn* que crea la hipótesis débil h_t (paso 2c). La pseudoperdida ε_t (basado en el entrenamiento de la serie de datos original S y la distribución de pesos D_t) es calculado en el paso d . En el paso e los parámetros de actualización de los pesos α es calculado como $\varepsilon_t/(1 - \varepsilon_t)$. A continuación la distribución de los pesos para siguiente iteración D_{t+1} es actualizada (paso 2f) y normalizado (paso 2g). Después de T iteraciones del paso 2, la hipótesis final $H(x)$ es devuelta como un voto ponderado de las T hipótesis débiles del paso 3.

Entrada: *Secuencia $(x_1, y_1), \dots, (x_N, y_N)$ con clases minoritarias $y^r \in Y, |Y| = 2$.*

Inicio: *Aprendizaje debil **WeakLearn***

Número de iteraciones T

porcentaje deseado del total de casos representado por la clase minoritaria N

Para las clases minoritarias N

1. Inicializar $D_1(i) = 1/m \forall i$

Repetir: *Para $t = 1, \dots, T$.*

a Crear series de entrenamiento temporal S'_t con distribución D'_t usando submuestreo aleatorio.

*b Llamar **WeakLearn** proporcionando con ejemplos S'_t y sus pesos D'_t .*

c Regresar una hipótesis $h_t: X \times Y \rightarrow [0, 1]$.

d Calcular las pseudopérdidas (para S y D_t):

$$\varepsilon_t = \sum_{(i,y):y_i \neq y} D_t(i)(1 - h_t(x_i, y_i) + h_t(x_i, y)).$$

e Calcular los parametros de la actualización de los pesos

$$\alpha_t = \varepsilon_t / (1 - \varepsilon_t).$$

f Actualizar D_t

$$D_{t+1}(i) = D_t(i) \alpha_t^{\frac{1}{2}(1+h_t(x_i,y_i)-h_t(x_i,y:y \neq y_i))}$$

g Normalizar $D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_t}$

Salida:

Hipotesis final: $H(x) = \operatorname{argmax}_{y \in Y} \sum_{t=1}^T h_t(x, y) \log \frac{1}{\alpha_t}$.

Figura 3.5. Esquema del algoritmo RUSBoost.

3.6 Referencias

- Ali K. M., Pazzani M. J. (1996). Error Reduction through Learning Multiple Descriptions, *Machine Learning*, 24: 3, 173-202.
- ASCE. (2000), Artificial neural networks in hydrology: II. Hydrologic applications, *J. Hydrol. Eng.*, 5(2), 124– 137.
- Bauer, E. and Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, 35: 1-38.
- Bruzzone L., Cossu R., Vernazza G. (2004). Detection of land-cover transitions by combining multiclass classifiers, *Pattern Recognition Letters*, 25(13): 1491-1500.
- Chawla N. V., Lazarevic A., Hall L. O. and Bowyer K. (2003). SMOTEBoost: Improving prediction of the minority class in boosting, in *Proc. Principles Knowl. Discov. Databases*, pp. 107–119.
- Chawla. N. V., Hall L. O., Bowyer K.W., and Kegelmeyer W. P. (2002). SMOTE: Synthetic minority oversampling technique,” *J. Artif. Intell. Res.*, vol. 16, pp. 321–357.
- Dasarathy B.V. and Sheela B.V. (1979). Composite classifier system design: Concepts and methodology, *Proceedings of the IEEE*, vol. 67, no. 5, pp. 708-713.
- Domingos, P. (1996). Using Partitioning to Speed Up Specific-to-General Rule Induction. In *Proceedings of the AAAI-96 Workshop on Integrating Multiple Learned Models*, pp. 29-34, AAAI Press.
- Freund Y. and Schapire R. (1996). Experiments with a new boosting algorithm, in *Proc. 13th Int. Conf. Mach. Learn.*, 1996, pp. 148–156.
- Freund, Y., and R.E. Schapire. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119-139.
- Govindaraju, R. S., and A. R. Rao (Eds.) (2000), *Artificial Neural Networks in Hydrology*, 329 pp., Kluwer Acad., Norwell, Mass.
- Hansen, L. K., and Salamon, P. (1990). Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(10), 993-1001.
- Leigh W., Purvis R., Ragusa J. M. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural networks, and genetic algorithm:a case study in romantic decision support, *Decision Support Systems* 32(4): 361-377.
- Lin H., Kao Y., Yang F., Wang P. (2006). Content-Based Image Retrieval Trained By Adaboost For Mobile Application, *International Journal of Pattern Recognition and Artificial Intelligence*, 20(4):525-541.
- Maimon O., Rokach L. (2004). Ensemble of Decision Trees for Mining Manufacturing Data Sets, *Machine Engineering*, vol. 4 No1-2.
- Mangiameli P., West D., Rampal R. (2004). Model selection for medical diagnosis decision support systems, *Decision Support Systems*, 36(3): 247-259.
- Merkwirth C., Mauser H., Schulz-Gasch T., Roche O., Stahl M., Lengauer T. (2004). Ensemble methods for classification in cheminformatics, *Journal of Chemical Information and Modeling*, 44(6):1971-1978.
- Mitchell, T. (1997). *Machine Learning*, McGraw-Hill.
- Opitz, D. and Maclin, R. (1999). Popular Ensemble Methods: An Empirical Study, *Journal of Artificial Research*, 11: 169-198.
- Polikar R. (2006). Ensemble Based Systems in Decision Making, *IEEE Circuits and Systems Magazine*, vol.6, no. 3, pp. 21-45.
- Quinlan, J. R. (1996). Bagging, Boosting, and C4.5. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 725-730, 1996.
- Schapire, R.E. (1990). The strength of weak learnability. *Machine Learning*, 5:197-227.
- Sharkey, A. J. C. (Ed.) (1999), *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems*, Springer-Verlag, New York.
- Tan A. C., Gilbert D., Deville Y. (2003). Multi-class Protein Fold Classification using a New Ensemble Machine Learning Approach. *Genome Informatics*, 14:206-217.
- Tukey J.W. (1977). *Exploratory data analysis*, Addison-Wesley, Reading, Mass.
- Tumer, K. and Ghosh J. (1996). Error Correlation and Error Reduction in Ensemble Classifiers, *Connection Science*, Special issue on combining artificial neural networks: ensemble approaches, 8 (3-4): 385-404.
- Warmuth, M., Liao, J., and Ratsch, G. (2006). Totally Corrective Boosting Algorithms that Maximizing the Soft Margin. In *Proc. ICML '06*, pages 1001 - 1008. ACM Press, 2006.

CAPÍTULO**CUATRO**

MODELOS MODULARES

En este capítulo se exploran los modelos modulares MNN. El diseño modular de la red neuronal (MNN) se basa en la estructura de la corteza visual en los mamíferos. Las células nerviosas que transmiten señales desde la retina están conectadas a varias redes diferentes en la corteza visual en la parte posterior de la cabeza. Cada red se especializa en una etapa de procesamiento diferente, convirtiendo una secuencia de patrones de puntos de la retina en algo que se puede interpretar como una imagen. Durante el desarrollo de la primera infancia, las células nerviosas en la corteza visual compiten entre sí para ver cuál va a responder a las señales de cada ojo. En una red MNN cada una de las redes compite entre sí aprendiendo del conjunto de entrenamiento. Cada módulo aprende a contribuir en la red como una pieza que forma parte a la solución. Las redes neuronales modulares MNN se conocen como un tipo de máquina comité llamado mezcla de expertos. La red tiene una estructura básica MLP con una capa de entrada y salida, pero una serie de "redes de expertos locales" que residen entre las capas. Estos expertos locales son redes MLP y cada uno es conectado a la capa de entrada. La capa de salida de cada experto está conectada a una "red gating" que determina cuál de los expertos locales produce una estimación más cercana a la salida deseada. Los pesos de las conexiones en la red de expertos se actualizan para reforzar la asociación. Por lo tanto, el MNN es capaz de subdividir los datos de entrenamiento en regiones con patrones similares. Se propone un modelo regional basado en el modelo Modular mediante RNAs para pronóstico de series de escurrimiento y lluvia. El modelo Modular-Regional brinda un prometedor modelo para el pronóstico hidrológico.

4.1 Introducción

En la predicción de los eventos hidrometeorológicos extremos el hombre debe de estar preparado ante eventos como: inundaciones, sequías, olas de calor etc. Dichos eventos exigen la necesidad de buenos modelos de predicción además de la importancia que ello tiene en la gestión de los recursos hídricos. Tanto el escurrimiento como la lluvia así como todos los componentes meteorológicos que se tienen en una cuenca son difíciles de estudiar puesto que no se tiene aún claro la influencia que tiene cada uno de estos componentes por ello es complicado y existe cierto escepticismo en los modelos físicos (Grayson et al., 1992) sin embargo los modelos no lineales pueden involucrar de una forma más aproximada dicho comportamiento y de esta forma poder hacer una mejor aproximación en el pronóstico. Las redes neurales artificiales RNAs con conexiones hacia adelante cuando están completas se les conoce también como redes singulares. Generalmente este tipo de redes son las más usadas y llegan a tener éxito siempre y cuando exista una buena relación entre la entrada y la salida de las variables estudiadas, sin embargo esto no puede ser adecuado cuando se tiene información en el espacio o en el tiempo fragmentado. De acuerdo con Ramachandra (2000) las relaciones funcionales entre la lluvia y el escurrimiento es muy probable que sean diferentes para escurrimientos bajos, medios y altos. La capacidad de infiltración, el almacenamiento local y la cubierta vegetal juegan un papel importante en los escurrimientos bajos, así como el escurrimiento base pues dependen totalmente de ellos. El escurrimiento base puede tener un tiempo de escurrimiento del orden de varios días o meses dependiendo de la cuenca. Por otro lado los eventos correspondientes al escurrimiento alto se generan normalmente por tormentas intensas que sobrepasan completamente la capacidad de infiltración, almacenamiento en depresiones y rugosidad de la superficie. Son bastante efímeras y se alimentan principalmente por el rápido escurrimiento. Los eventos de escurrimientos altos también pueden ser generados por eventos de lluvia de mediana intensidad que se producen en sucesión constante. El primer evento de lluvia puede satisfacer todo el déficit de humedad de la superficie y del subsuelo y el segundo evento puede contribuir en casi toda su totalidad al escurrimiento. Por esta razón RNAs requieren un mayor esfuerzo para estos casos durante su entrenamiento. Para que una RNAs tenga éxito en este tipo de casos debe de tener una serie de datos bastante extensa. Por lo que su proceso de entrenamiento será muy lento ya que la red intenta aprender todos los matices del patrón de entrada tratando de englobarlo en un solo modelo. Debido a este tipo de casos nacen las redes neuronales artificiales modulares RNAM.

4.2 Concepto de modularidad

Los modelos RNAs inspirados en la función del cerebro y más aún en recientes avances en neuropsicología y en neurobiología existen evidencia de modularidad en las funciones cerebrales. Por ejemplo en neuropsicología una lesión cerebral circunscrita podría causar un trastorno de lenguaje sin embargo deja otras funciones cognitivas del cerebro intactas, por otra parte en neurobiología se ha apreciado que las regiones en cerebros de humanos y animales se organizan en módulos especializados y funcionalmente segregadas (Montcastle, 1978; Eccles, 1984 y Edelman, 1987).

En los sistemas computacionales en general se considera que se tiene una arquitectura modular si esta se puede separar o dividir en dos o más subsistemas o módulos en el cual cada subsistema o módulo individual evalúa distintas entradas o las mismas entradas manteniendo comunicación con los otros subsistemas. La salida final del sistema modular depende de la integración de cada una de las salidas de los subsistemas las cuales se combinan mediante un subsistema o módulo llamado red computada.

En un sentido más amplio la modularidad implica que hay una considerable y visible división funcional o estructural entre los diferentes módulos de un sistema computacional. El enfoque de diseño del sistema modular tiene algunas ventajas evidentes, como la sencillez y economía del diseño, la eficiencia computacional, la tolerancia a fallos y una mejor extensibilidad. La ventaja más importante de un sistema modular es su estrecha analogía biológica. Los recientes avances en las ciencias neurobiológicas han fortalecido la creencia en la existencia de la modularidad en los niveles funcionales y estructurales en el cerebro con ello se refuerza el éxito de aprendizaje de este tipo de sistemas modulares.

En un sistema modular una tarea compleja es descompuesta en subtareas que a su vez son realizadas por módulos los cuales se unen para dar una única solución y de esta forma se resuelve la tarea compleja esto hace que el sistema en conjunto sea más fácil de entender modificar y extender (Hrycej, 1992; Gallinari, 1995). Además que el tiempo de entrenamiento del sistema puede reducirse como resultado de la descomposición modular (Pratt et al., 1991). La descomposición de un problema en componentes modulares puede llevarse a cabo de forma automática, explícita o por medio de descomposición en clases (Lu & Ito, 1998). Cuando la descomposición es en módulos explícita, esto por lo general se basa en la comprensión sólida del problema. La división en subtareas es conocida a priori antes del entrenamiento y así se puede mejorar el rendimiento en el aprendizaje. Algunas veces los módulos pueden ser soluciones especializadas a la misma tarea de manera que se obtiene el mejor rendimiento en la tarea cuando el módulo más adecuado dada las circunstancias se ha seleccionado. La descomposición en clases implica dividir un problema en subproblemas basados en relaciones de clases. En la descomposición automática se caracteriza por la partición ciega. Los fines de la descomposición automática son mejorar el rendimiento.

En un sistema modular cada módulo presenta las siguientes características.

1. Los módulos tienen un dominio específico y se han especializado arquitecturas computacionales para reconocer y responder a ciertos subconjuntos o subtareas de la tarea general.
2. Cada módulo es típicamente independiente de otros módulos en su funcionamiento y no influyen ni son influenciados por otros módulos.
3. Los módulos generalmente tienen una arquitectura más simple en comparación al sistema como un todo. Así un módulo puede responder a la entrada dada más rápido que un sistema monolítico complejo.
4. Las respuestas de los módulos individuales son simples y tienen que ser combinados por algún mecanismo de integración con el fin de generar la respuesta del sistema complejo.

En general las ventajas de un sistema modular radican en su capacidad de generalización así como su rendimiento computacional.

4.3 Redes neuronales artificiales modulares RNAM

Son evidentes las ventajas de los modelos modulares especialmente en los procesos cerebrales, esto se ha convertido en la corriente principal en las áreas de investigación de la neurociencia cognitiva. Las técnicas de diseño modular están ganando popularidad. El uso de redes neuronales modulares RNAM para regresión y clasificación se pueden considerar como un modelo competidor a RNAs convencionales, pero con más ventajas. Dos de las ventajas más importantes son una base neurobiológica estrecha y una mayor flexibilidad en el diseño. Otra motivación de las RNAM es ampliar y explotar las capacidades y arquitecturas básicas de las redes neuronales artificiales más comúnmente utilizados que son inherentemente de naturaleza modular. RNAM pueden aplicarse a una gran variedad de problemas complejos como suceden muy a menudo en la naturaleza. En muchos casos RNAM resultan más convenientes de RNAs. Por ejemplo, en el caso de aproximación de funciones. Las funciones continuas a trozos no pueden ser en general modeladas con precisión por las redes neuronales artificiales monolíticas RNAs. Pero, por otro lado, las redes neuronales modulares RNAM han demostrado ser muy eficaces y precisas cuando se utiliza para la aproximación de estos tipos de funciones (Haykins, 1994). Algunas de las principales ventajas de aprender los sistemas modulares son extensibilidad, el aprendizaje gradual, continua adaptación, la economía del aprendizaje, re-aprendizaje, y la eficiencia computacional.

Una definición formal de RNAM es la siguiente (Jordan & Jacobs, 1991a; Osherson et al., 1990; Jordan & Jacobs, 1991b). Una red neuronal se dice que es modular si el cálculo realizado por la red se puede descomponer en dos o más módulos (subsistemas) que operan en entradas distintas sin comunicarse entre sí. Las salidas de los módulos están mediadas por la integración de una unidad que no está permitido para alimentar información de vuelta a los módulos.

4.4 Modelo matemático de una RNAM

El algoritmo general para el modelo Modular ha sido discutido por Jacobs y Jordan (1993), Jordan y Jacobs (1994) y Haykin (1994). Este algoritmo fue adaptado para problemas de lluvia escurrimiento. El diagrama esquemático de una red neuronal artificial modular RNAM se muestra en la figura 3. El Modelo Modular está compuesto de diferentes RNAs feedforward o de alimentación hacia adelante en cada una de ellas da una salida que se ajusta mediante una red especial de tipo compuerta denominada Gating, la cual produce una salida o pesos que ajustan a cada módulo. Esto es la suma ponderada de las respuestas de los módulos forma la salida deL modelo Modular. Es entonces que la salida de la red Gating puede ser considerada como la probabilidad asignada en particular a un módulo, esto es mediante la función softmax implicando que siempre la suma de las salidas de la red Gating siempre sumaran la unidad.

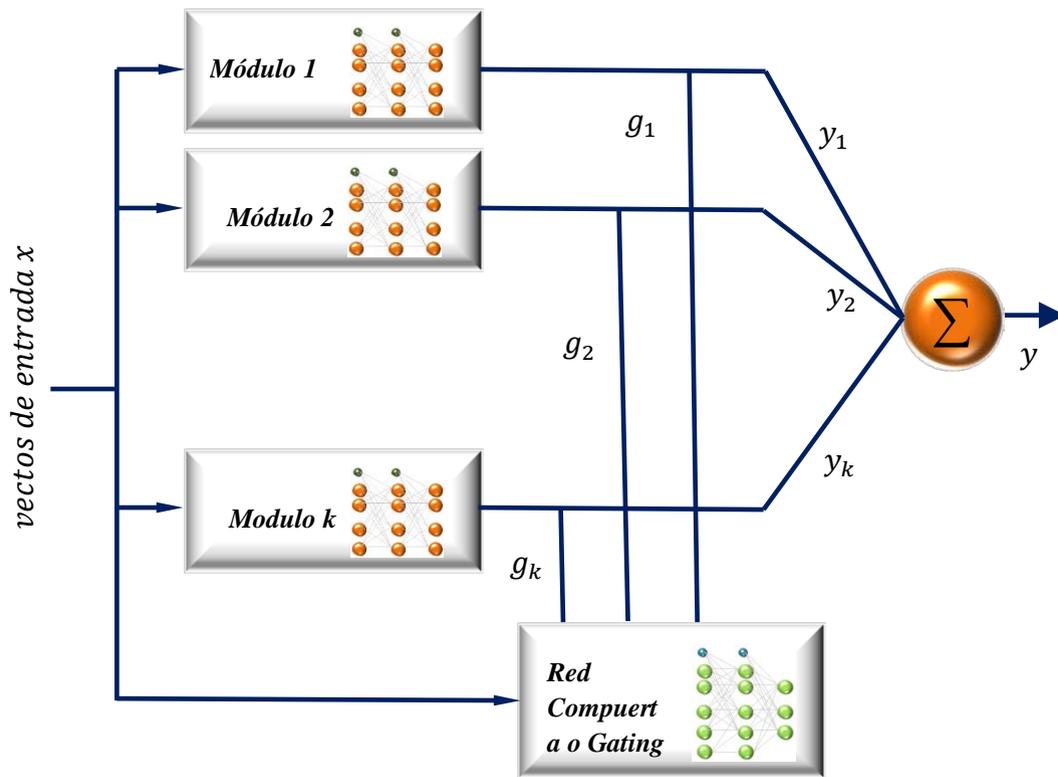


Figura 4.1. Arquitectura del modelo Modular con el algoritmo de aprendizaje de retropropagación.

En general el algoritmo para el modelo Modular fue adaptado para problemas de lluvia-escurrimiento. Cada vector de entrenamiento p -dimensional por x y la respuesta de salida por d denotando la descarga deseada. Cada uno de los vectores de entrada es presentado a la red Gating junto con los k módulos simultáneamente. La salida de cada módulo es representada por y_i entonces la salida de la RNAM es representada por y como sigue.

$$y = \sum_{i=1}^k g_i y_i \quad (4.1)$$

Aquí g_i denota los pesos asignados por la red Gating a las salidas de los módulos, la suma de los productos de las salidas por los pesos dan como resultado la salida del modelo Modular. Dado un vector de entrada x se conceptualiza la respuesta deseada como el resultado de una transformación determinística con ruido adherido entonces esto se puede expresar como sigue.

$$d = F_i(x) + \varepsilon_i, i = 1,2,3, \dots k. \quad (4.2)$$

En la ecuación anterior la forma de la función F no es conocida y ε_i se asume como una variable aleatoria Gaussiana con media cero y varianza σ_i^2 . La distribución de cada vector

de respuesta deseada es una combinación lineal de la distribución Gaussiana k , y se expresa como sigue.

$$f(d - y_i | x) = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^k \frac{g_i}{\sigma_i} \exp \left[-\frac{1}{2\sigma_i^2} (d - y_i)^2 \right] \quad (4.3)$$

El objetivo final del entrenamiento es encontrar los pesos w para todos los nodos de los módulos, una serie de pesos a para la red Gating y un vector σ conteniendo las varianzas σ_i^2 asociadas con cada módulo, estos pueden ser tratados como parámetros desconocidos que serán obtenidos maximizando la función de verosimilitud. La función log-verosímil es como sigue.

$$l(w, a, \sigma) = \ln \sum_{i=1}^k \frac{g_i}{\sigma_i} \exp \left[-\frac{1}{2\sigma_i^2} (d - y_i)^2 \right] \quad (4.4)$$

Maximizando el lado derecho de la ecuación es equivalente a encontrar la máxima verosimilitud y estimar los parámetros w, a y σ . Después que el entrenamiento es llevado a cabo la predicción del modulo i es interpretada como la media condicional dado el vector de entrada. Similarmente las salidas g_i de la red Gating son interpretadas como las probabilidades previas condicionales que el vector de entrada pertenezca a la i -ésimo clases y por lo tanto deberían ser predichas por el modulo i -ésimo. Los valores de σ_i son estimados de lo bien que cada módulo pueda predecir dentro de su propia clase. Hasta ahora no existe restricción alguna en cuanto a la arquitectura de los módulos individuales. Las únicas restricciones que se han impuesto son que la red Gating tiene k nodos en su salida igual al número de módulos del modelo Modular.

$$0 < g_i < 1, \quad i = 1, 2, 3, \dots k. \quad (4.5)$$

$$\sum_{i=1}^k g_i = 1 \quad (4.6)$$

u_i denota la fuerza de la señal de entrada recibida por la i -ésimo neurona en la capa de salida de la red Gating. Para satisfacer las ecuaciones anteriores la función softmax (Bridle, 1990) puede ser usada como salida en la red Gating.

$$g_i = \frac{\exp(u_i)}{\sum_{j=1}^k \exp(u_j)} \quad (4.7)$$

Las probabilidades posteriores dependen del número de módulos en el modelo Modular. En este caso h_i es la probabilidad posterior que el i -ésimo modulo genera como salida, en un sentido bayesiano se puede formular como sigue.

$$h_i = \frac{\frac{g_i}{\sigma_i} \exp \left[-\frac{1}{2\sigma_i^2} (d - y_i)^2 \right]}{\sum_{i=1}^k \frac{g_i}{\sigma_i} \exp \left[-\frac{1}{2\sigma_i^2} (d - y_i)^2 \right]}, i = 1,2,3 \dots k. \quad (4.8)$$

4.5 Referencias

- Bridle John S. (1990). Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems*, number 2, pages 211–217. Morgan Kaufmann, San Mateo.
- Eccles J. C. (1984). The cerebral neocortex: A theory of its operation. In E. G Jones and A. Peters, editors, *Cerebral Cortex: Functional Properties of Cortical Cells*, volume 2. Plenum Press.
- Edelman G. M. (1987). *Neural Darwinism: Theory of Neural Group Selection*. Basic Books.
- Gallinari P. (1995). Modular neural net systems, training of, *The handbook of Brain Theory and Neural Networks*, M. A. Arbib, Ed., Cambridge MA: The MIT Press, pp. 582-585, 1995.
- Govindaraju, R.S., Ramachandra Rao, A. (2000). *Artificial Neural Networks in Hydrology*. Kluwer Academic Publishing, The Netherlands.
- Grayson, RB., I.D. Moore and TA McMahon (1992), Physically based hydrologic modelling 2. Is the concept realistic? *Water Resour. Res.*, 26(10), 2659-2666.
- Haykins Simon. (1984). *Neural Networks, A comprehensive Foundation*. Macmillan College Publishing Company, New York, NY.
- Hrycej T. (1992). *Modular Learning in Neural Networks*. John Wiley, Chichester.
- Jordan M. I. And Jacobs R. A. (1991b). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science*, 15:219–250.
- Jordan M. I. and. Jacobs R. A. (1991a). A competitive modular connectionist architecture. In *Advances in Neural Information Processing Systems 3*, pages 767–773.
- Lu B. and Ito M. (1998). Task decomposition and module combination based on class relations: A modular neural network for pattern classification. Technical Report BMC TR-98-1, Bio-Mimetic Control Research Centre, Nagoya, Japan.
- Mountcastle V. B. (1978). An organizing principle for cerebral function: The unit module and the distributed system. In G. M. Edelman and V. B Mountcastle, editors, *The mindful brain: Cortical organization and the group selective theory of higher brain function*, page 7. MIT Press, Cambridge, MA.
- Osherson, S. Weinstein, and M. Stoli. (1991). Modular learning. *Computational Neuroscience*, pages 369–377, 1990.
- Pratt L.Y., Mostow J., and Kamm C.A. (1991). Direct transfer of learned information among neural networks. In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAI-91)*, pages 584-589. Anaheim, CA.

CAPÍTULO**CINCO**

REGIONALIZACION DE CUENCAS

En este capítulo se exploran diferentes métodos de análisis clúster para la formación de regiones mediante atributos de la misma. El análisis clúster trata de investigar, interpretar y clasificar los datos dados en grupos similares o grupos. Los datos dentro de un grupo deben ser lo más similar posibles y los puntos de los diferentes grupos deben de ser lo más disimiles posibles. En este capítulo, primero se introduce una breve descripción de varios atributos utilizados en la regionalización mediante el análisis clúster. Después de esto, se amplía el análisis clúster. Posteriormente, se describen los pasos para la regionalización. Finalmente la sección concluye con una discusión de los temas en análisis de conglomerados.

5.1 Introducción

Debido a la escasez de datos en muchos sitios es difícil realizar el diseño hidrológico. El análisis regional presenta una alternativa y una mejora en el análisis hidrológico fortaleciendo la información. Un grupo de cuencas que son lo suficientemente homogéneas en los mecanismos de generación de respuestas de flujos constituye una región homogénea. Para identificar regiones homogéneas se utiliza un procedimiento conocido como regionalización.

En el análisis regional los grupos que constituyen regiones homogéneas a menudo están constituidos por cuencas geográficamente contiguas basado en las fronteras políticas, administrativas, o fisiográficas. Sin embargo, esta práctica es criticada porque la delimitación de las regiones que utilizan estos factores no garantiza la homogeneidad hidrológica. En consecuencia, varios enfoques en el campo de la regionalización se han desarrollado que buscan similitud entre sitios mediante la examinación de sus atributos, como características fisiográficas, ubicación geográfica, y estadísticos de las series analizadas.

La regionalización de las cuencas hidrográficas incluyen: (i) el método de residuales (Thomas y Benson, 1970; Wandle, 1977; Glatfelter, 1984; Choquette, 1988); (ii) el análisis de correlación canónica (Ribeiro-Corréa et al, 1995; Cavadias, 1989, 1990; Cavadias et al, 2001); (iii) la región de influencia (región of influence, ROI) enfoque y sus extensiones (Burn, 1990a, b; Zrinji y Burn, 1994; Cunderlik and Burn, 2006a); (iv) el enfoque jerárquico y su extensión al marco ROI (Gabriele y Arnell, 1991; Zrinji y Burn, 1996); y (v) el análisis clúster (Mosley, 1981; Tasker, 1982; Acreman y Sinclair, 1986; Wiltshire, 1986; Bhaskar y O'Connor, 1989; Burn, 1989; Nathan y McMahon, 1990; Hosking y Wallis, 1997; Hall y Minns, 1999; Hall et al., 2002; Jingyi y Hall, 2004; Rao y Srinivas, 2006a, b). Javelle et al. (2002) desarrollando análisis regional frecuencia duración de flujos (FDF) curvas basadas en el método del índice de la avenida (Dalrymple, 1960) para describir el régimen de flujos en una cuenca. Shu y Burn (2004) usó un sistema experto difuso con mejoramiento genético para el análisis de frecuencia regional. Una detallada comparación de algunos de los enfoques utilizados generalmente en la regionalización de las cuencas se encuentra en Cunnane (1988) y GREHYS (1996), mientras que Bobée y Rasmussen (1995) proporcionan una revisión de la literatura relevante.

En el método de los residuales (method of residuals, MOR) el enfoque hacia el análisis de frecuencia regional, las regiones son formadas utilizando los signos positivos y negativos de la residuos extraídos de un modelo de regresión regional, el cual relaciona los cuantiles de flujos en cada estación hidrométrica con las características de las cuencas hidrográficas. Este enfoque de regionalización es ampliamente utilizado por el Servicio Geológico de Estados Unidos (USGS). Este método delimita regiones de inundación de una manera bastante arbitraria y las regiones son con frecuencia arregladas para que coincidan con las fronteras geográficas o límites hidrológicos y áreas políticas o administrativas. Entonces es de esperarse que las regiones delimitadas por este método den como resultado cuencas cuyas características de frecuencia de inundación podrían no ser similares (Wiltshire, 1986; Bhaskar y O'Connor, 1989).

En el análisis de correlación canónica (canonical correlation analysis, CCA) con un enfoque en el análisis regional de frecuencia (Cavadias, 1989, 1990), las cuencas de drenaje están representadas como puntos en espacios de pares de variables canónicas relación-inundaciones no correlacionadas y pares de variables canónicas relación-cuenca no correlacionadas para examinar la similitud en los patrones de puntos correspondientes en estos espacios. Si los patrones de punto son lo suficientemente similares, las regiones se forman en los espacios de las variables canónicas relación-inundación. El enfoque fue originalmente basado en la opinión visual subjetiva de patrones de agrupamiento que pueden estar disponibles. Ribeiro-Corréa et al. (1995) y Cavadias (1995) extensión del enfoque para determinar vecindades hidrológicas homogéneas y fue aplicado a la regionalización de flujos. El problema con este enfoque es que a la regionalización es que la similitud en los patrones de los puntos no puede ser encontrada (Bobée and Rasmussen, 1995).

El enfoque de la región de influencia (región of influence, ROI) (Burn, 1990a, 1990b) permite que cada sitio tenga su propia región. La región de influencia de un sitio en específico consiste en aquellos sitios en la región de estudio cuya distancia al objetivo es un espacio de atributos multidimensionales ponderados que no sobrepasen un umbral elegido.

Para regiones fijas, Gabriele y Arnell (1991) propusieron el enfoque jerárquico para análisis de frecuencia regional que representa explícitamente la variabilidad espacial de las diferentes características de los flujos. Zrinji and Burn (1996) incorporaron el concepto de enfoque jerárquico dentro del marco de región de influencia para definir una serie de regiones de influencia para cada sitio en oposición a una sola región de influencia. Cada enfoque para la regionalización tiene sus beneficios y limitaciones. Sin embargo, debido a las limitaciones por la escasez de datos y la subjetividad en la selección de los atributos, pesos, valores de umbrales y medidas de distancia, no existe un criterio establecido por el cual la superioridad de algún método en particular de regionalización pueda ser claramente establecida.

5.2 Atributos usados en la regionalización

Dentro del análisis *clúster* se forman vectores con similitudes. Generalmente los atributos que se utilizan para la regionalización de las cuencas hidrográficas incluyen: 1) características fisiográficas; 2) características de cubierta; 3) características asociadas con el uso del suelo; 4) características de drenaje de la cuenca; 5) características geográficas; 6) características meteorológicas; 7) características geológicas; 8) medidas de respuesta de la cuenca así como del pico de avenida y 9) estacionalidad y estadísticas direccionales de los flujos.

En la práctica, las regiones homogéneas son formadas utilizando un enfoque de regionalización mediante el uso de los estadísticos de la serie de flujos. Sin embargo se deberían utilizar más características o atributos en la regionalización. Un inconveniente en el uso de estadísticas de flujos como atributos para formar regiones homogéneas es que dichas regiones pueden parecer homogéneas pero no necesariamente son eficaces para el análisis regional (Burn et al., 1997). Del mismo modo, la formación de las regiones no debe basarse exclusivamente en las características fisiográficas de cuencas. Esto se debe a que la similitud fisiográfica no necesariamente implica similitud en la respuesta hidrológica de la cuenca. Por lo tanto para la formación de regiones homogéneas es razonable incluir algunos atributos que se estiman a partir de datos medidos en los sitios de aforo, siempre que estas medidas no están altamente correlacionadas con los valores propios (Hosking y Wallis, 1997, pp. 54-55). Los descriptores de estacionalidad de los flujos son menos propensos a errores y son más robustos que las medidas basadas en datos de la magnitud de flujos. Otras características del lugar en estudio que se pueden utilizar para formar regiones homogéneas pueden basarse en estimaciones suficientemente precisos que pueden utilizarse como si fueran cantidades deterministas, por ejemplo, la lluvia media anual.

5.3 Clasificación de algoritmos clúster

Los algoritmos de clasificación generalmente se dividen en dos (Jain y Dubes, 1988): agrupación jerárquica y la particional. El procedimiento de agrupación jerárquico proporciona una secuencia anidada de particiones mientras que la agrupación particional genera una sola partición de los datos en un intento de recuperar la partición natural de los datos. Los algoritmos de agrupación jerárquica se pueden subdividir en dos categorías: de aglomeración y divisivos. Para un conjunto dado de vectores de características N , los procedimientos de agrupamiento jerárquico aglomerativas comienzan con un número de grupos N . Los grupos aislados son aquellos que consisten en un solo vector de características. Para evaluar la disimilitud entre los grupos se utiliza la distancia euclidiana. Los grupos que son menos

disímiles son encontrados y fusionados. Esto proporciona N-2 grupos aislados y un clúster con dos vectores de características. El proceso de identificación y fusión de los dos grupos más cercanos se repite hasta que se obtiene el número deseado de grupos. Por otro lado el procedimiento de agrupamiento jerárquico divisivos comienzan con un único clúster formado por todos los N de vectores de características. El vector de característica que tiene la mayor disimilitud a otros vectores de la agrupación es identificado y se separa para formar un grupo disidente. Los valores de disimilitud de los vectores de características restantes del clúster original son examinados para determinar si alguno puede agregarse al grupo disidente. En este paso se divide el clúster o grupo original en dos partes. El grupo más amplio se somete al procedimiento mencionado en el siguiente paso. El algoritmo termina cuando se obtiene el número deseado de grupos.

El proceso de agrupación jerárquica (tanto de aglomeración y divisivo) puede ser representado como una secuencia anidada o árbol, llamado dendograma, que muestra cómo se relacionan los grupos que se forman en los diferentes pasos del proceso. El inconveniente de los algoritmos de agrupamiento jerárquico es que los grupos resultantes por lo general no son óptimos debido a que los vectores de características formados en las primeras etapas no se pueden mover a otro grupo. Los algoritmos de agrupamiento jerárquico divisivos siempre son divididos en grupos o clúster. En contraste, los algoritmos aglomerativos siempre se fusionan en grupos o clúster. En los procedimientos de agrupación particional se intenta recuperar la agrupación natural presente en los datos a través de una sola partición. Los algoritmos de agrupamiento basado en prototipos son la clase más popular de los métodos de agrupamiento particional que consideran el prototipo, como grupo centroide, como representado del clúster.

Los algoritmos de agrupación también se pueden clasificar como hard clustering y fuzzy clustering. En hard clustering, cada vector de características se asigna a uno de los grupos con un grado de pertenencia igual a uno. Esto se basa en el supuesto de que los vectores característicos pueden ser divididos dentro de grupos no solapados con fronteras bien definidos entre ellos. Esto es natural para los grupos compactos y bien separados sin embargo, en muchas situaciones realistas se tienen vectores con semejanza parcial a varios grupos y por lo tanto no se pueden justificar plenamente la asignación de un vector de características a un clúster. La teoría de conjuntos difusos o fuzzy clustering (Zadeh, 1965) es una forma natural de representar una situación de este tipo. Fuzzy clustering permite un vector de características de pertenecer a todos los grupos de forma simultánea con un cierto grado de pertenencia en el intervalo $[0, 1]$.

5.4 Pasos en la regionalización mediante el análisis clúster

- 1) ***Selección de atributos:*** en este paso se analizan los datos de distintas variables para identificar los atributos que influyen en la respuesta de la cuenca o región en estudio.
- 2) ***Preparación de vectores característicos:*** en este paso los datos disponibles para cada atributo son reescalados para anular las diferencias en sus varianzas y sus magnitudes relativas.
- 3) ***Formación de grupos o clúster:*** en este paso se particionan vectores característicos disjuntos o superpuestos mediante algún algoritmo clúster.

- 4) **Selección del número óptimo de regiones:** a partir de los grupos formados en el paso anterior, acá se interpretan visualmente y mediante índices de validez para determinar el número óptimo de regiones.
- 5) **Pruebas en las regiones homogéneas:** después de formar regiones en el paso anterior es necesario probar su homogeneidad mediante el uso de pruebas estadísticas que miden su grado de homogeneidad.
- 6) **Ajustes de regiones homogéneas:** las regiones con mayor posibilidad de ser homogéneas son ajustadas para poder mejorar su homogeneidad.

5.5 Clasificación mediante el algoritmo de Fisher

Este algoritmo calcula exactamente una partición óptima en k de un conjunto Ω . El objetivo es construir particiones de forma iterativa con un número de clases igual al índice de la iteración actual escogiendo de manera óptima cada clase agregada una partición existente.

Para el caso multidimensional se tiene el conjunto $\Omega = \{x_1, x_2, x_3, \dots, x_n\}$ con $x_i \in R^p$. Para poder hallar la partición $P(m, k) = (C_1, C_2, C_3, \dots, C_k)$ de Ω en k clases tal que $1 < k < m$. Con pesos de los individuos como $p_1, p_2, p_3, \dots, p_m$, siempre que $\sum_{j=1}^m p_j = 1$ y siendo $n_1, n_2, n_3, \dots, n_k$ el efectivo de cada clase, con $\sum_{i=1}^k n_i = m$. Se tiene el siguiente criterio a minimizar denotado como inercia intra-clases, esto es como sigue.

$$W[P(m, k)] = \sum_{i=1}^k I(C_i) = \sum_{i=1}^k \sum_{j \in C_i} p_j [x_j - G(C_i)]^2 \quad (5.1)$$

con

$$G(C_i) = \sum_{j \in C_i} \frac{p_j}{\mu_i} x_j \quad y \quad \mu_i = \sum_{j \in C_i} p_j \quad (5.2)$$

Finalmente P_l^j denota una partición óptima de $\{j, j+1, \dots, m\}$ en l clases. La partición P_k^1 de inercia intra-clase es la mínima calculada por el algoritmo de Fisher. De acuerdo con el criterio de inercia intra-clase, este es aditivo, lo que conduce a un proceso de optimización mediante la programación dinámica.

El algoritmo de Fisher es denotado por dos pasos donde se construye recursivamente una sucesión P_l^j de las particiones óptimas $\{j, j+1, \dots, m\}$ en l clases, como se menciono anteriormente. En forma concreta interesa encontrar P_k^1 , partición óptima de $\{1, 2, \dots, m\}$ en k clases. El algoritmo de Fisher es mostrado como sigue:

Paso 1: para $j = 1, 2, \dots, m$ se define $P_1^j = \{j, j + 1, \dots, m\}$ siendo la partición optima de $\{j, j + 1, \dots, m\}$ es una sola clase.

Paso 2: para $l = 2, \dots, k$ (etapa l)

Para $j = k - l + 1, \dots, m - l + 1$

Se busca un $h_0 \in \{j, \dots, m - l + 1\}$ minimizando

$$I(\{j + 1, \dots, i\}) + P_{l-1}^{i+1} \text{ donde } i \text{ varia en } \{j, \dots, m - l + 1\}.$$

Se define

$$P_l^j = \{j, j + 1, \dots, i_0\}, P_{l-1}^{i_0+1} \text{ donde } P_{l-1}^{i_0+1}$$

es la partición optima de

$\{i_0 + 1, \dots, m\}$ en $l - 1$ clases, en la etapa $l - 1$.

$$W[P_l^j] = I(\{j, \dots, i_0\}) + W[P_{l-1}^{j_0+1}]$$

5.6 Teoría de Momentos-L

L-momentos han sido ampliamente utilizados en el análisis de frecuencia de avenidas, Momentos-L y relaciones Momentos-L son más convenientes que los momentos de probabilidad pesada porque son más fáciles de interpretar como medidas de forma distribucional (Hoskin, 1994) similares a los productos de momentos ordinarios, Las principales ventajas sobre los momentos convencionales (producto) es que son capaces de caracterizar una gama más amplia de distribuciones. El propósito de Momentos-L y momentos de probabilidad pesada es para resumir distribuciones teóricas y muestras observadas. Por lo tanto Momentos-L pueden ser utilizados para la estimación de parámetros, estimación de intervalos y pruebas de hipótesis (Voguel, 1993). Detalles sobre el método de Momentos-L se pueden encontrar en Hosking y Wallis (1997). Brevemente, se trata de una modificación de los momentos de probabilidad pesada (MPP) método explorado por Greenwood et al. (1979), con la ventaja de ofrecer una descripción de la forma de una distribución de probabilidad mediante asimetría-L y curtosis-L. Los momentos-L son una combinación lineal de los momentos de probabilidad pesada y vienen dados por:

$$\lambda_{r+1} = \int_0^1 x(F)P_r^*(F)dF = \sum_{k=0}^r \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} \beta_k \quad (5.3)$$

con

$$P_r^*(F) = \sum_{k=0}^r \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} F_k \text{ and } \beta_r = \int_0^1 x(F)F_r dF \quad (5.4)$$

Donde $F(x)$ es una función de distribución acumulada (fda) y $x(F)$ la función cuantil. Las relaciones de Momentos-L son:

$$\tau = \frac{\lambda_2}{\lambda_1} \text{ y } \lambda_r = \frac{\lambda_r}{\lambda_2}, \quad r = 3, 4, \dots \quad (5.5)$$

Cada una análogas a las relaciones tradicionales, es decir τ es el coeficientes de variación ($L - CV$); τ_3 la asimetría-L y τ_4 la curtosis-L. Los parámetros son estimados al equiparar los Momentos-L ordinarios con la distribución de Momentos-L. En la práctica, los Momentos-L son estimados a partir de una muestra finita. Véase si $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ es una simple muestra de tamaño n ordenada. Los Momentos-L de una muestra están dados por:

$$l_{r+1} \sum_{k=0}^r \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!} b_k \quad (5.6)$$

Donde b_k es un estimador imparcial de β_k con:

$$b_k = n^{-1} \sum_{i=k+1}^n \frac{(i-1)(i-2) \dots (i-k)}{(n-1)(n-2) \dots (n-k)} x_{i:n} \quad (5.7)$$

Las relaciones de los Momentos-L muestrales están definidos como:

$$t = \frac{l_2}{l_1} \text{ y } t_r = \frac{l_r}{l_2}, \quad r = 3, 4, \dots \quad (5.8)$$

Que se utilizarán para el análisis de homogeneidad en el análisis de frecuencia regional.

5.7 Pruebas de homogeneidad regional

La siguiente se usa en general para el análisis de frecuencia regional. Supóngase que hay N sitios en la región con tamaños de muestra n_1, n_2, \dots, n_N , respectivamente. Las relaciones de los Momentos-L muéstrales en el sitio i son denotados por t^i, t_3^i y t_4^i etc. Los promedios ponderados de las relaciones de los Momentos-L están dados por:

$$\bar{t} = \frac{\sum_{i=1}^N n_i t^{(i)}}{\sum_{i=1}^N n_i} \text{ y } \bar{t}_r = \frac{\sum_{i=1}^N n_i t_r^{(i)}}{\sum_{i=1}^N n_i} \quad (5.9)$$

$$r = 3, 4, \dots$$

Detección de datos utilizando la medida de discordancia

Véase si $\mathbf{u}_i = [t^{(i)}, t_3^{(i)}, t_4^{(i)}]^T$ sea el vector que contiene los valores t, t_3 y t_4 para el sitio i donde superíndice T denota la transposición del vector o matriz. Quedando:

$$\bar{\mathbf{u}} = \sum_{i=1}^N \mathbf{u}_i / N \quad (5.10)$$

Sea el promedio regional (no ponderado). La medida de discordancia para el sitio i está definida entonces como sigue:

$$D_i = \frac{1}{3} N(\mathbf{u}_i - \bar{\mathbf{u}})^T \mathbf{A}^{-1} (\mathbf{u}_i - \bar{\mathbf{u}}) \quad (5.11)$$

Donde

$$\mathbf{A} = \sum_{i=1}^N (\mathbf{u}_i - \bar{\mathbf{u}})^T (\mathbf{u}_i - \bar{\mathbf{u}}) \quad (5.12)$$

Obviamente, un valor grande de D_i indica la discordancia del sitio i con otros sitios. Hosking & Wallis (1997) encontraron que no hay un número fijo que sugieran que D_i sea un valor grande y sugirieron algunos valores críticos para la prueba de discordancia que dependen del número de sitios en cada región (ver Tabla 5.1).

Tabla 5.1. Valores críticos de D_i para la prueba de discordancia (Hosking & Wallis, 1997).

Número de sitios	5	6	7	8	9	10	11	12	13	14	≥ 15
Valor critico	1.333	1.648	1.917	2.140	2.329	2.491	2.632	2.757	2.869	2.971	3.000

5.7.1 Prueba de homogeneidad usando la medida de heterogeneidad

Se debe tomar en cuenta que en una región homogénea todos los sitios deben de tener la misma población de acuerdo a Momentos-L. Una simple medida de dispersión de los Momentos-L muestrales es la desviación estándar de todos los sitios L-CV:

$$V = \left\{ \sum_{i=1}^N n_i (t^i - t^R)^2 / \sum_{i=1}^N n_i \right\}^{\frac{1}{2}} \quad (5.13)$$

Hosking & Wallis (1997) construyeron un estadístico H como:

$$H = \frac{V - \mu_V}{\sigma_V} \quad (5.14)$$

Para medir la heterogeneidad entre los sitios de la región, donde μ_V y σ_V son la media y la desviación estándar, respectivamente, de N_{sim} valores simulados de V . Las simulaciones se realizaron utilizando una distribución flexible de las relaciones regionales promedio de las relaciones de Momentos-L \bar{t} , \bar{t}_3 , y \bar{t}_4 . Después Hosking & Wallis (1993, 1997), usaron la distribución kappa de cuatro parámetros con la función cuantil como sigue:

$$x(F) = \xi + \alpha \frac{\left\{ 1 - \left[\frac{1 - F^h}{h} \right]^k \right\}}{k} \quad (5.15)$$

En las simulaciones (Hosking, 1994). Con el fin de obtener valores confiables de μv y σv , el número de simulaciones tiene que ser grande, en este estudio se utilizó $N_{sim} = 500$. La región se considera “aceptablemente homogénea” si $H < 1$, “posiblemente heterogénea” si $1 \leq H < 2$, y “definitivamente heterogénea” si $H \geq 2$.

5.8 Ajuste de regiones

Si las regiones que se han obtenido a partir del análisis clúster no son estadísticamente homogéneas, entonces pueden volverse a ajustar para mejorar su homogeneidad. Este paso en la regionalización se justifica por que las regiones en general no se esperan sean homogéneas cuando dichas regiones son formadas a partir de atributos. Sin embargo, las modificaciones no deben ser sustanciales si los atributos utilizados para análisis clúster incluyen un número razonable de variables causales que afectan el flujo de respuesta de las cuencas y si un algoritmo clúster eficiente es utilizado para la regionalización.

Las opciones sugeridas por Hosking y Wallis (1997) para el ajuste de las regiones resultantes mediante el algoritmo de agrupación incluyen: (i) la eliminación (o eliminar) uno o más sitios de la serie de datos; (ii) la transferencia (o movimiento) de uno o más sitios de una región a otras regiones; (iii) dividir una región para formar dos o más nuevas regiones; (iv) permitir que un sitio para ser compartido por dos o más regiones; (v) disolver regiones mediante la transferencia de sus sitios a otras regiones; (vi) la fusión de una región con otra; (vii) la fusión de dos o más regiones y grupos; y (viii) la obtención de más datos y la redefinición de las regiones. Entre estas, las tres primeras opciones son útiles en la reducción de los valores de las medidas de heterogeneidad de una región, mientras que las opciones (iv) a (vii) ayudan a asegurar que cada región es suficientemente grande.

5.9 Regionalización mediante el análisis Fuzzy C-Medias

El algoritmo Fuzzy C-Medias fue propuesto por Dunn (1974) y extendido por Bezdek (1981) este último muy popular. Este algoritmo que está basado en optimización iterativa de una función objetiva fuzzy. Este algoritmo es muy útil para particionar N cuencas en una región en c grupos difusos o fuzzy clusters.

Si se tiene a y_k que denota $k - th$ vector de características que representa $k - th$ cuencas en función de n dimensiones en el espacio con etiquetas de los ejes en las coordenadas (y_1, \dots, \dots, y_n) es decir $[y_{1k}, \dots, \dots, y_{nk}] \in R^n$, donde y_{ik} denota los valores del atributo i en y_k . Los atributos del vector característico y_k son reescalados como.

$$x_{ik} = \frac{w_i}{\sigma_i} [f(y_{ik})] \quad \text{para } 1 \leq i \leq n, 1 \leq k \leq N \quad (5.16)$$

Donde x_{ik} denota el valor reescalado de y_{ik} ; w_i es el peso asignado al grupo i ; σ_i se refiere a la desviación estándar del atributo i ; $f(\cdot)$ Representa la función de transformación y N representa el número de n dimensiones vectores característicos. El reescalado de los atributos es necesario debido a las diferencias en su varianza, magnitudes e importancia relativa.

La serie de los N vectores característicos reescalados pueden ser representados como una matriz de datos X de tamaño $n \times N$.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1N} \\ \dots & \ddots & \dots \\ x_{n1} & \dots & x_{nN} \end{bmatrix} \quad (5.17)$$

Además sea $V = (v_1 \dots \dots v_c)$ representan una c -tupla de prototipos v_i , cada uno de los cuales caracteriza el centroide de uno de los c clusters. El algoritmo Fuzzy c -Medias particiona la matriz X en c subconjuntos minimizando la siguiente función objetivo.

$$J(U, V; X) = \sum_{i=1}^c \sum_{k=1}^N (u_{ik})^\mu d^2(x_k, v_i) \quad (5.18)$$

Sujeta a las siguientes condiciones

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \in \{1, \dots, N\} \quad (5.19)$$

$$0 < \sum_{k=1}^N u_{ik} < N \quad \forall i \in \{1, \dots, c\} \quad (5.20)$$

Donde $u_{ik} \in [0,1]$ denota la pertenencia o grado de pertenencia del k -th vector de características reescalado x_k en la i -th grupo difuso o fuzzy cluster; U es la matriz de partición difusa que contiene los miembros de cada vector de características reescalado en cada grupo difuso; el parámetro $\mu \in [1, \infty]$ se refiere al peso del exponente para cada miembro difuso; $d^2(x_k, v_i)$ es la distancia de k -th vector de características reescalado x_k al centroide del i -th grupo v_i . Cuando se usan prototipos de puntos la forma general de la medida de distancia se da por la siguiente ecuación.

$$U = \begin{bmatrix} u_{11} & \dots & u_{1k} & \dots & u_{1N} \\ \vdots & & \vdots & & \vdots \\ u_{i1} & \dots & u_{ik} & \dots & u_{iN} \\ \vdots & & \vdots & & \vdots \\ u_{c1} & \dots & u_{ck} & \dots & u_{cN} \end{bmatrix}_{c \times N} \quad (5.21)$$

$$d^2(x_k, v_i) = (x_k - v_i)^T A_i (x_k - v_i) \quad (5.22)$$

Donde la norma A_i es una matriz simétrica definida positiva asociada con el grupo i , para la estimación de la distancia euclidiana entre x_k y v_i , $A_i = I \quad \forall i$, donde I es la matriz identidad.

La primera restricción de la ecuación 5.21 requiere que los miembros de un vector característico de entradas elegido sobre todos los c grupos difusos deben de sumar 1.0.

Es significativo asignar valores de pertenencia muy pequeños a un vector de características si es representante de una cuenca cuya respuesta hidrológica es bastante diferente a la respuesta hidrológica de las otras cuencas consideradas para la agrupación. Sin embargo, la primera restricción no permite las iteraciones del procedimiento de Fuzzy c-Medias para converger a una solución para el que la pertenencia del vector de características en todos los grupos c no sume 1. Así, en la solución del algoritmo Fuzzy c-Medias, existe la posibilidad de que ciertos sitios que no encajan en ninguna de las regiones identificadas todavía tendrían valores de pertenencia considerables en todos los grupos, de tal manera que sumen uno. Esto, a su vez, afecta a la homogeneidad de las agrupaciones resultantes. Para resolver este problema, los procedimientos para ajuste de regiones de Hosking y Wallis (1997), podrían ser útiles.

5.10 Referencias

- Acreman MC, Sinclair CD (1986) Classification of drainage basins according to their physical characteristics: An application for flood frequency analysis in Scotland. *Journal of Hydrology* 84(3–4): 365–380.
- Bezdek JC (1981) Pattern recognition with fuzzy objective function algorithms. Plenum Press, 4: 95–104.
- Bhaskar NR, O'Connor CA (1989) Comparison of method of residuals and cluster analysis for flood regionalization. *Journal of Water Resources Planning and Management* 115(6): 793–808.
- Bobée B, Rasmussen P (1995) Recent advances in flood frequency analysis. U.S. National Report to International Union of Geodesy and Geophysics (1991–1994). *Reviews in Geophysics* 33: 1111–1116.
- Burn DH (1989) Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management* 115(5): 567–582.
- Burn DH (1990a) An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrological Sciences Journal* 35(2): 149–165.
- Burn DH (1990b) Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research* 26(10): 2257–2265.
- Cavadias GS (1989) Regional flood estimation by canonical correlation. Paper Presented to the Annual Conference of the Canadian Society of Civil Engineering, St. John's, Newfoundland, Canada.
- Cavadias GS (1989) Regional flood estimation by canonical correlation. Paper Presented to the Annual Conference of the Canadian Society of Civil Engineering, St. John's, Newfoundland, Canada.
- Cavadias GS (1990) The canonical correlation approach to regional flood estimation. In: Beran MA, Brilly M, Becker A, Bonacci O (Eds.), *Proceedings of the Ljubljana Symposium on Regionalization in Hydrology*. IAHS Publication No. 191, Wallingford, England, pp. 171–178.
- Cavadias GS (1995) Regionalization and multivariate analysis: The canonical correlation approach. *Proceedings of the UNESCO International Conference on Statistical and Bayesian Methods in Hydrological Sciences*, Paris, p. 19.
- Cavadias GS, Ouarda TBMJ, Bobée B, Girard C (2001) A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. *Hydrological Sciences Journal* 46(4): 499–512.
- Choquette AF (1988) Regionalization of peak discharges for streams in Kentucky. *Water Resources Investigation Report 87-4209*. US Geological Survey, Louisville District, Louisville, KY.
- Cunderlik JM, Burn DH (2006a) Switching the pooling similarity distances: Mahalanobis for Euclidean. *Water Resources Research* 42: W03409, doi:10.1029/2005WR004245.
- Cunnane C (1988) Methods and merits of regional flood frequency analysis. *Journal of Hydrology* 100(1–3): 269–290.

- Dalrymple T (1960) Flood frequency analysis. U.S. Geological Survey, Water Supply Paper 1543-A., U.S. Department of the Interior, Washington, DC.
- Dunn JC (1974) Well separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*
- Gabriele S, Arnell N (1991) A hierarchical approach to regional flood frequency analysis. *Water Resources Research* 27(6): 1281–1289.
- Glatfelter DR (1984) Techniques for estimating magnitude and frequency of floods on streams in Indiana. US Geological Survey, Water Resources Investigations Report 84-4134.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C. & Wallis, J. R. (1979) Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour. Res.* 15, 1049–1054.
- GREHYS (Groupe de recherche en hydrologie statistique) (1996) Presentation and review of some methods for regional flood frequency analysis. *Journal of Hydrology* 186: 63–84.
- Hall MJ, Minns AW (1999) The classification of hydrologically homogeneous regions. *Hydrological Sciences Journal* 44(5): 693–704.
- Hall MJ, Minns AW, Ashrafuzzaman AKM (2002) The application of data mining techniques for the regionalization of hydrological variables. *Hydrology and Earth System Sciences* 6(4): 685–694.
- Hosking JRM, Wallis JR (1997) *Regional frequency analysis: An approach based on L-moments*. Cambridge University Press, New York.
- Hosking, J. R. M. & Wallis, J. R. (1993) Some statistics useful in regional frequency analysis. *Water Resour. Res.* 29(2),271–281.
- Hosking, J. R. M. & Wallis, J. R. (1997) *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press: Cambridge, UK.
- Hosking, J. R. M. (1994) The four-parameter Kappa distribution. *IBM J. Res. Develop.* 38(3), 251–258.
- Hosking, J. R. M. (1994) The four-parameter Kappa distribution. *IBM J. Res. Develop.* 38(3), 251–258.
- Jain AK, Dubes RC (1988) *Algorithms for clustering data*. Prentice-Hall Inc., Englewood Cliff, NJ.
- Zadeh LA (1965) Fuzzy sets. *Information and Control* 8(3): 338–353.
- Javelle P, Ouarda, TBMJ, Lang M, Bob'ee B, Gal'ea G, Gr'esillon J-M (2002) Development of regional flow-duration–frequency curves based on the index-flood method. *Journal of Hydrology* 258(1–4): 249–259.
- Jingyi Z, Hall MJ (2004) Regional flood frequency analysis for the Gan-Ming river basin in China. *Journal of Hydrology* 296: 98–117.
- Mosley MP (1981) Delimitation of New Zealand hydrological regions. *Journal of Hydrology* 49: 173–192.
- Nathan RJ, McMahon TA (1990) Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology* 121: 217–238. New York.
- Rao AR, Srinivas VV (2006a) Regionalization of watersheds by hybrid cluster analysis. *Journal of Hydrology* 318(1–4): 37–56.
- Rao AR, Srinivas VV (2006b) Regionalization of watersheds by fuzzy cluster analysis. *Journal of Hydrology* 318(1–4): 57–79.
- Ribeiro-Corréa B, Cavadias GS, Clement B, Rousselle J (1995) Identification of hydrological neighbourhoods using canonical correlation analysis. *Journal of Hydrology* 173: 71–89.
- Shu C, Burn DH (2004) Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement. *Journal of Hydrology* 291(1–2): 132–149.
- Tasker GD (1982) Comparing methods of hydrologic regionalization. *Water Resources Bulletin* 18(6): 965–970.
- Thomas DM, Benson MA (1970) Generalization of streamflow characteristics from drainage basin characteristics. US Geological Survey Water-Supply Paper 1975. US Government Printing Office, Washington, DC.
- Vogel, R.M., McMahon, T.A. and Chiew, F.H.S., 1993b. Flood-flow frequency model selection in Australia. *Journal of Hydrology* 146, pages 421–449.
- Wandle SW Jr (1977) Estimating the magnitude and frequency of floods on natural-flow streams in Massachusetts, US Geological Survey Water Resources Investigations Report, pp. 77–39.
- Wiltshire SE (1986) Regional flood frequency analysis II. Multivariate classification of drainage basins in Britain. *Hydrological Sciences Journal* 31(3): 335–346.
- Zrinji Z, Burn DH (1994) Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology* 153: 1–21.
- Zrinji Z, Burn DH (1996) Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management* 122(4): 245–252.

CAPÍTULO**SEIS**

EXPERIMENTOS Y RESULTADOS

En este capítulo se realiza una comparación del rendimiento entre los modelos tradicionales y los de enfoque regional propuesto para predicción de lluvia y de escurrimiento en diferentes escalas temporales. Los modelos tradicionales y los del enfoque regional propuesto están basados en RNAs, Ensamblados y Modulares. Concretamente se comparó si los modelos de predicción con enfoque regional son más precisos que los de enfoque tradicional. La precisión se evaluó mediante las medidas de eficiencia tradicionales MSE, RMSE, MAE, MAPE, índice de concordancia d y el coeficiente R^2 . Los resultados encontrados en este trabajo demuestran que el enfoque regional propuesto es útil y prometedor en el pronóstico de la lluvia y de escurrimientos en escalas mensuales, trimestral y semestral, ya que los modelos regionales han sido superiores en gran parte a los modelos tradicionales de predicción. En recientes años los modelos utilizados para pronóstico de lluvia y escurrimiento basados en RNAs han demostrado ser una poderosa herramienta. Es por eso que el enfoque regional propuesto se puede considerar como una nueva herramienta para el pronóstico hidrológico. El impulso para el desarrollo de estos modelos ha sido el papel fundamental que juega el pronóstico en los recursos hídricos. Las aplicaciones prácticas incluyen el diseño y operación de estructuras hidráulicas, control de avenidas, gestión de suministro de agua, generación de energía eléctrica, diseño de sistemas de irrigación, mitigación de impactos generados por inundaciones y sequías, entre muchas otras aplicaciones. El propósito fundamental en esta investigación es mejorar el pronóstico hidrológico mediante la regionalización. Se espera que esta investigación siga impulsando al desarrollo y la investigación en el área hidrológica.

6.1 Introducción

Existe un gran interés en pronosticar a corto y largo plazo con el menor error posible, puesto que todo el tiempo se pretende prever condiciones a futuro, una de las variables climáticas más importantes en Hidrología son el escurrimiento y la lluvia ya que son importantes en la planeación, en la operación y en general para disponibilidad hídrica.

Se sabe que la atmósfera es un sistema caótico: ya que cualquier cambio de condiciones aún siendo estas apenas diferentes evolucionarán hasta que, eventualmente, diverjan (Lorenz, 1963,1969). Aunque se pudiera tener un modelo matemático de pronóstico perfecto de las variables hidrológicas, debido a la imposibilidad de tener datos confiables y perfectos de las mediciones de estas variables en el instante t , es imposible determinar el pronóstico exacto en un tiempo $t + 1$ y por lo tanto mientras más lejano sea el pronóstico, mayor será la incertidumbre.

La naturaleza caótica de la atmósfera, hace que sea imposible realizar pronósticos precisos de lluvia y por ende de escurrimientos con suficiente anticipación mayores a su umbral de predictibilidad determinístico, el cual es del entorno de unos pocos días (Lorenz, 1982).

Bajo este entorno se plantea la pregunta, ¿son posibles los pronósticos de escurrimientos, lluvia u otras variables hidrológicas con anticipación mayor a unos pocos días? La respuesta es sí, aunque bajo ciertas condiciones, en esta tesis doctoral se abordan los pronósticos mensuales desde el punto de vista estadístico (modelo de predicción orientado por datos) (Wang, 2006).

En todas las pruebas que se ensayaron con los distintos modelos se trató de encontrar un patrón respecto de los datos anteriores en la serie histórica, basándose únicamente en las relaciones matemáticas existentes de los valores actuales respecto de los anteriores. Esto es pronosticando el mes siguiente Q_{t+1} dado los meses anteriores $Q_{t-k}, Q_{t-k+1}, \dots, Q_{t-1}, Q_t$ siendo k el número de retrasos.

En la aplicación del enfoque Regional propuesto, se compararon los modelos tradicionales RNAs, Ensamblados y Modulares basados en RNAs con el enfoque Regional para los modelos RNAs, Ensamblados y Modulares basados en RNAs. Se han tomado series mensuales de lluvia y escurrimiento de los países de Australia y Canadá debido a que tienen una gran cantidad de información disponible, pues se ha procurado que la cantidad de vacíos en los registros sean menores al 10% y de esta forma garantizar resultados confiables.

La utilización de modelos Regionales ha demostrado ser una herramienta eficaz para el pronóstico de lluvias y de escurrimientos, de ahí que se propone como una nueva herramienta en el pronóstico hidrológico.

6.2 Estado del arte de la última década en el pronóstico de lluvia

En la última década se han aplicado con éxito RNAs al pronóstico de lluvia, por ejemplo French et al., 1992. Utilizaron un modelo de RNAs para pronosticar los campos de intensidad en el espacio y tiempo. RNAs fue capaz de pronosticar los campos de intensidad de la lluvia de una hora. Los resultados de RNAs los compararon con los generados por el modelo de

simulación matemática y encontraron que RNAs es capaz de encontrar la compleja relación que describe la evolución espacio-temporal de las precipitaciones. Hungs et al., 2009. Aplicaron un modelo de redes neuronales artificiales RNAs para mejorar el pronóstico de la lluvia, usando datos horarios de lluvia en 75 estaciones de lluvia en Bangkok, Tailandia. Ellos encontraron que el enfoque RNAs es superior al modelo convencional de persistencia además que el parámetro de entrada más importante aparte de la lluvia es la temperatura. Kin et al., 2001. Aplicaron RNAs para pronosticar la distribución espacial de la precipitación para una cuenca urbana. Ellos aplicaron tres modelos diferentes de RNAs, red neuronal multicapa feedforward, recurrente y de retraso o delay. Finalmente encontraron que RNAs proporciona predicciones razonables de altura de lluvia. Khalida et al., 2013. Aplicaron un modelo RNAs de retardo de tiempo TDNN para pronóstico de lluvia en la región de Tizi-Ouzou en Argelia. El ancho de ventana utilizada fue de dos a cuatro meses. Los resultados indican que RNAs tiene un adecuado potencial en el pronóstico de lluvias. Nastos et al., 2013. Construyeron modelos predictivos basados en RNAs. Ellos pronosticaron la intensidad de la lluvia en mm/día en Atenas, Grecia. Los resultados se refieren a lluvia máximas, medias y mínimas proyectadas de forma mensual para cuatro meses. Encontraron que los modelos basados en RNAs son bastantes confiables para el pronóstico de intensidades de lluvias para los siguientes cuatro meses. Los resultados mediante RNAs de la intensidad de la lluvia pronosticada en comparación con la correspondiente observada les han resultado muy bien, pudiendo ser altamente competitivos los modelos RNAs. Farajzadeh et al., 2014. Pronosticaron las entradas de la cuenca del lago Urmia situado en el noroeste de Iran. Ellos utilizaron un modelo de red neuronal artificial feedforward y un modelo ARIMA para pronosticar precipitación mensual, en ese estudio encontraron que ambos modelos tienen resultados similares sin diferencia significativa por lo tanto RNAs es una herramienta competitiva en el pronóstico de precipitación. Abbot & Marohasy, 2014. Hicieron uso de las redes neuronales artificiales RNAs para pronóstico de precipitaciones estacionales encontrando que se destaca el valor de la oscilación del pacifico interdecadal. Este índice de acuerdo con los autores no se había utilizado en el pronóstico en Queensland. Utilizaron tres regiones geográficamente distintas y demuestran que RNAs es superior al modelo Predictive Ocean Atmosphere Model for Australia (POAMA). El cual es el modelo general de circulación actualmente utilizado para pronósticos oficiales de precipitaciones estacionales.

6.3 Estado del arte de la última década en el pronóstico de escurrimientos

En la modelación de escurrimientos las RNAs han alcanzado una gran gama de aplicaciones, por ejemplo Zealand et al., 1999. Investigaron la utilidad de redes neuronales artificiales (RNAs) para la predicción a corto plazo de escurrimientos comparando las capacidades de RNAs con los enfoques convencionales en el pronóstico de escurrimientos en una parte del rio Winnipeg en el noroeste de Ontario, Canadá. Ellos utilizaron series de cuartos de mes de escurrimientos y encontraron que RNAs superan a los modelos convencionales de pronóstico. Cigizoglu, 2003. Investigó la aplicabilidad de RNAs para predicción estimación y

extrapolación de datos de escurrimiento diario en ríos de la región del Mediterráneo oriental de Turquía. Encontró evidentemente que la aplicación de RNAs da mejores resultados que los modelos convencionales. Mohammadi et al., 2005. Analizaron tres métodos diferentes de pronóstico de escurrimientos en el reservorio Amir Kabir en la cuenca del río Karaj, localizada al noroeste de Tehran (Iran). Los modelos analizados fueron, redes neuronales artificiales RNAs, ARIMA y análisis de regresión, finalmente encontraron que los mejores resultados fueron presentados por RNAs por lo tanto es una herramienta eficaz en el pronóstico de escurrimientos. Abdel et al., 2006. Modelaron escurrimientos en dos ríos de USA, el río Agua Negra (Black Water) y el río Gila. Utilizaron modelos basados en RNAs pronosticaron escurrimientos diarios. Ellos probaron diferentes tipos de arquitecturas de RNAs comparándolo con el modelo convencional AR encontrando que RNAs es el mejor predictor además de recomendarlo para el pronóstico de escurrimientos siendo una herramienta exitosa. Diamantopoulou et al., 2006. Usaron un modelo de red neuronal artificial de tres capas denominado Time Delay Artificial Neural Network (TDANN) para pronóstico de escurrimientos diarios en el río Aliakmon en el norte de Grecia. Ellos determinaron el número de nodos de las capas ocultas basándose en el valor máximo del coeficiente de correlación. Han demostrado que TNANN tiene un gran potencial de aplicabilidad en muchas áreas hidrológicas y ambientales. Demirel et al., 2009. Analizaron el pronóstico de escurrimiento basados en el suelo y agua mediante la herramienta SWAT y RNAs en escurrimientos diarios en la cuenca Pracana en Portugal. Encontraron que RNAs tuvo más éxito que el modelo SWAT esto demostró que RNAs pueden ser herramientas poderosas en el pronóstico de escurrimientos diarios. Demirel & Booi, 2009. Investigaron la selección de un modelo apropiado de pronóstico para flujos bajos en el Río Meuse comparando los modelos ARMAX multivariado, un modelo de regresión lineal y RNAs. Ellos encontraron que RNAs fue el mejor en comparación con los otros modelos para pronosticar flujos bajos en siete días. Pierini et al., 2012. Analizaron series temporales de escurrimientos en embalses aguas arriba y en la planta hidroeléctrica en Buta Ranquil. Ellos compararon redes neuronales artificiales RNAs y modelos autorregresivos AR, ellos encontraron que el mejor modelo para pronóstico fue RNAs. Arslan, 2013. Investigó varios modelos de RNAs para simulación y predicción de caudales aguas abajo en el río Eufrates. El aplico diferentes tipos de arquitecturas de modelos RNAs. Las series pronosticadas fueron mensuales de escurrimientos la conclusión final fue que RNAs son prometedoras para pronóstico de escurrimientos.

6.4 Modelos de predicción en combinación con técnicas de agrupamiento

El objetivo principal de desarrollar un modelo regional vía RNAs es reducir la incertumbre en los pronósticos de la lluvia y el escurrimiento. Si un modelo de predicción basado en RNAs aprende correctamente las características de los datos, entonces se ve que RNAs logra una adecuada generalización. La forma matemática del modelo regional RNAs es como sigue:

$$y_k(t) = f_2 \left[\sum_{j=1}^{n2} w_{kj} f_1 \left(\sum_{i=1}^{n1} w_{ji}^h x_i(t) + w_{j0} \right) + w_{k0} \right] \quad (6.1)$$

Donde

$$x_i(t) = P_t, P_{t-1}, P_{t-2}, \dots, P_{t-r} \quad y \quad y_k(t) = P_{t+1} \quad (6.2)$$

donde x_i es la entrada en el nodo i de la capa de entrada y representa a la estación i de un grupo homogéneo con $n1$ estaciones para dicho grupo, cada entrada x_i esta representada por los datos de lluvia P_t hasta P_{t-r} donde r indica el retraso en cada serie computada de cada estación, y y_k es la salida en el nodo k de la capa de salida, y_k representa a la lluvia P_{t+1} pronosticada en un dato hacia adelante correspondiente a la estación k de un grupo homogéneo con $n1$ estaciones respectivamente. Las funciones de transferencia no lineales más usadas son: la función sigmoideal y la tangente hiperbólica, en esta tesis ambas funciones se usaron optando por la que brindó mejores resultados en cada entrenamiento.

$$\text{Funcion sigmoideal: } f(s_j) = \frac{1}{1 + e^{-s_j}} \quad (6.3)$$

$$\text{Funcion tangente hiperbolica: } f(s_j) = \frac{1 - e^{-2s_j}}{1 + e^{-2s_j}} \quad (6.4)$$

Donde $s_j \in [-\infty, \infty]$. La función sigmoideal está acotada entre $(0, 1)$ y la función tangencial hiperbólica entre $(-1, 1)$.

En redes neuronales artificiales, el entrenamiento es un proceso de optimización no lineal, donde el error entre la salida del modelo y la salida deseada se reduce al mínimo por un algoritmo predeterminado que cambia varias veces los valores de los pesos de las conexiones de las RNAs. Los parámetros de las RNAs se ajustan durante el proceso de entrenamiento a través de la minimización del error cuadrático medio (MSE). En general, en redes neuronales hay un paso crucial (1) la selección de los pesos iniciales y (2) los criterios de detención de aprendizaje (Ray & Klindworth, 2000). Los valores de peso deben de estar cerca de su nivel óptimo ya que esto afecta el rendimiento de la red neuronal. Pocos estudios se han llevado a cabo para encontrar buenos pesos iniciales. En general, los pesos se inicializan mediante la generación de números aleatorios (Haykin, 1999). En este estudio se han empleado para los modelos RNAs y Regionales pesos iniciales al azar. Para detener el proceso de entrenamiento se limitó el número de iteraciones o se estableció un nivel de error aceptable en la fase de entrenamiento.

No hay seguridad de que los parámetros estén cerca de los valores óptimos que serán encontrados durante la fase de aprendizaje a pesar de que el número de iteraciones este limitada a un valor predefinido. Para evitar el sobreentrenamiento, se ha adoptado dos criterios para detener el proceso de formación: 1) RMSE está predefinido y el entrenamiento se realiza

hasta que el RMSE disminuye hasta un valor de umbral. 2) En base a la examinación preliminar, se observó que el error de la red neuronal disminuye tan bajo como el umbral de RMSE dentro de las 2000 épocas con buenos pesos iniciales y sin exceso de entrenamiento, sin embargo, los valores de umbral nunca pueden ser alcanzados para los pesos iniciales malos, incluso después de un gran número de épocas.

Los procedimientos de entrenamiento y validación de arquitecturas de la red específica fueron repetidos con el fin de manejar las incertidumbres de los pesos iniciales y los criterios de suspensión.

En esta investigación preliminar se encontró que entre 5 y 10 ensayos son suficientes para encontrar el mejor resultado. Las eficiencias de rendimiento de cada ensayo fueron registradas y comparadas. El resultado con el más alto valor R era considerado el óptimo de la red.

Para la división de datos en el entrenamiento y validación de la red, se eligió el 90 % para el entrenamiento y el 10% para la validación de los modelos de RNAs y Regionales. Para este propósito se utilizó la función `trainlm` de MATLAB. Esta función actualiza los pesos y valores de sesgo de acuerdo con el algoritmo Levenberg -Marquardt (The MathWorks, 2012).

El resultado de combinar RNAs, Modelos Modulares y Ensamblados con Momentos-L es un modelo regional que se resume en cuatro pasos. Es importante el diseño de la red neuronal para conseguir buenos resultados. En la técnica de Momentos-L se han formado grupos iniciales mediante la técnica de c-Medias y mediante el algoritmo de Fisher para después refinarla mediante Momentos-L. El modelo fue alimentado con series de escurrimientos y lluvia mensuales, trimestrales y semestrales con un retraso de hasta cuatro datos (P_4). Los pasos para aplicar el modelo regional son como sigue.

Paso 1). Normalizar cada uno de los datos de las series estudiadas mediante la siguiente expresión.

$$x_{norm} = \frac{x - min_A}{max_A - min_A} (nuevomax_A - nuevomin_A) + nuevomin_A \quad (6.5)$$

La variable normalizada x_{norm} se encuentra en el rango $[nuevomax_A - nuevomin_A]$, para este caso se ha utilizado el rango $[0, 1]$.

Paso 2). Los datos que alimentaran a la red son la lluvia acumulada mensual de un retraso de hasta 4 y de diferentes estaciones. Para encontrar las estaciones idóneas es necesario agruparlas para obtener regiones homogéneas esto es mediante el algoritmo de Fisher y posteriormente mediante Momentos-L. Las características físicas, tales como la longitud, la latitud, la elevación y la lluvia media anual de las estaciones seleccionadas en el sitio de estudio fueron usadas en la formación de regiones homogéneas.

Paso 3). Diseño de la red RNAs-Regional, Ensamblado-Regional y Modular-Regional. La red utilizada es todos los modelos es de una capa oculta y varias salidas que corresponden al número de estaciones involucradas en el grupo homogéneo. El número de neuronas en la capa

oculta se obtuvo mediante la siguiente expresión, que de acuerdo a Gnana y Depa (2013) presentó los mejores resultados.

La expresión es $\frac{4^n + 18}{n - 1}$ donde n representa el número de entradas en el modelo. El modelo fue entrenado mediante el algoritmo de Levenberg Marquardt (LM) y hasta 2000 épocas. Las funciones de transferencia utilizadas fueron sigmoïdal y tangente hiperbólica.

Paso 4). Determinar los criterios de evaluación de la red. Para este trabajo se emplearon los errores MSE, RMSE, MAE y d , los cuales fueron descritos con anterioridad. La validación del modelo ha sido con el 10% del total de la serie, de esta manera se comparó la capacidad de predicción del modelo. En el siguiente diagrama de flujo se describe el enfoque regional propuesto.

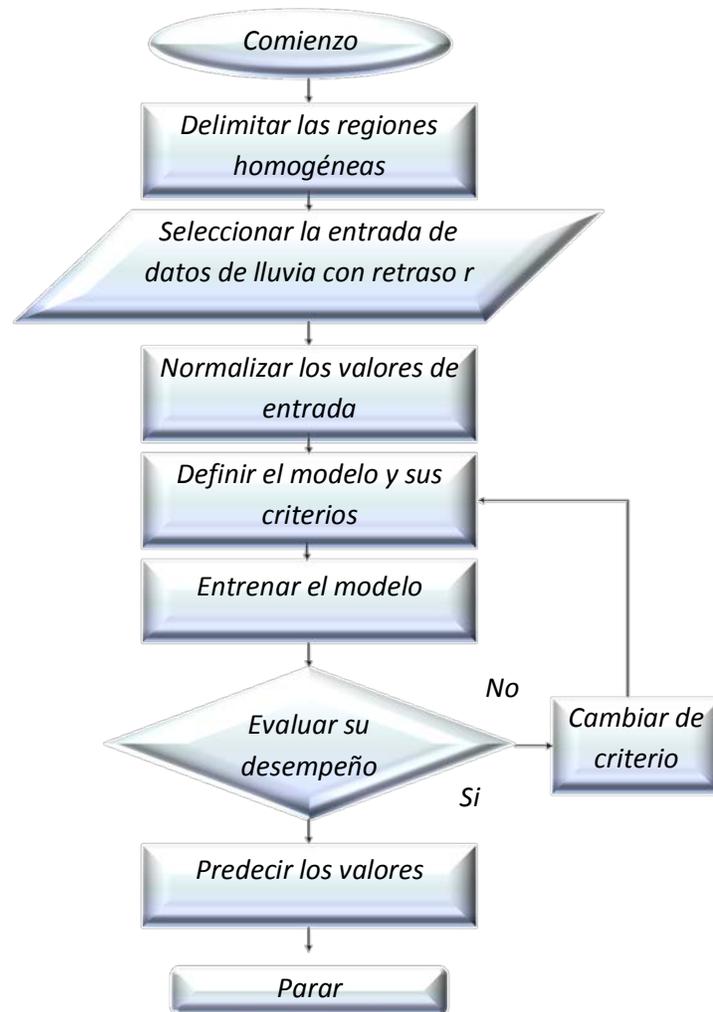


Figura 6.1. Diagrama de flujo del modelo Regional.

6.5 Medidas de eficiencia

Para medir la eficiencia de los modelos regionales propuestos y el modelo tradicional RNAs se emplearon cuatro pruebas de bondad de ajuste: error absoluto medio (MAE), la raíz del error cuadrático medio (RMSE) y porcentaje del error absoluto medio (MAPE) se utilizaron como criterios de rendimiento. El índice de concordancia (d) fue elegido como un criterio para comparar modelos. Los criterios de rendimiento se definen como sigue:

Tabla 6.1. Medidas de eficiencia de los modelos

Medida de error	Nombre	Fórmula	
Error cuadrático medio	MSE	$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	(6.6)
Raíz del error cuadrático medio	RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	(6.7)
Error absoluto medio	MAE	$\frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	(6.8)
Porcentaje del error absoluto medio	MAPE	$\frac{100}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{y_i}$	(6.9)
El índice de concordancia	<i>d</i>	$1 - \frac{\sum_{i=1}^n y_i - \hat{y}_i ^2}{\sum_{i=1}^n y_i - \bar{y}_i ^2 + \hat{y}_i - \bar{y}_i ^2}$	(6.10)

Donde *n* es el número total de salidas pronosticadas, *y_i* es el *i*ésimo escurrimeintos pronosticado mediante los modelos tradicionales o regionales, *ŷ_i* es la correspondiente *i*ésimo observacion. *ȳ_i* Es la media de lso valores pronosticados y observados respectivamente. Para una buena prediccion, el índice de concordancia tiene que estar cercano a 1 y los MSE, RMSE, MAE y MAPE tienen que ser pequeños.

6.6 Comparación de los modelos de predicción en lluvias

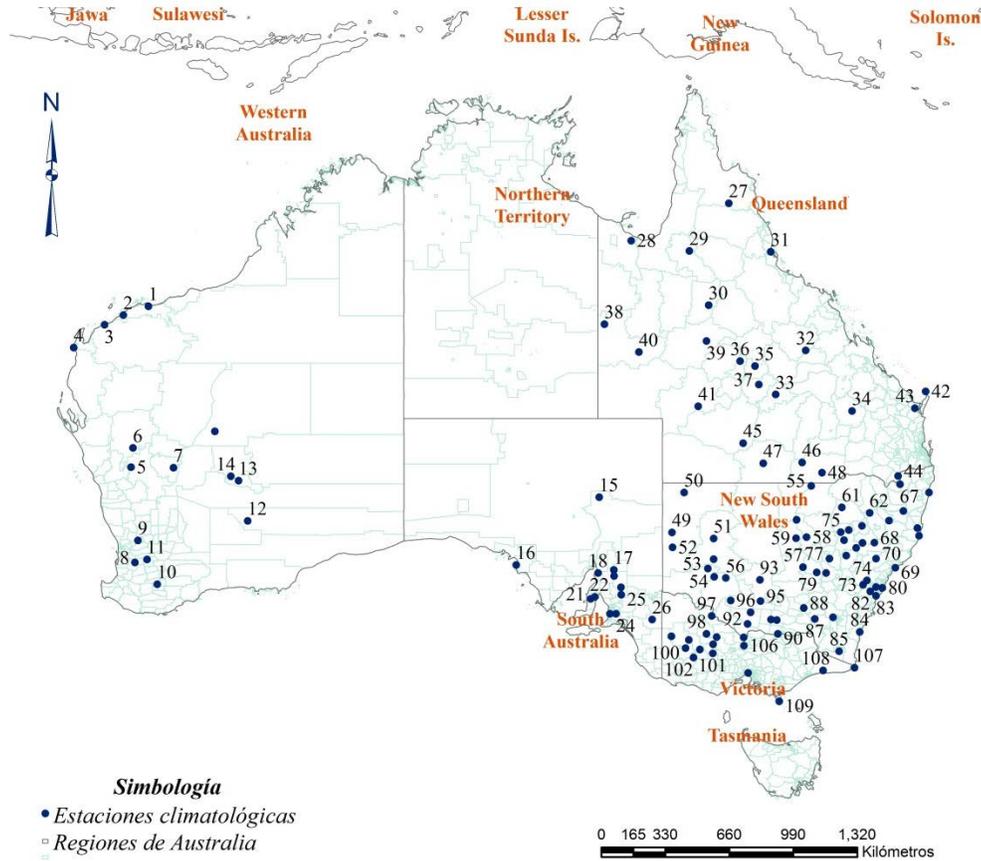


Figura 6.2. Ubicación de las estaciones climatológicas de estudio en Australia.

En este estudio se utilizaron datos de lluvia mensual tomados de 109 estaciones climatológicas en Australia. Los datos fueron tomados de NOAA - National Oceanic and Atmospheric Administration. La ubicación de las estaciones se muestra en la Figura 6.2. Los datos observados tienen una longitud de registro de 106 años (1272 meses) con observaciones comprendidas entre el periodo 1907 y 2012, las estaciones fueron escogidas cuidando que la cantidad de datos vacíos no fuera mayor del 10%. Los datos observados comienzan en enero y el último mes es diciembre. Las características de las estaciones se presentan en la Tabla 6.2.

Tabla 6.2. Ubicación de las estaciones climatológicas estudiadas en Australia.

Número	Código	Nombre	Latitud	Longitud	Elevación
1	ASN0004035	ROEBOURNE	-20.78	117.15	12
2	ASN0005008	MARDIE	-21.19	115.98	11
3	ASN0005016	ONSLow	-21.64	115.11	4
4	ASN0005020	NINGALOO	-22.70	113.67	10
5	ASN0007027	GABYON	-28.25	116.34	340
6	ASN0007064	MURGOO	-27.36	116.43	303
7	ASN0007197	CHALLA	-28.28	118.31	400
8	ASN0009507	BANNISTER	-32.68	116.52	350
9	ASN00010111	NORTHAM	-31.65	116.66	170
10	ASN00010579	KATANNING COMPARISON	-33.69	117.56	310
11	ASN00010626	PINGELLY	-32.53	117.08	297
12	ASN00012013	BULONG	-30.75	121.75	380
13	ASN00012046	LEONORA	-28.88	121.33	376
14	ASN00012176	STURT MEADOWS	-28.68	120.97	400
15	ASN00017031	MARREE COMPARISON	-29.65	138.06	50
16	ASN00018079	STREAKY BAY	-32.80	134.21	13
17	ASN00019062	YONGALA	-33.03	138.75	512
18	ASN00021043	PORT PIRIE NYRSTAR	-33.17	138.01	4
19	ASN00021062	HALLETT (OLD CANOWIE)	-33.30	138.76	575
20	ASN00021086	BURRA (WORLDS END)	-33.84	139.07	293
21	ASN00022008	MAITLAND	-34.37	137.67	185
22	ASN00022021	ARDROSSAN (WINULTA)	-34.28	137.86	135
23	ASN00023721	HAPPY VALLEY RESERVOIR	-35.06	138.56	170
24	ASN00023733	MOUNT BARKER	-35.06	138.85	360
25	ASN00024511	EUDUNDA	-34.18	139.08	415
26	ASN00025509	LAMEROO COMPARISON	-35.33	140.52	99
27	ASN00028004	PALMERVILLE	-16.00	144.08	203.8
28	ASN00029004	BURKETOWN POST OFFICE	-17.74	139.55	5.5
29	ASN00029012	CROYDON TOWNSHIP	-18.20	142.24	115.5
30	ASN00030045	RICHMOND POST OFFICE	-20.73	143.14	211.1
31	ASN00032004	CARDWELL MARINE PDE	-18.26	146.02	4.5
32	ASN00035019	CLERMONT SIRIUS ST	-22.82	147.64	260
33	ASN00035069	TAMBO POST OFFICE	-24.88	146.26	395.1
34	ASN00035070	TAROOM POST OFFICE	-25.64	149.80	199
35	ASN00036007	BARCALDINE POST OFFICE	-23.55	145.29	266.9
36	ASN00036066	BEACONSFIELD	-23.33	144.60	230
37	ASN00036143	BLACKALL TOWNSHIP	-24.42	145.47	284
38	ASN00037043	URANDANGI	-21.61	138.31	174
39	ASN00037051	WINTON POST OFFICE	-22.39	143.04	181.9
40	ASN00038003	BOULIA AIRPORT	-22.91	139.90	161.8
41	ASN00038024	WINDORAH POST OFFICE	-25.42	142.66	126.3
42	ASN00039085	SANDY CAPE LIGHTHOUSE	-24.73	153.21	109
43	ASN00040126	MARYBOROUGH	-25.52	152.72	10.3
44	ASN00041095	STANTHORPE LESLIE	-28.66	151.93	787
45	ASN00044004	BEECHAL	-27.14	144.74	218
46	ASN00044010	BOLLON MARY ST	-28.03	147.48	182.9
47	ASN00044026	CUNNAMULLA POST OFFICE	-28.07	145.68	188.7
48	ASN00044166	YAMBURGAN	-28.51	148.40	170
49	ASN00046003	YANCO GLEN (CORONA)	-31.29	141.44	233
50	ASN00046037	TIBOOBURRA POST OFFICE	-29.43	142.01	183
51	ASN00046043	WILCANNIA (REID ST)	-31.56	143.37	75
52	ASN00047007	BROKEN HILL (PATTON)	-31.98	141.47	315
53	ASN00047013	POONCARIE (KARPA KORA)	-32.97	143.11	72
54	ASN00047024	POONCARIE (MULURULU)	-33.34	143.40	75

55	ASN00048168	ANGLEDPOOL (ANGLEDPOOL	-29.12	147.90	150
56	ASN00049008	HATFIELD (CLARE)	-33.40	143.94	77
57	ASN00050028	TRUNDLE (MURRUMBOGIE)	-32.90	147.52	264
58	ASN00051034	WARREN (MUMBLEBONE)	-31.50	147.69	192
59	ASN00051039	NYNGAN AIRPORT	-31.55	147.20	173
60	ASN00051057	MARRA CREEK (WOMBOIN)	-30.70	147.22	131
61	ASN00053034	WEE WAA (PENDENNIS)	-30.12	149.32	168
62	ASN00054003	BARRABA POST OFFICE	-30.38	150.61	500
63	ASN00055018	MULLALEY (GARRAWILLA)	-31.17	149.65	375
64	ASN00055023	GUNNEDAH POOL	-30.98	150.25	285
65	ASN00055057	WILLOW TREE (VALAIS)	-31.77	150.29	465
	ASN00056028	URALLA (SALISBURY	-30.73	151.51	1018
66	ASN00056032	TENTERFIELD (FEDERATION	-29.05	152.02	838
67	ASN00061051	MURRURUNDI POST OFFICE	-31.77	150.84	466
68	ASN00061055	NEWCASTLE NOBBYS	-32.92	151.80	33
69	ASN00061086	JERRYS PLAINS POST OFFICE	-32.50	150.91	90
70	ASN00062009	CASSILIS (DALKEITH)	-32.00	149.99	420
71	ASN00062013	GULGONG POST OFFICE	-32.36	149.53	475
72	ASN00063039	KATOOMBA (MURRI ST)	-33.71	150.31	1015
73	ASN00063118	BILPIN (FERN GROVE)	-33.52	150.49	610
74	ASN00064008	COONABARABRAN (NAMOI	-31.27	149.27	505
75	ASN00064013	BINNAWAY (HAWTHORNE)	-31.64	149.43	410
76	ASN00065000	ARTHURVILLE (CRAMOND)	-32.50	148.75	305
77	ASN00065022	MANILDRA (HAZELDALE)	-33.16	148.59	530
78	ASN00065026	PARKES (MACARTHUR	-33.14	148.16	324
79	ASN00066062	SYDNEY (OBSERVATORY	-33.86	151.21	39
80	ASN00067019	PROSPECT RESERVOIR	-33.82	150.91	61
81	ASN00068007	CAMDEN (BROWNLOW HILL)	-34.03	150.65	61
82	ASN00068024	DARKES FOREST (KINTYRE)	-34.23	150.91	370
83	ASN00069018	MORUYA HEADS PILOT	-35.91	150.15	17
84	ASN00070009	BUKALONG STATION	-36.80	149.20	785
85	ASN00070032	FAIRLIGHT STATION	-35.23	148.91	610
86	ASN00072000	ADELONG POST OFFICE	-35.31	148.06	333
87	ASN00073025	OLD JUNEE (MILLBANK)	-34.79	147.56	270
88	ASN00074026	URANA (BUTHERWAH)	-35.35	146.31	120
89	ASN00074034	COROWA AIRPORT	-35.99	146.36	143
90	ASN00074087	URANA (NOWRANIE)	-35.33	146.03	115
91	ASN00074128	DENILQUIN (WILKINSON ST)	-35.53	144.95	93
92	ASN00075032	HILLSTON AIRPORT	-33.49	145.52	122
93	ASN00075049	MAUDE (NAP NAP)	-34.45	144.17	75
94	ASN00075064	GROONGAL (GUNDALINE)	-34.48	145.55	107
95	ASN00075075	CONARGO (WILLURAH)	-35.00	145.09	99
96	ASN00076046	NYAH (YARRABY TANK)	-35.17	143.28	80
97	ASN00077030	NARRAPORT	-36.00	143.04	98
98	ASN00079016	WARRANOOKE	-36.73	142.73	150
99	ASN00079023	HORSHAM POLKEMMET RD	-36.66	142.07	128
100	ASN00079039	REDBANK	-36.91	143.34	260
101	ASN00079046	WARTOOK RESERVOIR	-37.09	142.43	440
102	ASN00080009	ST ARNAUD (COONOOER	-36.48	143.35	205
103	ASN00080015	ECHUCA AERODROME	-36.16	144.76	96
104	ASN00080029	LAKE MARMAL	-36.15	143.52	110
105	ASN00081008	COLBINABBIN	-36.54	144.78	144
106	ASN00084016	GABO ISLAND LIGHTHOUSE	-37.57	149.92	15.2
107	ASN00084030	ORBOST (COMPARISON)	-37.69	148.46	41
108	ASN00085096	WILSONS PROMONTORY	-39.13	146.42	95
109	ASN00004035	ROEBOURNE	-20.78	117.15	12

Tabla 6.3 Relaciones de los radios Momentos-L, estadístico de discordancia D_i y medidas de homogeneidad H y H^* de las diferentes regiones homogéneas.

Grupo	Numero	Código	Media anual	Desv. Estándar	t1	t3	t4	D	H	H*
G1	1	ASN00004035	3022.2	1644.5	0.309	0.113	0.091	1.19	-0.3858	0.7195
	2	ASN00005008	2904.7	1675.6	0.316	0.157	0.152	1.19		
	3	ASN00005016	2891.7	1884.1	0.357	0.205	0.129	0.10		
	4	ASN00005020	2429.1	1131.7	0.262	0.115	0.119	1.33		
G2	5	ASN00007027	2603.0	999.6	0.214	0.117	0.151	1.03	-0.7718	0.5897
	6	ASN00007064	2257.7	944.2	0.233	0.117	0.120	0.65		
	7	ASN00007197	2181.0	1013.9	0.258	0.169	0.121	1.60		
	8	ASN00012046	2383.4	1094.9	0.259	0.158	0.106	0.34		
	9	ASN00012176	2159.2	1058.8	0.273	0.171	0.109	1.07		
	10	ASN00013012	2548.7	1230.7	0.270	0.144	0.076	1.25		
G3	11	ASN00009507	6415.4	1449.9	0.124	0.092	0.183	1.15	-2.0565	0.2267
	12	ASN00010111	4313.4	1042.1	0.136	0.087	0.125	0.77		
	13	ASN00010579	4819.8	961.3	0.112	0.053	0.114	0.62		
	14	ASN00010626	4465.8	994.6	0.123	0.029	0.194	1.12		
G4	15	ASN00017031	1637.0	828.7	0.273	0.246	0.150	2.23	-1.2776	0.4519
	16	ASN00041095	7445.0	1720.1	0.131	0.021	0.097	0.57		
	17	ASN00044004	3522.5	1568.5	0.248	0.144	0.123	0.51		
	18	ASN00044010	4603.5	1750.2	0.206	0.152	0.144	0.60		
	19	ASN00044026	3727.8	1604.7	0.240	0.124	0.137	1.50		
	20	ASN00044166	4581.2	1512.8	0.181	0.130	0.172	1.25		
	21	ASN00046037	2327.4	1335.2	0.309	0.229	0.129	0.86		
	22	ASN00048168	4608.3	1612.2	0.197	0.128	0.108	0.99		
	23	ASN00056032	8494.0	1951.8	0.129	0.099	0.123	0.88		
	24	ASN00058012	14694.9	3699.0	0.138	0.162	0.166	0.62		
G5	25	ASN00018079	3763.9	910.0	0.137	0.068	0.102	1.32	0.5468	1.4281
	26	ASN00019062	3702.5	996.5	0.149	0.116	0.153	1.30		
	27	ASN00021043	3522.2	1034.0	0.161	0.122	0.160	1.33		
	28	ASN00021062	5068.3	1212.4	0.134	0.013	0.151	0.32		
	29	ASN00021086	3180.0	924.3	0.161	0.049	0.162	0.73		
G6	30	ASN00061055	10922.9	2493.3	0.129	0.119	0.099	1.31	-0.2611	0.8652
	31	ASN00061086	6499.3	1600.2	0.140	-0.035	0.088	1.19		
	32	ASN00063118	13155.8	3443.3	0.147	0.085	0.113	1.16		
	33	ASN00047000	2386.1	1155.3	0.246	0.196	0.235	1.14		
	34	ASN00049008	2973.0	1144.2	0.208	0.170	0.157	0.20		
G7	35	ASN00062013	6469.1	1947.5	0.167	0.097	0.114	0.42	-1.3795	0.4183
	36	ASN00063039	13861.9	3757.8	0.153	0.034	0.114	1.23		
	37	ASN00065000	5764.7	1757.8	0.169	0.080	0.143	0.27		
	38	ASN00065022	6727.5	1848.4	0.152	0.054	0.178	1.74		
	39	ASN00065026	5904.8	1811.7	0.173	0.105	0.105	0.76		
	40	ASN00066062	12040.0	3127.8	0.145	0.121	0.130	1.51		
	41	ASN00067019	8651.8	2531.9	0.161	0.122	0.141	1.44		
	42	ASN00075032	3711.2	1285.3	0.193	0.086	0.147	0.56		
	43	ASN00047013	2401.0	1009.1	0.221	0.167	0.186	1.66		
	44	ASN00047024	2879.4	1227.2	0.229	0.197	0.183	0.38		
	45	ASN00050028	4747.4	1692.4	0.201	0.101	0.099	1.02		
G8	46	ASN00022008	4999.1	1057.2	0.118	0.092	0.146	0.47	0.8634	1.6398
	47	ASN00022021	3934.7	1018.1	0.146	0.065	0.121	1.63		
	48	ASN00023721	6416.6	1359.0	0.121	0.031	0.096	0.95		
	49	ASN00023733	7535.5	1611.8	0.119	0.077	0.150	1.22		
	50	ASN00024511	4506.0	1133.0	0.138	0.063	0.199	0.92		
	51	ASN00025509	3804.1	962.2	0.142	0.015	0.150	0.80		
G9	52	ASN00028004	10415.9	3095.5	0.166	0.133	0.115	1.53	-0.3415	0.8226

	53	ASN00029004	7940.2	3757.7	0.250	0.259	0.173	0.88		
	54	ASN00029012	7492.0	2600.7	0.191	0.108	0.144	1.40		
	55	ASN00030045	4842.7	1948.6	0.223	0.148	0.154	0.68		
	56	ASN00032004	20993.9	5973.0	0.161	0.060	0.138	1.00		
G10	57	ASN00035019	6514.0	2362.7	0.204	0.126	0.132	1.33		
	58	ASN00036007	5021.6	2227.7	0.245	0.160	0.122	0.86		
	59	ASN00036066	4378.7	1982.1	0.245	0.209	0.139	1.33	-0.6489	0.5383
	60	ASN00037051	4171.3	1984.6	0.253	0.198	0.182	0.20		
	61	ASN00038003	2573.4	1536.1	0.312	0.274	0.180	1.28		
G11	62	ASN00035069	5267.7	1835.1	0.190	0.152	0.138	0.33		
	63	ASN00035070	6630.7	2044.3	0.162	0.175	0.195	1.62		
	64	ASN00036143	5262.4	1980.7	0.205	0.189	0.151	0.25		
	65	ASN00038024	2852.7	1491.2	0.280	0.180	0.139	0.47	0.5912	1.4454
	66	ASN00039085	12677.0	3052.2	0.133	0.089	0.139	1.67		
	67	ASN00040126	11241.8	3083.3	0.154	0.070	0.116	1.67		
G12	68	ASN00051034	4396.9	1664.1	0.209	0.080	0.154	1.35		
	69	ASN00051039	4395.1	1641.1	0.209	0.105	0.119	1.43		
	70	ASN00051057	3970.1	1445.2	0.205	0.044	0.103	1.39		
	71	ASN00053034	5632.4	1769.4	0.176	0.078	0.111	0.45		
	72	ASN00054003	6898.4	1776.1	0.146	0.012	0.090	1.21		
	73	ASN00055018	6381.9	1903.5	0.164	0.135	0.112	1.72		
	74	ASN00055023	6160.4	1678.9	0.154	-0.013	0.104	0.21		
	75	ASN00055057	7329.0	2092.2	0.161	0.074	0.108	0.37	-1.6146	0.3809
	76	ASN00056028	7724.5	1584.2	0.116	0.025	0.118	0.69		
	77	ASN00057023	8994.9	1945.6	0.119	0.085	0.176	2.78		
	78	ASN00059017	12139.1	3700.6	0.167	0.179	0.159	1.85		
	79	ASN00061051	8316.0	2042.7	0.139	0.048	0.105	0.64		
80	ASN00062009	6272.8	1751.3	0.155	0.050	0.140	0.12			
81	ASN00064008	7497.9	2287.5	0.169	0.062	0.148	0.48			
82	ASN00064013	5845.2	1826.4	0.175	0.105	0.106	0.30			
G13	83	ASN00069018	9614.0	3062.3	0.175	0.186	0.163	0.73		
	84	ASN00070009	6108.9	1622.7	0.147	0.137	0.122	1.48		
	85	ASN00084016	9291.2	2187.7	0.131	0.201	0.129	1.54	1.0833	1.8128
	86	ASN00084030	8461.3	1887.4	0.125	0.114	0.159	0.25		
	87	ASN00085096	10476.3	1841.3	0.101	0.034	0.055	0.67		
	88	ASN00086071	6458.0	1392.1	0.124	0.013	0.054	1.32		
G14	89	ASN00068007	7456.5	2410.9	0.178	0.135	0.130	2.42		
	90	ASN00068024	14244.9	4503.7	0.175	0.135	0.126	0.85		
	91	ASN00070032	8209.3	2303.6	0.158	0.051	0.122	1.39		
	92	ASN00072000	7951.5	2094.7	0.150	0.014	0.120	0.55		
	93	ASN00073025	5091.2	1476.3	0.165	0.074	0.112	0.16	-1.3893	0.3461
	94	ASN00074026	4401.8	1419.2	0.177	0.119	0.145	0.62		
	95	ASN00075049	3188.1	1188.2	0.196	0.148	0.218	0.82		
	96	ASN00075064	3813.8	1303.2	0.189	0.130	0.147	1.44		
	97	ASN00075075	3735.0	1331.1	0.196	0.104	0.177	1.03		
	98	ASN00076046	3300.7	1071.6	0.183	0.071	0.136	0.72		
G15	99	ASN00074034	5452.5	1647.4	0.170	0.111	0.117	0.79		
	100	ASN00077030	3558.7	1038.2	0.164	0.019	0.161	2.26		
	101	ASN00078038	3712.4	1028.9	0.154	0.018	0.161	0.78		
	102	ASN00078043	3974.8	1076.9	0.149	0.065	0.175	1.09		
	103	ASN00079016	4436.5	1132.1	0.143	0.012	0.147	0.50		
	104	ASN00079023	4481.3	1039.4	0.131	0.005	0.146	0.54	0.7649	1.5676
	105	ASN00079046	8736.7	1739.6	0.113	-0.023	0.099	1.01		
	106	ASN00080009	4310.1	1275.7	0.166	0.061	0.151	0.71		
	107	ASN00080015	4256.1	1285.3	0.166	0.054	0.170	2.36		
	108	ASN00080029	3745.5	1167.9	0.177	0.043	0.130	0.41		
	109	ASN00081008	5152.2	1506.7	0.164	0.006	0.117	0.45		

La identificación de regiones homogéneas es usualmente el paso más difícil y requiere la mayor cantidad de juicio o experiencia en su formación. Siguiendo a Hosking y Wallis (1997) el valor de H es usado para suponer el grado de heterogeneidad de las regiones. Si la región es no homogéneamente aceptable se debe considerar redefinirla. La región se puede dividir en dos o más subregiones, algunos sitios pueden ser retirados de la región o una reagrupación completamente diferente de los sitios. Sin embargo, cuando los grupos son no homogéneos puede ser debido a la presencia de un pequeño número de sitios atípicos en la región. En tales casos, las características del lugar de estos sitios deben ser examinadas cuidadosamente. Es posible volver a asignar estos sitios a regiones diferentes para conseguir nuevos grupos homogéneos, pero a veces puede parecer que no hay razón física del porque los sitios son atípicos siendo estos diferentes del resto de los sitios en la región.

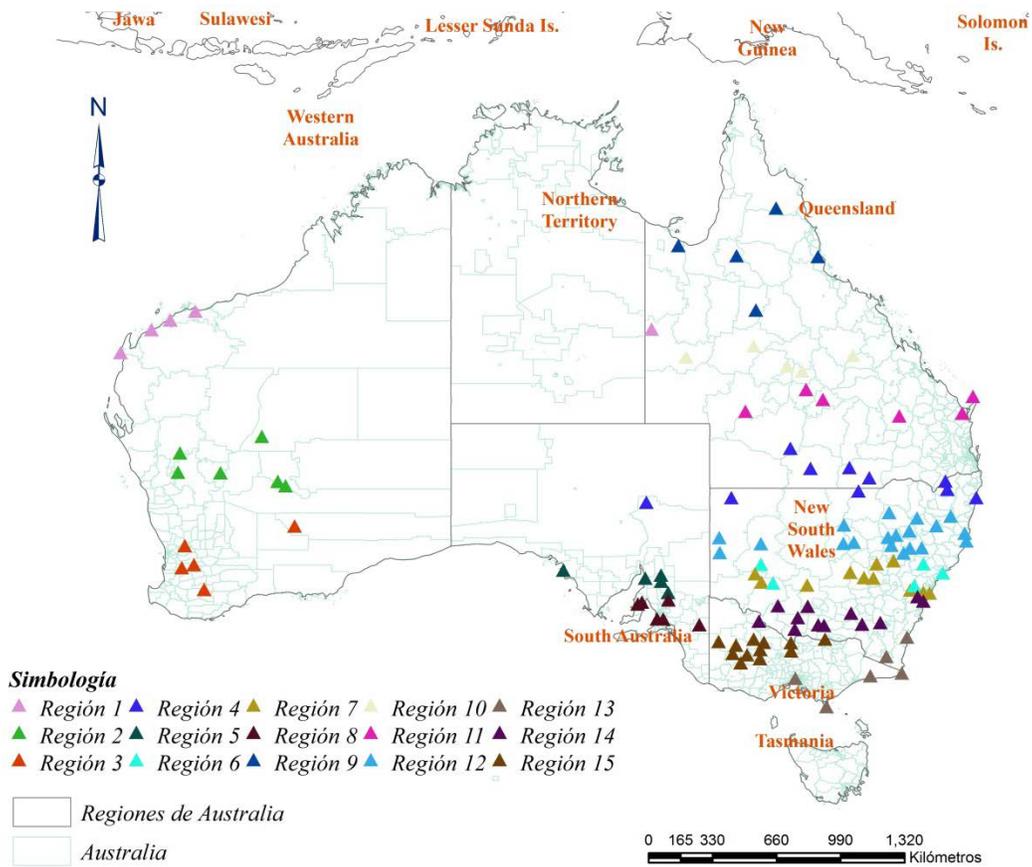


Figura 6.3. Delimitación de regiones homogéneas mediante Momentos-L.

El resultado de la formación de regiones homogéneas en las series de lluvia media mensual fue de 15 regiones. Esto fue mediante Momentos-L. La Figura 6.3 muestra las agrupaciones construidas.

6.7 Comparación de los modelos RNAs y RNAs-Regional

Se utilizaron dos modelos RNAs y RNAs-Regional para estimar la lluvia mensual en las diferentes regiones homogéneas de Australia. Se eligieron períodos de entrenamiento y validación para la aplicación. El 90% de los datos de toda la serie de lluvia fue utilizado para el entrenamiento y el 10% restante de los datos de toda la serie para validación. Se encontró que en el RMSE el 59% de las estaciones utilizadas presentó un mejor ajuste mediante el modelo RNAs-Regional que con el modelo tradicional de RNAs pues solo el 41% de las estaciones presentaron un adecuado ajuste.

El MAE obtenido fue mejor en el 53% para el RNAs-Regional y 47% para RNAs, el MAPE también fue menor para el modelo RNAs-Regional con 53% y solo el 47% para RNAs de las estaciones utilizadas. Por último el índice de concordancia d favoreció al modelo RNAs-Regional en un 75% de todas las estaciones estudiadas y en un 25% para RNAs. Regionalización combinada con RNAs reduce el RMSE, MAE, MAPE y el índice de concordancia d . en la siguiente tabla se presentan como ejemplo del análisis los resultados obtenidos para las estaciones ASN00007027, ASN00007064 y ASN00007197 las cuales pertenecen al grupo G2. Dichos resultados muestran la eficiencia del modelo RNAs-Regional ante el modelo RNAs tradicional.

Tabla 6.4. RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs- Regional para las regiones estudiadas (periodo de prueba).

	RNAs- Regional			RNAs		
	ASN00007027	ASN00007064	ASN00007197	ASN00007027	ASN00007064	ASN00007197
MSE	64969.80	58470.09	55809.14	73167.99	63819.16	58783.88
RMSE	254.89	241.81	236.24	270.50	252.62	242.45
MAE	186.04	180.36	174.51	195.60	186.26	175.39
MAPE	86.75	96.38	96.99	117.74	167.75	115.77
d	0.12	0.10	0.06	0.14	0.11	0.08

Las figuras 6.4 a 6.6 muestran el ajuste de los modelos RNAs y RNAs-Regional en la región G2. Los valores de las pruebas mediante RMSE, MAE, MAPE y el índice de concordancia d de los modelos RNAs y RNAs-Regional en dicha región indican que en todas las pruebas de bondad de ajuste excepto en el índice de concordancia el modelo RNAs-Regional siempre tuvo mejor desempeño que el modelo RNAs tradicional.

En este trabajo se encontró un desempeño eficiente en los modelos RNAs- Regionales ya que los resultados conseguidos demuestran su gran potencial. Con ello se propone una herramienta alternativa para pronóstico de lluvia acumulada mensual.

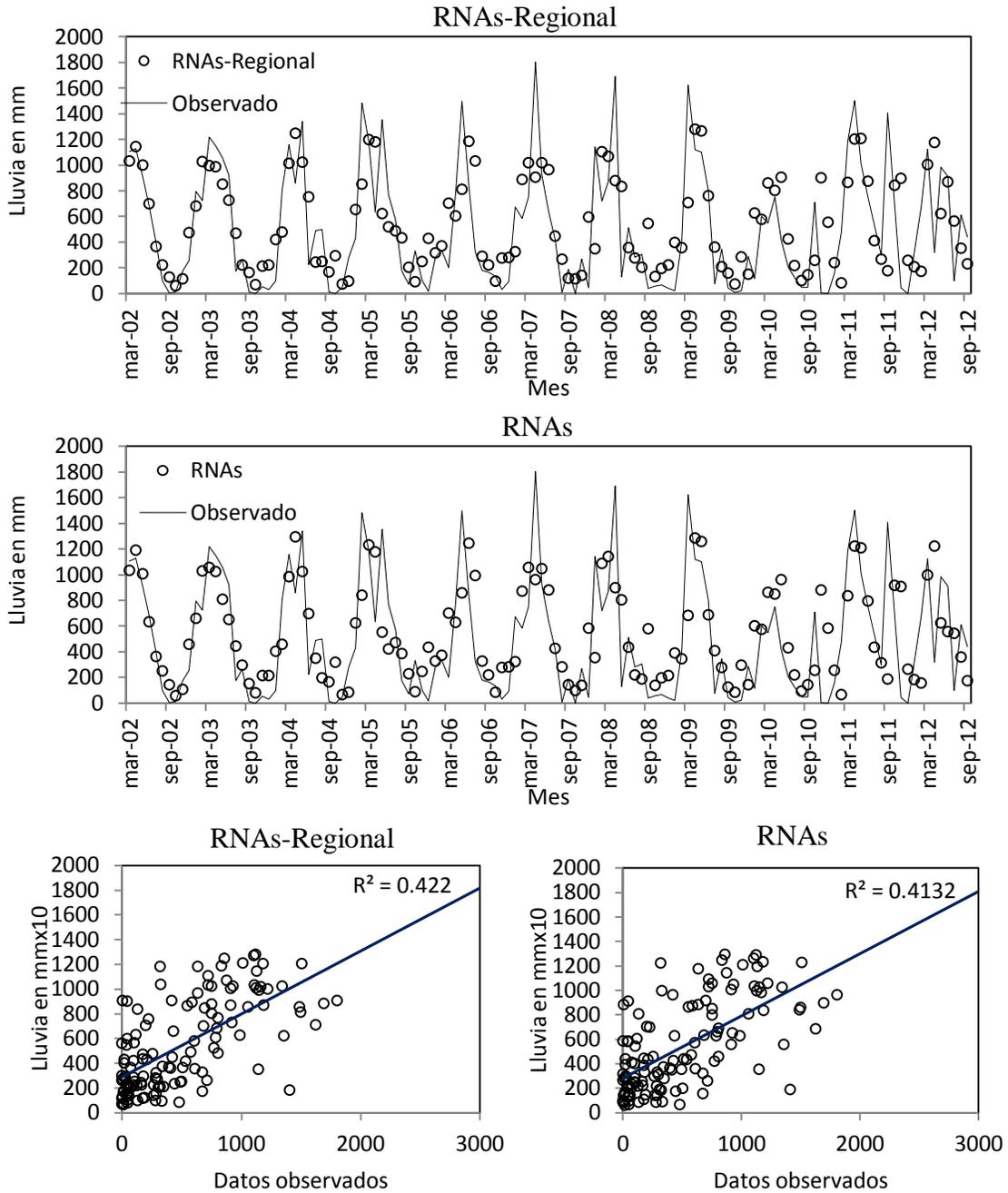


Figura 6.4. Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007027.

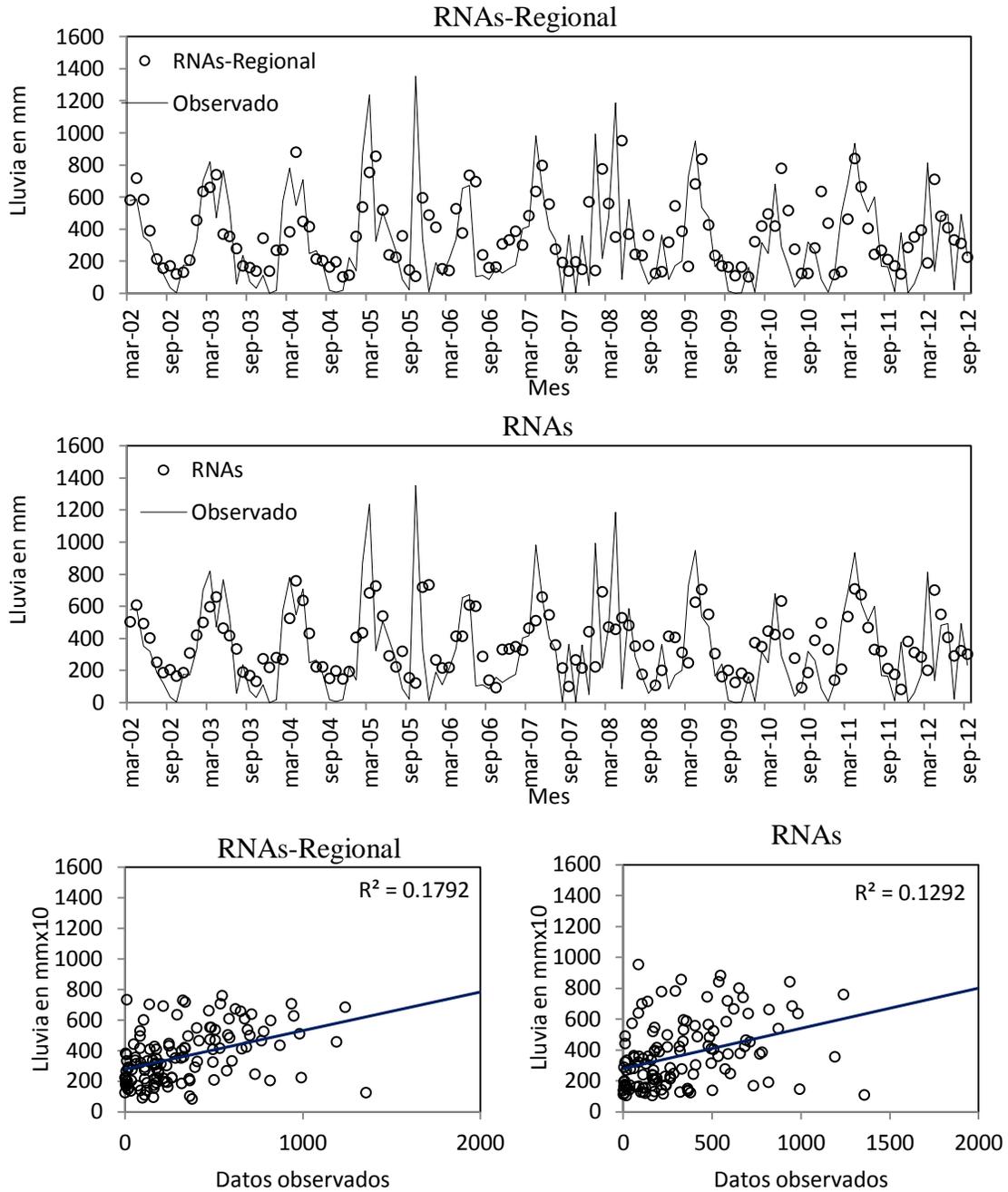


Figura 6.5. Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007064.

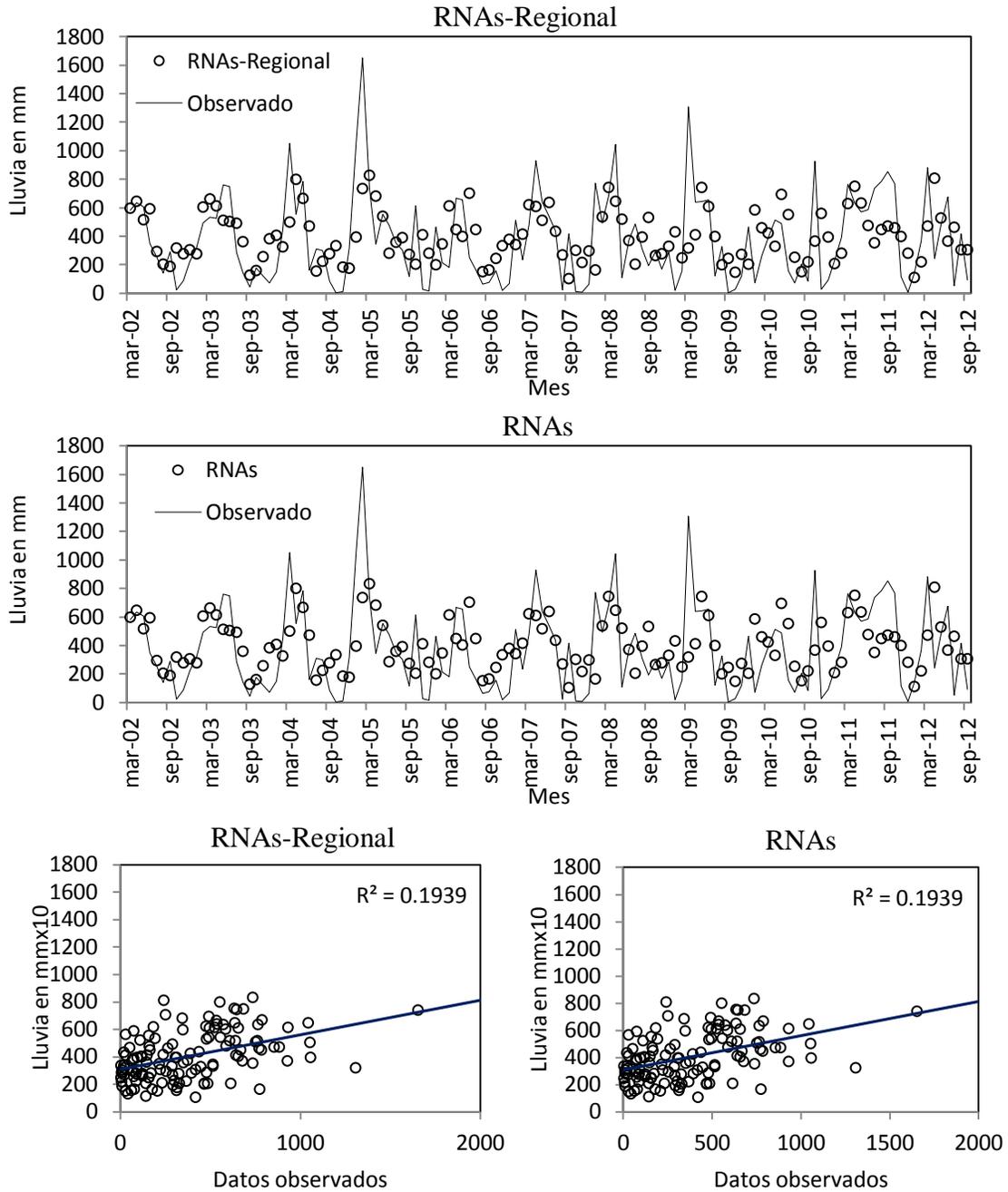


Figura 6.6. Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007197.

6.8 Comparación de los modelos Ensamblado y Ensamblado-Regional

En la comparación de los modelos Ensamblados para predicción de lluvias se utilizaron dos modelos de predicción. Un modelo Ensamblado basado en RNAs y un modelo Ensamblado-Regional también basado en RNAs. El modelo Ensamblado se puede considerar como tradicional y el modelo regional es el propuesto en este trabajo. Ambos modelos se utilizaron para estimar lluvia mensual en las diferentes regiones homogéneas de Australia. El entrenamiento de los modelos fue con el 90% de los datos de toda la serie de lluvia y el 10% restante de los datos se utilizó para validación. Los resultados indican que en el RMSE el 57% de las estaciones utilizadas presentaron mejor ajuste mediante el modelo Ensamblado regional que con el modelo Ensamblado tradicional ya que solo el 43% de las estaciones presentaron un adecuado ajuste.

El MAE obtenido fue mejor en el 56% para el Ensamblado-Regional y 44% para Ensamblado, el MAPE también fue menor para el modelo Ensamblado-Regional con 58% y solo el 42% para el Ensamblado de las estaciones utilizadas. Por último el índice de concordancia d favoreció al modelo Ensamblado-Regional en un 83% de todas las estaciones estudiadas y en un 17% para Ensamblado. Regionalización combinada con Ensamblado reduce el RMSE, MAE, MAPE y el índice de concordancia d . En la siguiente tabla se presentan como ejemplo del análisis los resultados obtenidos para las estaciones ASN00007027, ASN00007064 y ASN00007197 las cuales pertenecen al grupo G2. Dichos resultados muestran la eficiencia del modelo Ensamblado-Regional ante el modelo Ensamblado tradicional.

Tabla 6.5. RMSE, MAE e índice de concordancia d de los modelos Ensamblado y Ensamblado-Regional para las regiones estudiadas (periodo de prueba).

	Ensamblado-Regional			Ensamblado		
	ASN00007027	ASN00007064	ASN00007197	ASN00007027	ASN00007064	ASN00007197
MSE	56084.57	50065.14	52688.41	56817.95	50479.01	96053.07
RMSE	236.82	223.75	229.531	238.37	224.68	309.92
MAE	169.30	156.54	568.88	172.21	150.22	175.00
MAPE	78.67	86.63	86.88	80.65	109.84	95.12
d	0.39	0.37	0.00	0.38	0.38	0.19

Las figuras 6.7 a 6.9 muestran el ajuste de los modelos Ensamblado y Ensamblado-Regional en la región G2. Los valores de las pruebas mediante RMSE, MAE, MAPE y el índice de concordancia d de los modelos RNAs y RNAs-Regional en dicha región indican que en todas las pruebas de bondad de ajuste excepto en el índice de concordancia el modelo Ensamblado-Regional siempre tuvo mejor desempeño que el modelo Ensamblado tradicional. Se obtuvo un desempeño eficiente en este trabajo mediante el modelo Ensamblado-Regional ya que los resultados conseguidos demuestran su gran potencial. Con ello se propone una herramienta alternativa para pronóstico de lluvia acumulada mensual.

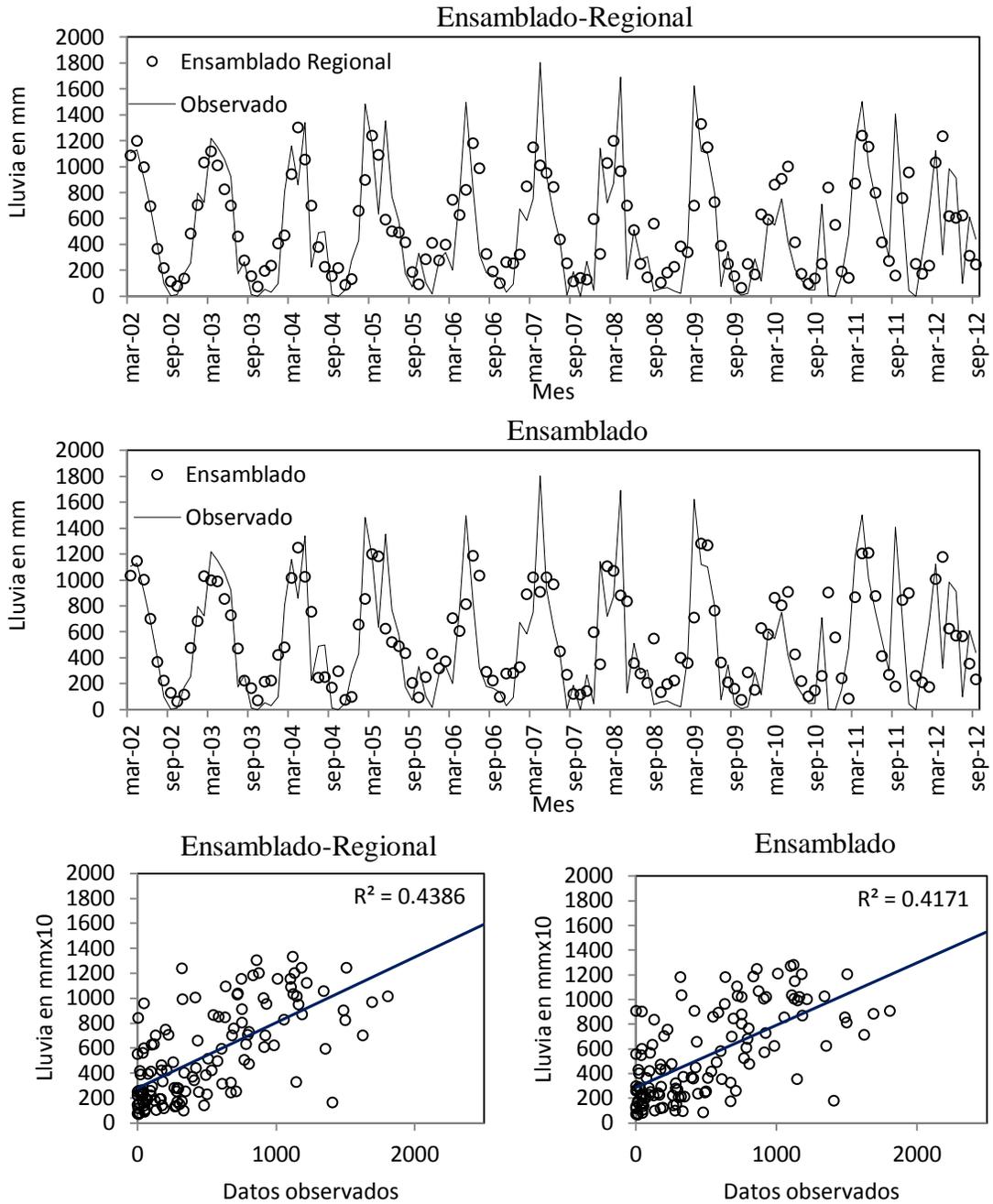


Figura 6.7. Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007027.

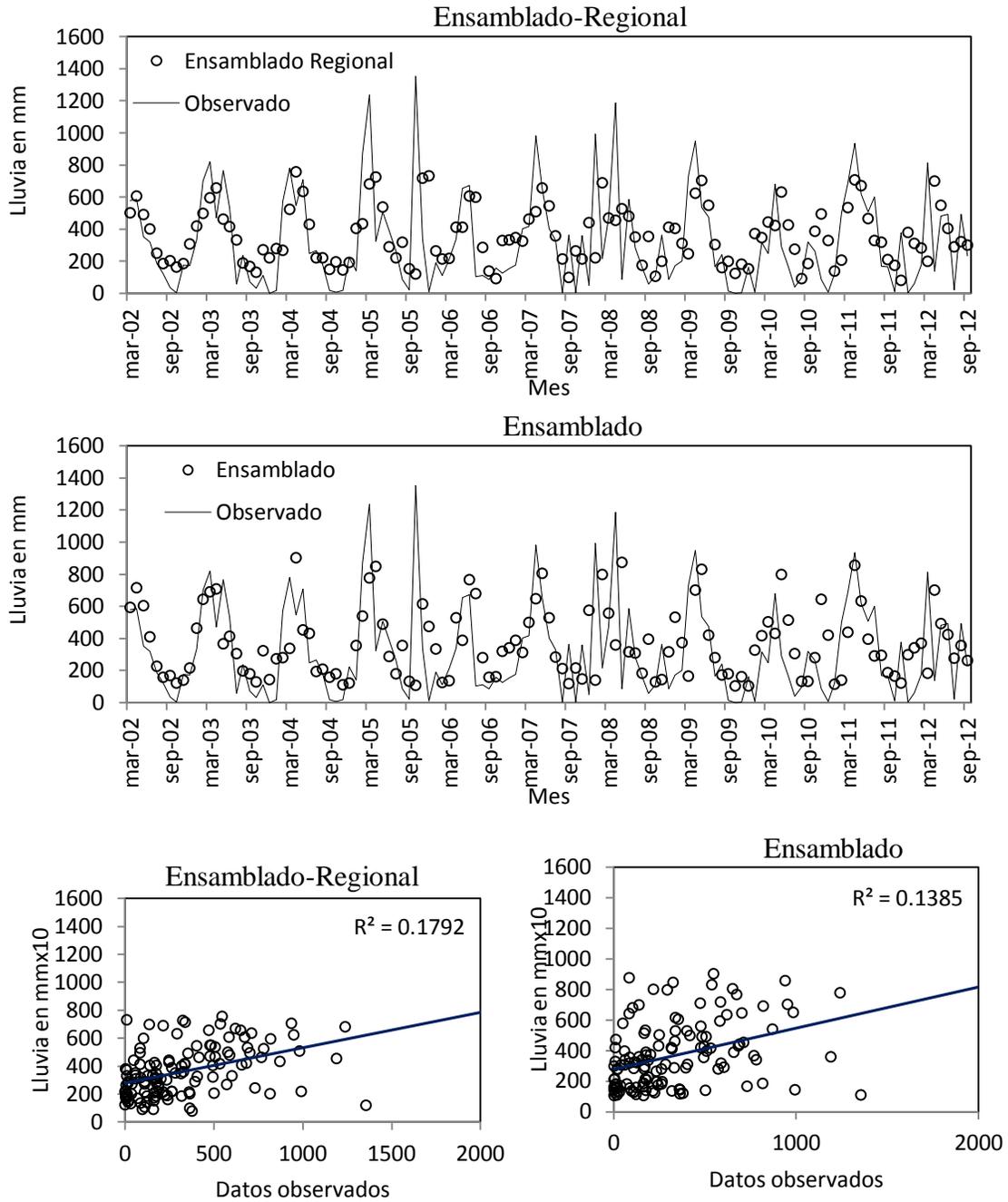


Figura 6.8. Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007064.

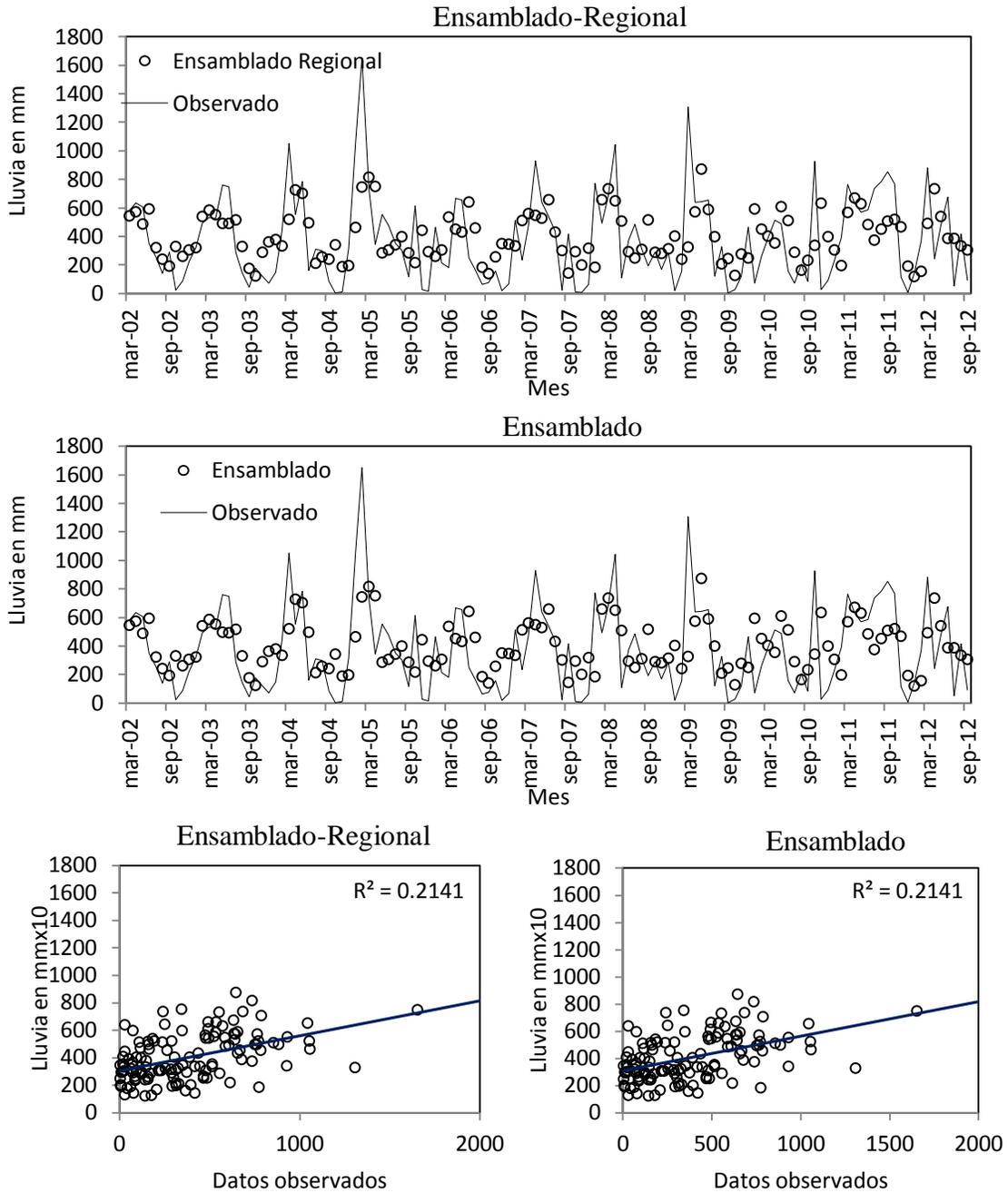


Figura 6.9. Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007197.

6.9 Comparación de los modelos Modular y Modular-Regional

En los modelos Modular y Modular-Regional que fueron utilizados para estimar la lluvia mensual en las diferentes regiones homogéneas de Australia, se eligieron para los períodos de entrenamiento y validación, el 90% y el 10% respectivamente. Los resultados demostraron que en el RMSE el 56% de las estaciones utilizadas presentaron un mejor ajuste mediante el modelo Modular-Regional que con el modelo tradicional Modular. Siendo en este último el 44% de las estaciones.

El MAE obtenido fue mejor en el 55% para el Modular-Regional y 45% para Modular, el MAPE también fue menor para el modelo Modular-Regional con 54% y solo el 46% para Modular de las estaciones utilizadas. Por último el índice de concordancia d favoreció al modelo Modular-Regional en un 69% de todas las estaciones estudiadas y en un 31% para RNAs. La regionalización combinada con RNAs reduce el RMSE, MAE, MAPE y el índice de concordancia d . En la siguiente tabla se presentan como ejemplo del análisis los resultados obtenidos para las estaciones ASN00007027, ASN00007064 y ASN00007197 las cuales pertenecen al grupo G2. Dichos resultados muestran la eficiencia del modelo Modular-Regional ante el modelo Modular tradicional.

Tabla 6.6. RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs- Regional para las regiones estudiadas (periodo de prueba).

	Modular-Regional			Modular		
	ASN00007027	ASN00007064	ASN00007197	ASN00007027	ASN00007064	ASN00007197
MSE	56508.85	51407.33	106167.97	56817.81	51441.16	93211.32
RMSE	237.72	226.73	325.83	238.36	226.81	305.31
MAE	169.56	154.48	177.07	172.26	156.42	171.88
MAPE	75.81	99.82	97.83	80.68	93.40	93.18
d	0.39	0.37	0.17	0.38	0.36	0.20

Las figuras 6.10 a 6.12 muestran el ajuste de los modelos Modular y Modular-Regional en la región G2. Los valores de las pruebas mediante RMSE, MAE, MAPE y el índice de concordancia d de los modelos Modular y Modular-Regional en dicha región indican que en todas las pruebas de bondad de ajuste excepto en el índice de concordancia el modelo Modular-Regional siempre tuvo mejor desempeño que el modelo Modular tradicional.

Se encontró un desempeño eficiente en este trabajo mediante los modelos Modular-Regional ya que los resultados conseguidos demuestran su gran potencial. Con ello se propone una herramienta alternativa para pronóstico de lluvia acumulada mensual.

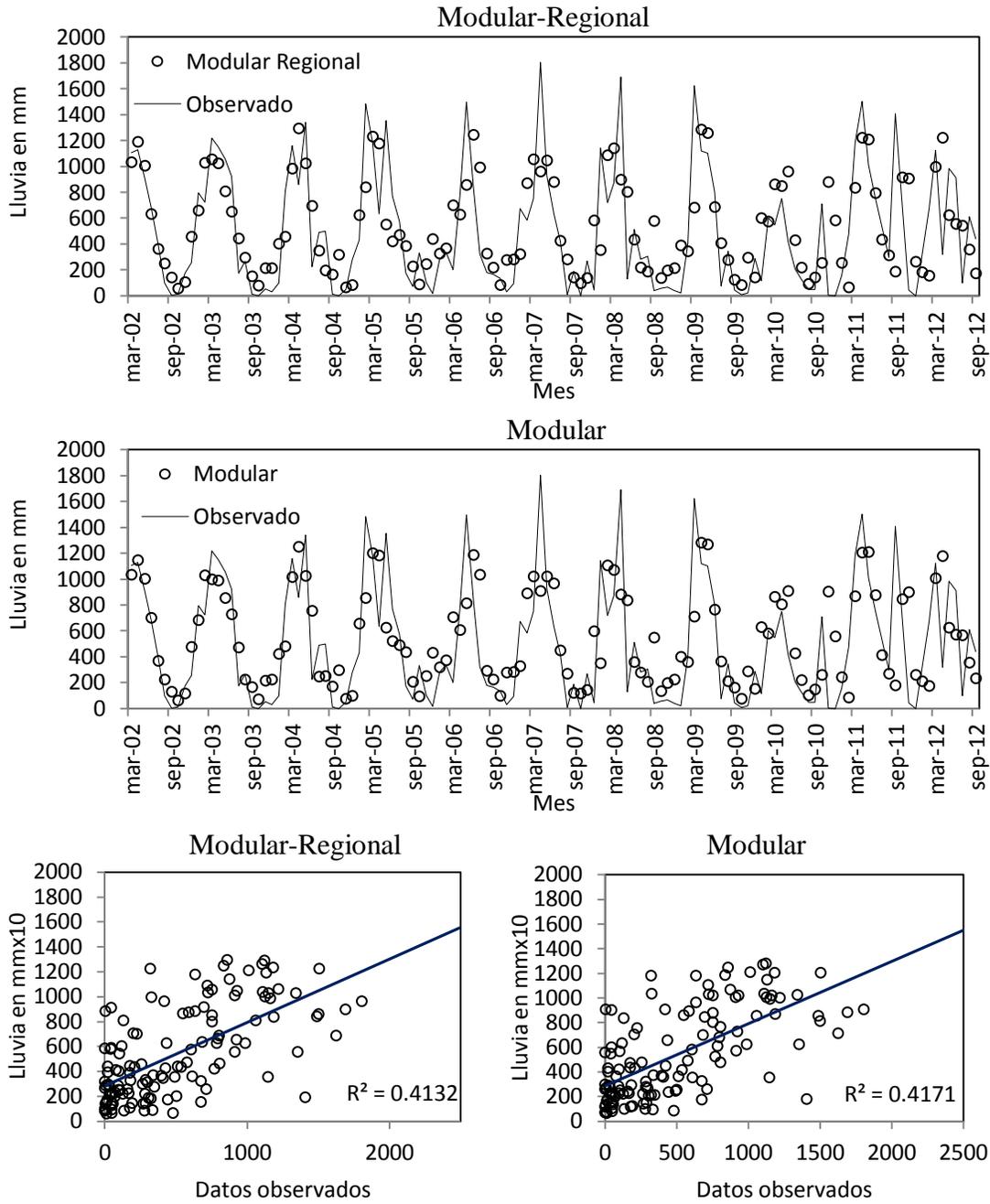


Figura 6.10. Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007027.

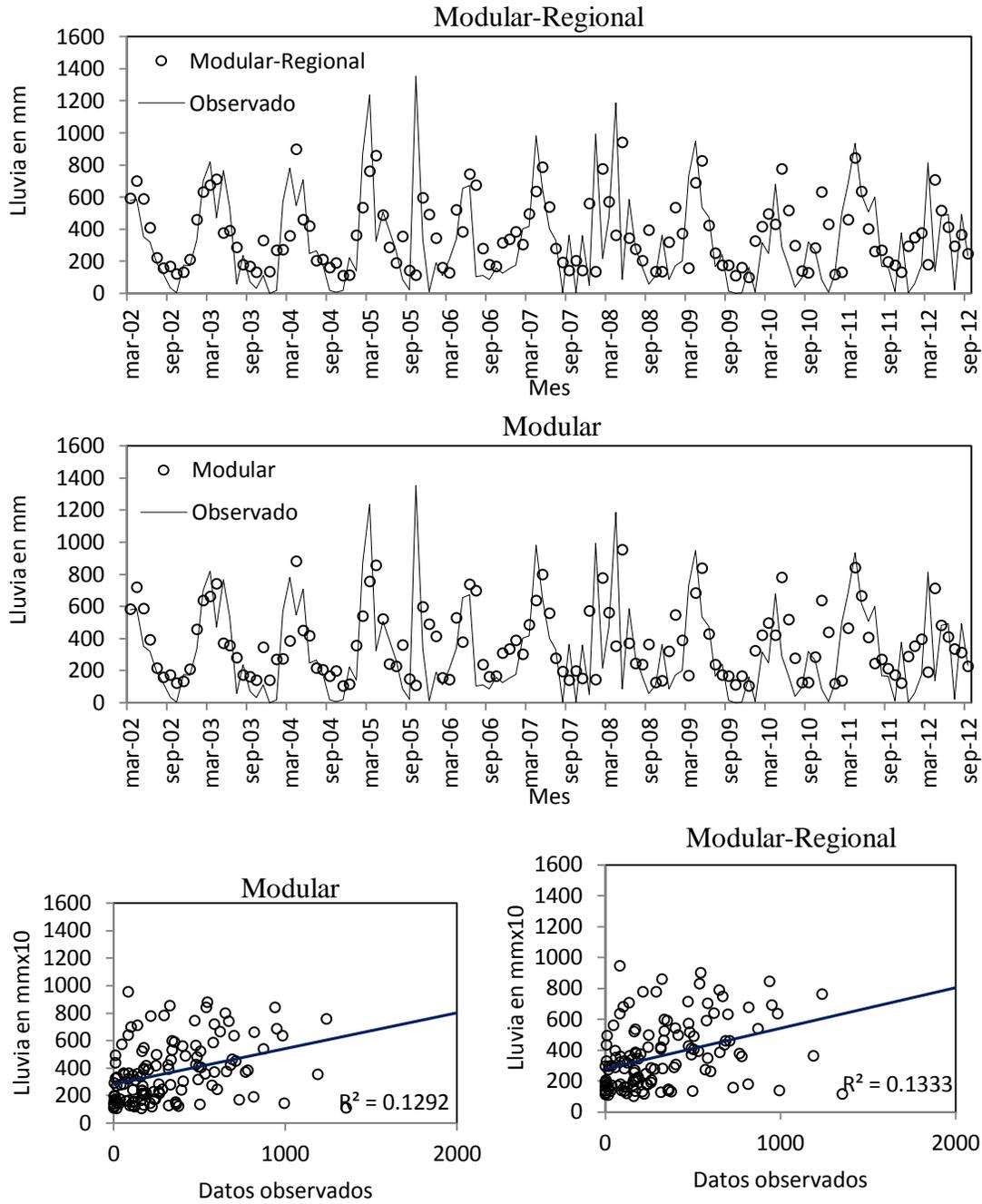


Figura 6.11. Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007064.

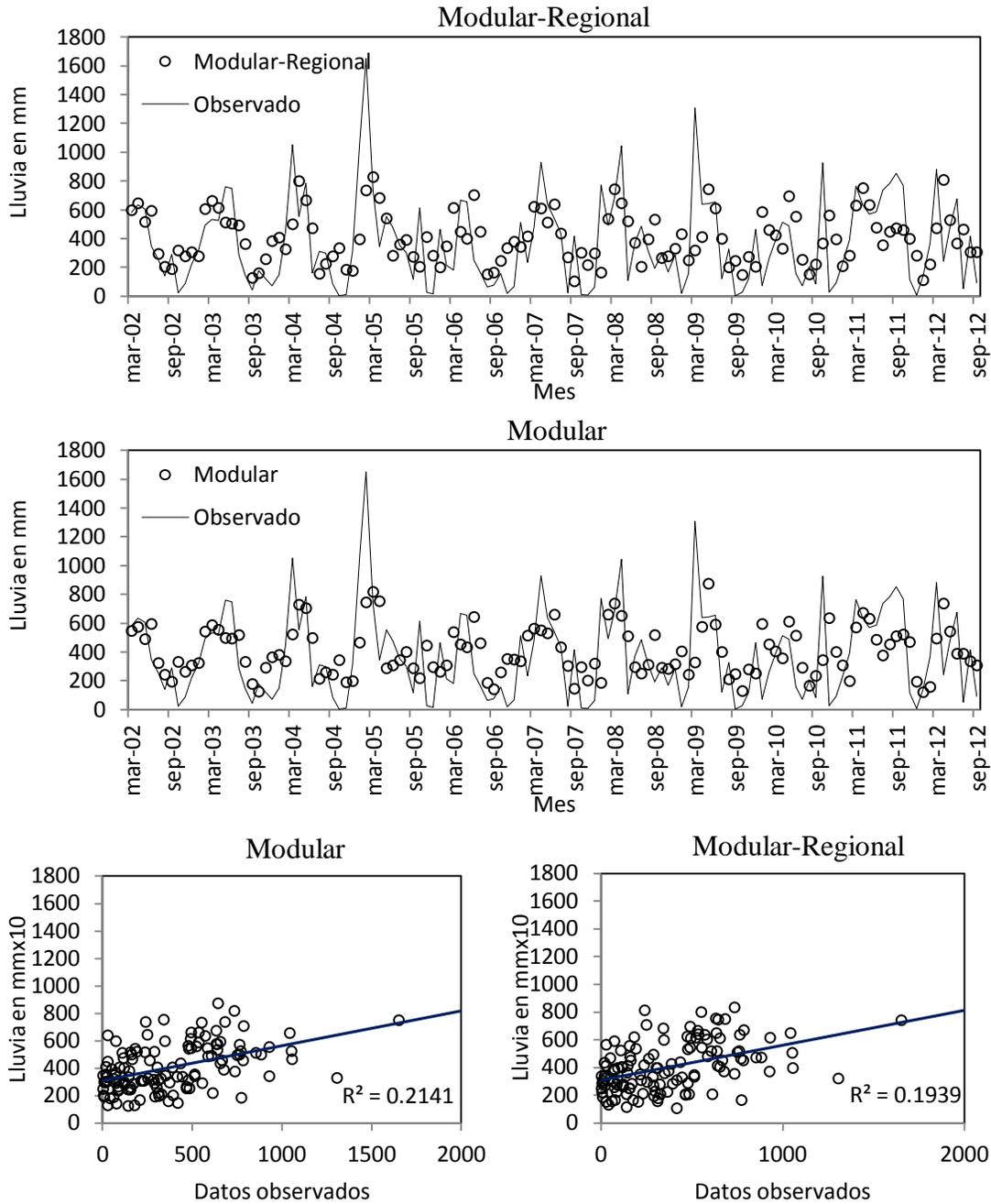


Figura 6.12. Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007197.

6.10 Comparación de modelos de predicción en escurrimientos

Los modelos se aplicaron a escurrimientos mensuales del Suroeste y Sureste de Canadá. Los ríos canadienses descargan alrededor del 9% del suministro de agua renovable en el mundo, por esta razón el pronóstico de escurrimientos de ríos son un componente importante de la gestión de recursos hídricos y de vital importancia tanto desde el punto de vista de suministro del agua y protección contra inundaciones. La gestión del agua se puede lograr con una planificación adecuada de las operaciones de los sistemas fluviales, de embalses y con un servicio de predicción de escurrimientos eficiente. En este estudio, se obtuvieron los datos de escurrimientos mensuales de 90 estaciones en las regiones este y oeste de Canadá a partir de los datos hidrométricos del CD-ROM de Environment Canadá (HYDAT, 2013). La ubicación de las estaciones se muestra en la Figura 6.13. Los datos observados comprenden entre 99 y 114 años de longitud (1188 y 1368 meses) con observaciones comprendidas entre el periodo 1900 y 2008 con máxima longitud y entre 1915 a 2008 para la mínima longitud. Los datos observados comienzan en enero y el último mes es diciembre. Las características de las estaciones se presentan en la Tabla 6.7.

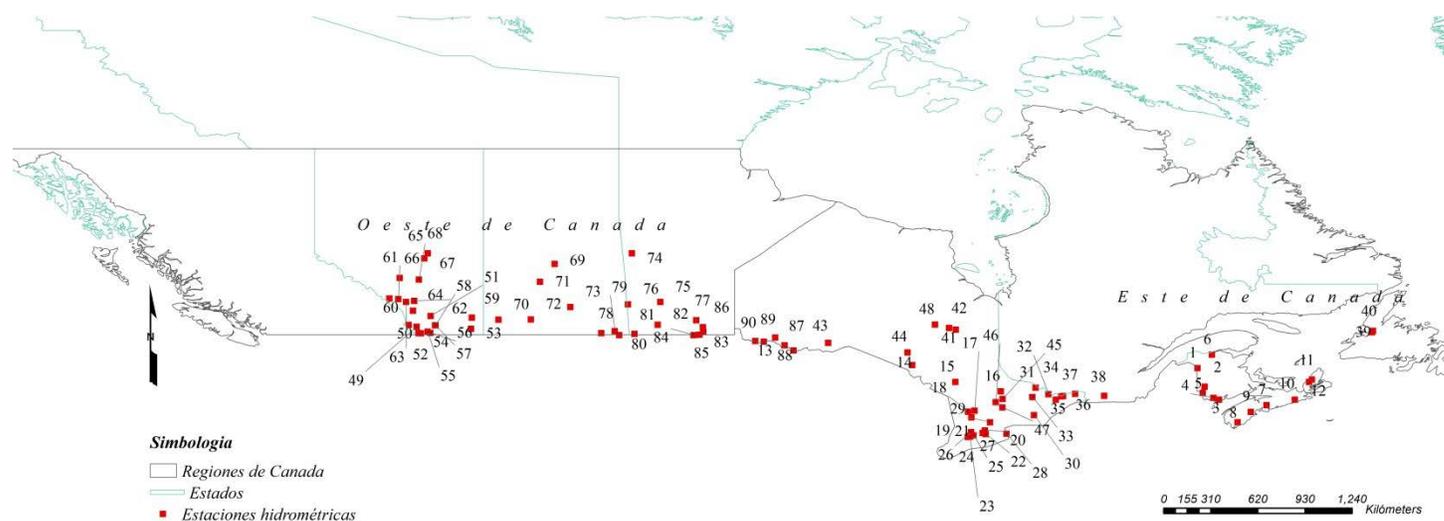


Figura 6.13 Localización de las estaciones hidrométricas en el sureste y suroeste de Canadá.

Tabla 6.7. Información de estaciones hidrométricas en el sureste y suroeste de Canadá.

Número	Código	Año	Nombre	Latitud	Longitud	Área de drenaje(km ²)	
1	01AF002	1930	2008	SAINT JOHN RIVER AT GRAND FALLS	47.04	-67.74	21900
2	01AK001	1918	2008	SHOGOMOC STREAM NEAR TRANS CANADA	45.94	-67.32	234
3	01AQ001	1916	2009	LEPREAU RIVER AT LEPREAU	45.17	-66.47	239
4	01AQ002	1917	2009	MAGAGUADAVIC RIVER AT ELMCROFT	45.27	-66.81	1420
5	01AR004	1928	2008	ST. CROIX RIVER AT VANCEBORO	45.57	-67.43	1080
6	01BE001	1918	2009	UPSALQUITCH RIVER AT UPSALQUITCH	47.83	-66.89	2270
7	01DG003	1921	2009	BEAVERBANK RIVER NEAR KINSAC	44.85	-63.66	97
8	01EC001	1915	2009	ROSEWAY RIVER AT LOWER OHIO	43.84	-65.37	495
9	01EF001	1915	2009	LAHAVE RIVER AT WEST NORTHFIELD	44.45	-64.59	1250
10	01EO001	1915	2009	ST. MARYS RIVER AT STILLWATER	45.17	-61.98	1350
11	01FB001	1916	2009	NORTHEAST MARGAREE RIVER AT MARGAREE	46.37	-60.98	368
12	01FB003	1918	2009	SOUTHWEST MARGAREE RIVER NEAR UPPER	46.22	-61.14	357
13	05PC019	1905	2009	RAINY RIVER AT FORT FRANCES	48.61	-93.40	38600

14	02BE002	1935	2008	MONTREAL RIVER NEAR MONTREAL RIVER	47.21	-84.62	2880
15	02CE002	1915	2009	AUX SABLES RIVER AT MASSEY	46.21	-82.07	1350
16	02EA005	1915	2009	NORTH MAGNETAWAN RIVER NEAR BURK'S	45.67	-79.38	321
17	02FB007	1915	2009	SYDENHAM RIVER NEAR OWEN SOUND	44.52	-80.93	181
18	02FC001	1914	2008	SAUGEEN RIVER NEAR PORT ELGIN	44.46	-81.33	3960
19	02FC002	1914	2009	SAUGEEN RIVER NEAR WALKERTON	44.12	-81.12	2150
20	02GA003	1913	2009	GRAND RIVER AT GALT	43.35	-80.32	3520
21	02GA010	1913	2008	NITH RIVER NEAR CANNING	43.19	-80.46	1030
22	02GB001	1913	2008	GRAND RIVER AT BRANTFORD	43.13	-80.27	5210
23	02GD001	1915	2009	THAMES RIVER NEAR EALING	42.97	-81.21	1340
24	02GD003	1915	2009	NORTH THAMES RIVER BELOW FANSHAWE DAM	43.04	-81.18	1450
25	02GD004	1938	2009	MIDDLE THAMES RIVER AT THAMESFORD	43.06	-80.99	306
26	02GD005	1938	2009	NORTH THAMES RIVER AT ST. MARYS	43.26	-81.15	1080
27	02GE002	1922	2009	THAMES RIVER AT BYRON	42.96	-81.33	3110
28	02HA003	1860	2009	NIAGARA RIVER AT QUEENSTON	43.16	-79.05	686000
29	02HB001	1915	2009	CREDIT RIVER NEAR CATARACT	43.84	-80.02	205
30	02HL001	1915	2008	MOIRA RIVER NEAR FOXBORO	44.25	-77.42	2620
31	02KB001	1915	2009	PETAWAWA RIVER NEAR PETAWAWA	45.89	-77.32	4120
32	02KC009	1921	2009	BONNECHERE RIVER NEAR CASTLEFORD	45.50	-76.56	2380
33	02KD004	1930	2009	MADAWASKA RIVER AT PALMER RAPIDS	45.33	-77.52	5800
34	02KF005	1915	2009	OTTAWA RIVER AT BRITANNIA	45.36	-75.81	90900
35	02KF006	1918	2009	MISSISSIPPI RIVER AT APPLETON	45.18	-76.12	2900
36	02LA004	1933	2008	RIDEAU RIVER AT OTTAWA	45.38	-75.70	3830
37	02LB005	1915	2008	SOUTH NATION RIVER NEAR PLANTAGENET	45.52	-74.98	3810
38	02OJ007	1937	2009	RICHELIEU (RIVIERE) AUX RAPIDES FRYERS	45.40	-73.26	22000
39	02YK001	1925	2009	HUMBER RIVER AT GRAND LAKE OUTLET	49.16	-57.42	5020
40	02YL001	1928	2009	UPPER HUMBER RIVER NEAR REIDVILLE	49.24	-57.36	2110
41	04LD001	1920	2009	GROUNDHOG RIVER AT FAUQUIER	49.31	-82.04	11900
42	04LF001	1918	2009	KAPUSKASING RIVER AT KAPUSKASING	49.41	-82.44	6760
43	02AB006	1926	2009	KAMINISTIQUIA RIVER AT KAMINISTIQUIA	48.53	-89.60	6475
44	02BD002	1920	2009	MICHIPICOTEN RIVER AT HIGH FALLS	47.96	-84.90	5130
45	02EB004	1915	2009	NORTH BRANCH MUSKOKA RIVER AT PORT	45.21	-79.28	1390
46	02EB006	1937	2009	MUSKOKA RIVER BELOW BALA	45.02	-79.68	4770
47	02EC002	1915	2009	BLACK RIVER NEAR WASHAGO	44.71	-79.28	1520
48	04LJ001	1920	2009	MISSINAIBI RIVER AT MATTICE	49.61	-83.27	8940
49	05AA004	1910	2010	PINCHER CREEK AT PINCHER CREEK	49.49	-113.95	158
50	05AA008	1910	2010	CROWSNEST RIVER AT FRANK	49.60	-114.41	403
51	05AC003	1918	2009	LITTLE BOW RIVER AT CARMANGAY	50.13	-113.12	2778
52	05AD003	1908	2009	WATERTON RIVER NEAR WATERTON PARK	49.11	-113.84	613
53	05AF010	1911	2009	MANYBERRIES CREEK AT BRODIN'S FARM	49.36	-110.73	338
54	05AD005	1911	2009	BELLY RIVER NEAR MOUNTAIN VIEW	49.10	-113.70	319
55	05AE002	1909	2009	LEE CREEK AT CARDSTON	49.20	-113.30	312
56	05AE005	1911	2010	ROLPH CREEK NEAR KIMBALL	49.13	-113.14	222
57	05AE006	1911	2009	ST. MARY RIVER NEAR LETHBRIDGE	49.57	-112.85	3527
58	05AE016	1919	2009	POTHOLE CREEK AT RUSSELL'S RANCH	49.57	-112.83	770
59	05AH005	1910	2009	SEVEN PERSONS CREEK AT MEDICINE HAT	50.02	-110.68	3276
60	05BC001	1910	2009	SPRAY RIVER AT BANFF	51.16	-115.55	751
61	05BE004	1923	2009	BOW RIVER NEAR SEEBE	51.12	-115.03	5170
62	05BJ001	1908	2009	ELBOW RIVER BELOW GLENMORE DAM	51.01	-114.09	1236
63	05BJ004	1934	2009	ELBOW RIVER AT BRAGG CREEK	50.95	-114.57	791
64	05BL007	1911	2010	STIMSON CREEK NEAR PEKISKO	50.43	-114.17	236
65	05CC002	1912	2009	RED DEER RIVER AT RED DEER	52.28	-113.82	11609
66	05DC001	1913	2009	NORTH SASKATCHEWAN RIVER NEAR ROCKY	52.38	-114.94	11007
67	05DF001	1911	2009	NORTH SASKATCHEWAN RIVER AT EDMONTON	53.54	-113.49	28096
68	05EA001	1914	2009	STURGEON RIVER NEAR FORT SASKATCHEWAN	53.83	-113.28	3247
69	05GG001	1910	2009	NORTH SASKATCHEWAN RIVER AT PRINCE	53.20	-105.77	131000
70	05HA003	1908	2009	BEAR CREEK NEAR PIAPOT	49.90	-109.11	253

71	05HG001	1911	2009	SOUTH SASKATCHEWAN RIVER AT SASKATOON	52.14	-106.64	141000
72	05JB001	1914	2009	NOTUKEU CREEK NEAR VANGUARD	49.92	-107.19	3540
73	05JF001	1911	2009	QU'APPELLE RIVER NEAR LUMSDEN	50.65	-104.85	17500
74	05KJ001	1913	2009	SASKATCHEWAN RIVER AT THE PAS	53.84	-101.21	389000
75	05LJ007	1922	2009	TURTLE RIVER NEAR LAURIER	50.95	-99.53	955
76	05ME001	1913	2009	ASSINIBOINE RIVER NEAR RUSSELL	50.81	-101.44	19400
77	05MJ001	1913	2009	ASSINIBOINE RIVER AT HEADINGLEY	49.87	-97.41	162000
78	05NB001	1911	2009	LONG CREEK NEAR ESTEVAN	49.10	-103.01	4840
79	05ND004	1913	2009	MOOSE MOUNTAIN CREEK NEAR OXBOW	49.23	-102.23	6050
80	05ND007	1930	2009	SOURIS RIVER NEAR SHERWOOD	48.99	-101.96	23100
81	05NF002	1935	2009	ANTLER RIVER NEAR MELITA	49.06	-101.05	3220
82	05NG001	1912	2009	SOURIS RIVER AT WAWANESA	49.60	-99.68	61100
83	05OC001	1912	2009	RED RIVER AT EMERSON	49.01	-97.22	102000
84	05OC004	1903	2008	PEMBINA RIVER AT NECHE	48.99	-97.56	8480
85	05OD001	1913	2009	ROSEAU RIVER NEAR DOMINION CITY	49.19	-96.98	5020
86	05OE001	1912	2009	RAT RIVER NEAR OTTERBURNE	49.46	-97.01	1420
87	05PA006	1921	2009	NAMAKAN RIVER AT OUTLET OF LAC LA CROIX	48.38	-92.18	13400
88	05PA012	1924	2009	BASSWOOD RIVER NEAR WINTON	48.08	-91.65	4510
89	05PB014	1914	2009	TURTLE RIVER NEAR MINE CENTRE	48.85	-92.72	4870
90	05PC018	1928	2009	RAINY RIVER AT MANITOU RAPIDS	48.63	-93.91	50200

6.11 Aplicaciones y resultados

Cuando no se cuenta con suficiente información es difícil extrapolar basándose en el comportamiento de la serie con la que se cuenta pues la incertidumbre es alta dependiendo de la cantidad de datos disponibles (Ashkar, et al.1993). Sin embargo este problema se puede resolver “paliando la carencia en el tiempo con la abundancia en el espacio” de acuerdo con Hosking (1997).

El desarrollo de los modelos regionales con L-momentos (Hosking y Wallis, 1993, 1997) incluye dos pasos: (1) delimitación regiones homogéneas en primera instancia detección de los grupos probables homogéneos mediante algún método datos utilizando la medida de discordancia D_i ; y las pruebas de homogeneidad utilizando la medida de heterogeneidad, H, (2) la selección del modelo de predicción. Estos dos pasos se han seguido para la delimitación de regiones homogéneas y aplicación del modelo en las regiones del oeste y del este de Canadá y de los métodos estadísticos utilizados se describen más adelante.

A continuación se utiliza el método de c-Medias para delimitar los grupos iniciales después se refinan mediante Momentos-L.

6.12 Análisis de grupos mediante c-Medias

En el método de las c-Medias los centros de los cúmulos se inicializan seleccionándolos al azar a partir del conjunto de datos. Después la formación de los grupos es mediante la asignación de cada punto al centro más cercano. Cuando el conjunto de datos se ha asignado, la posición promedio de los puntos de datos dentro de cada grupo se calcula y del centro del cúmulo se mueve entonces a la posición media. Este proceso de asignación y de promediados se repite hasta que todos los centros de los grupos ya no se mueven. El proceso entonces se dice que han convergido.

Después de examinar los datos y teniendo en cuenta los períodos en los que hubo vacíos, inicialmente se propuso la formación de grupos a través de la identificación de regiones mediante método de las c-Medias, la determinación de regiones se realiza por el método determinante. Las características tomadas para formar los grupos fueron físicas, tales como la longitud, la latitud, la elevación, el escurrimiento medio anual así como el área de drenaje de las subcuencas de aporte en cada estación.

Tabla 6.8. Transformación de las características de los sitios.

Características del sitio (X)	Variable de agrupación (Y)
Área de drenaje (km ²)	$Y = \log(X) \times 3 / s.d.(\log X)$
Elevación de la estación (m)	$Y = \sqrt{X} / s.d.(\sqrt{X})$
Latitud de la estación (grados)	$Y = X / s.d.(X)$
Longitud de la estación (grados)	$Y = X / s.d.(X)$
Escurrimiento medio anual (m ³)	$Y = X / s.d.(X)$

Como resultado de la agrupación mediante c-Medias el número de grupos encontrados fue de 6. El Software Matlab R.12 se utilizó en la agrupación de la serie de escurrimientos y en la determinación de una primera aproximación de las regiones homogéneas. La Figura 6.14 muestra las agrupaciones construidas mediante c-Medias estos grupos son los que han de conformar las entradas en el modelo regional.

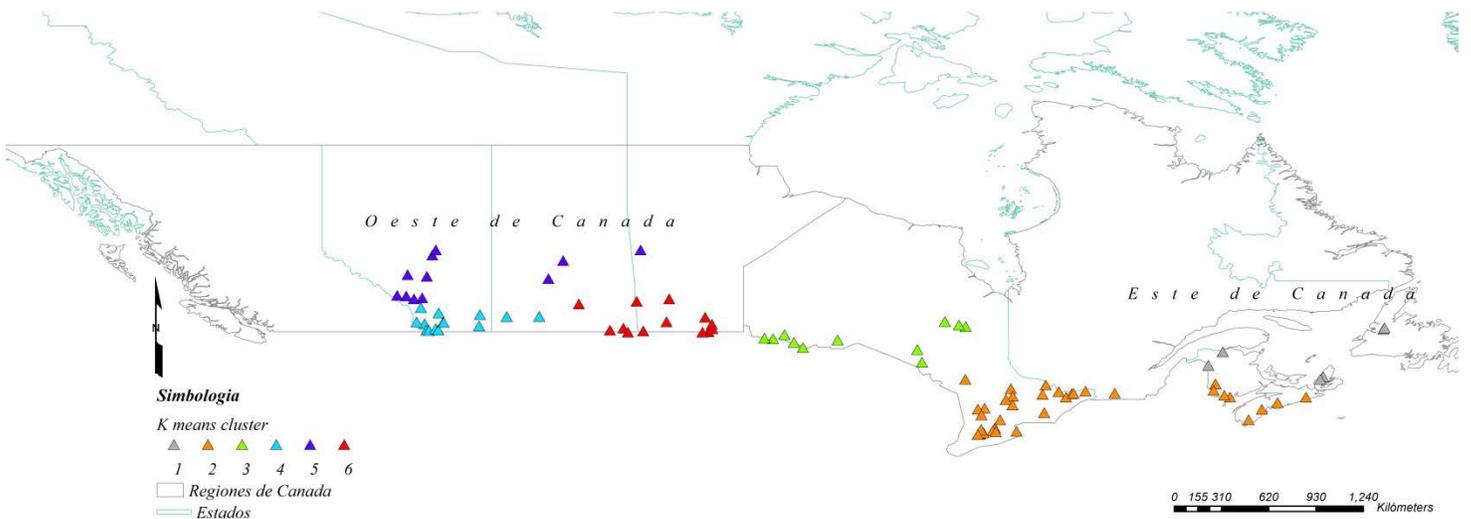


Figura 6.14 Localización de los sitios que conforman cada grupo delimitado mediante c-Medias

6.13 Selección de datos usando la medida de discordancia D_i

La medida de discordancia junto con las relaciones muestrales de Momentos-L de los 9 sitios en la región F se muestran en la Tabla 6.9 y se representan en la figura 6.15. El valor crítico, 2.491, se supera en un solo sitio, río Humber en la gran salida del lago (Humber river at grand lake outlet), con una medida de discordancia de 2.9995. Se puede observar en la figura. 6.15 que Humber river at grand lake outlet tiene la Asimetría-L más baja en comparación con los otros sitios. Por lo tanto, este sitio se excluye de la región homogénea.

Tabla 6.9. Radio de los Momentos-L y estadístico de discordancia D_i para escurrimientos en la región 9.

Sitio	Área de drenaje (km ²)	n_i	$t^{(i)}$	$t_3^{(i)}$	$t_4^{(i)}$	D_i
MOIRA RIVER NEAR FOXBORO	2620	94	0.123	-0.013	0.088	2.3388
PETAWAWA RIVER NEAR PETAWAWA	4120	95	0.125	0.059	0.144	1.3112
BONNECHERE RIVER NEAR CASTLEFORD	2380	89	0.183	0.046	0.160	0.6722
MADAWASKA RIVER AT PALMER RAPIDS	5800	80	0.125	0.047	0.079	0.1882
OTTAWA RIVER AT BRITANNIA	90900	95	0.127	-0.099	0.115	0.6781
MISSISSIPPI RIVER AT APPLETON	2900	92	0.143	-0.037	0.111	0.1573
RIDEAU RIVER AT OTTAWA	3830	76	0.155	-0.053	0.137	0.4832
SOUTH NATION RIVER NEAR	3810	94	0.172	0.045	0.111	0.7595
RICHELIEU (RIVIERE) AUX RAPIDES	22000	73	0.130	0.081	0.094	0.4119
HUMBER RIVER AT GRAND LAKE OUTLET	5020	85	0.338	0.180	0.117	2.9995

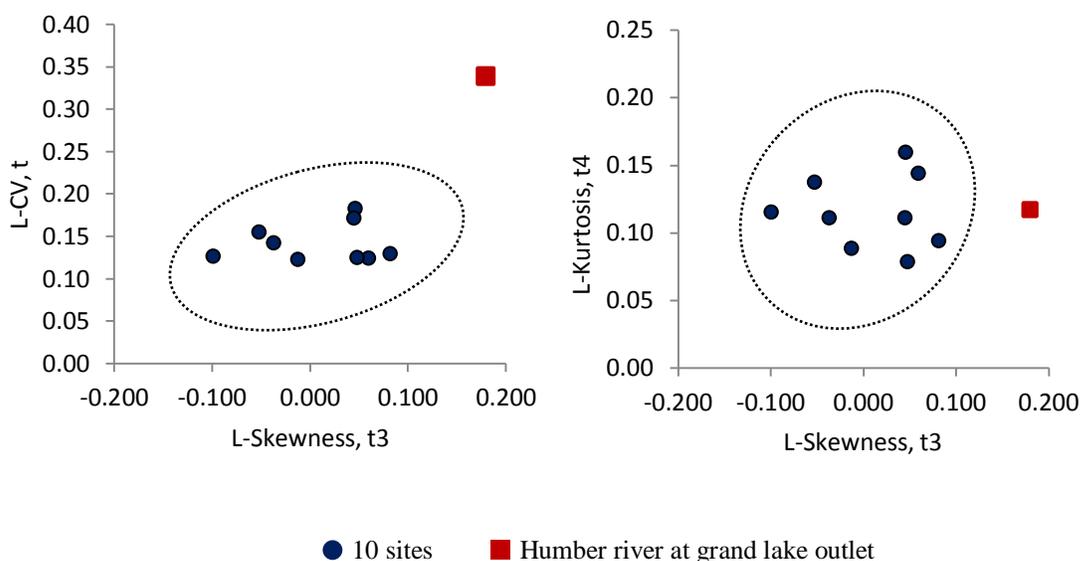


Figura 6.15 L-Radios de los Momentos-L de los 10 sitios en la región 9.

6.14 Heterogeneidad regional y formación de regiones homogéneas

El siguiente paso de la identificación de regiones homogéneas suele ser el más difícil y requiere una mayor cantidad de criterio. De acuerdo con Hosking y Wallis (1997) el valor de H para suponer el grado de heterogeneidad. Si la región no homogéneamente aceptable, se debe considerar redefinirla. La región se puede dividir en dos o más subregiones, algunos sitios pueden ser retirados de la región, o una asignación completamente diferente de los sitios a las regiones podría ser optada. Sin embargo, puede ser que la aparición de la heterogeneidad sea debida a la presencia de un pequeño número de sitios atípicos en la región. En tales casos, las características del lugar de estos sitios deben ser examinadas cuidadosamente. Es posible volver a asignar estos sitios para regiones de las que son más típicas, pero a veces puede parecer que no hay ninguna razón física para que los sitios sean atípicos siendo diferentes del resto de sitios en la región. Para el caso de la región F, los promedios de los pesos regionales

de los radios de los Momentos-L que fueron calculados son $\bar{t} = 0.1568$, $\bar{t}_3 = -0.01428$, $\bar{t}_4 = 0.0824$ con $V = 0.0275$. Los valores de los paametros correspondientes d ela distribución Kappa (Fig. 6) son $\xi = 0.1202$, $\alpha = 0.5952$, $k = 0.0094$ y $h = 0.1164$. Las estadísticas de resumen son $\mu V = 0.0212$ y $\sigma V = 0.0106$. Entonces se tiene a $H = 0.8089$, lo que indica que la región de estudio demuestra una homogeneidad aceptable.

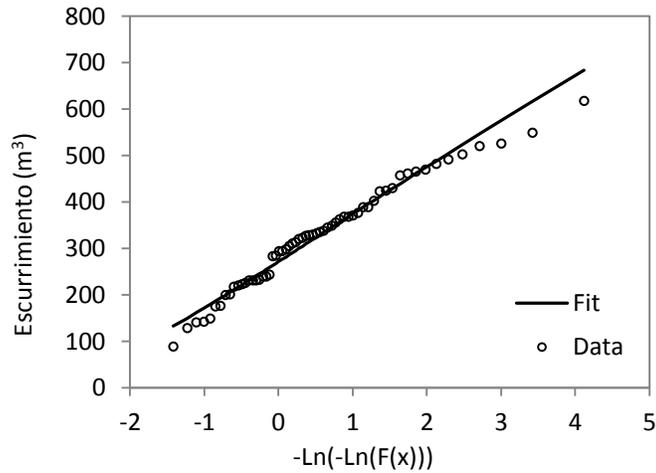


Figure 6.16 Ajuste de la distribución Kappa a ocho sitios de la región 9.

Las medidas de discordancia, heterogeneidad y pruebas de bondad de ajuste anteriormente descritas se aplicaron a los datos de cada región. El resumen estadístico de las regiones se encuentra en la Tabla 4. La medida de heterogeneidad H indica la obtención de regiones homogéneas las cuales se clasificaron en regiones de la A a la K . El valor más alto en el estadístico H los presenta la región H con un valor de 0.803 para dicho estadístico en general todos los grupos presentan valores menores a 1.00 por lo que se consideró a estas regiones como "homogéneas".

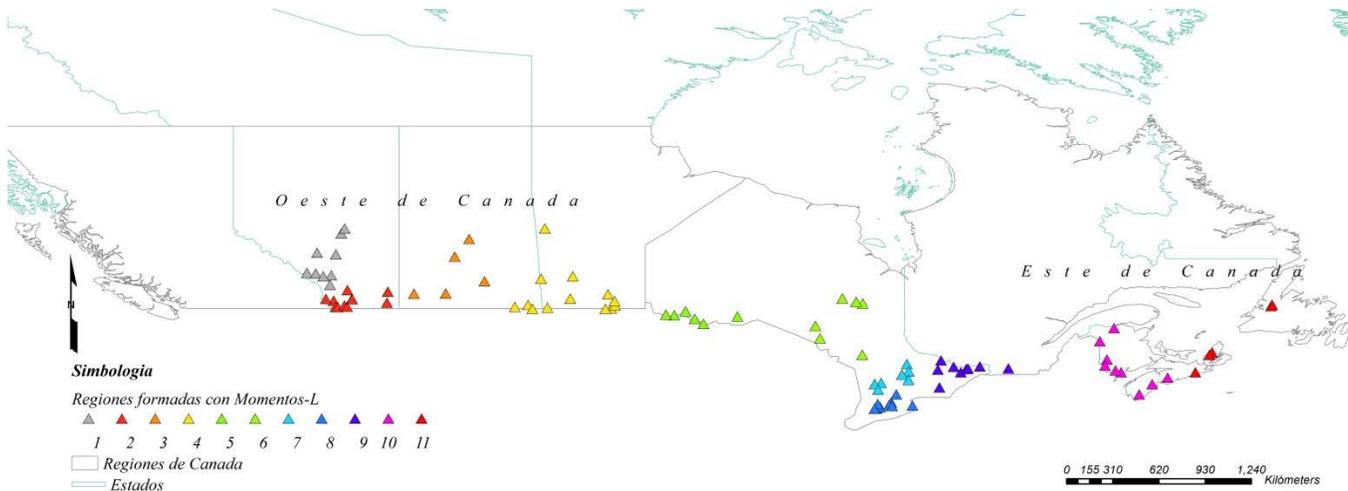


Figura 6.17 Localización de los sitios en cada grupo de estaciones en Canadá mediante c-Medias.

Tabla 6.10. Resumen de las regiones finales mediante Momentos-L.

Regiones	Número de sitios	\bar{t}	\bar{t}_3	\bar{t}_4	H	Regiones	Número de sitios	\bar{t}	\bar{t}_3	\bar{t}_4	H
1	9	0.235	0.252	0.188	0.803	7	7	0.125	0.026	0.157	-1.747
2	11	0.306	0.221	0.195	0.346	8	10	0.155	0.023	0.145	0.404
3	5	0.308	0.202	0.153	-3.105	9	9	0.143	0.009	0.116	-0.77
4	13	0.447	0.29	0.137	-1.29	10	9	0.129	0.047	0.087	0.108
5	6	0.168	0.062	0.142	0.171	11	5	0.084	0.031	0.136	-1.809
6	6	0.122	0.012	0.115	-0.314						

6.15 Comparación de los modelos RNAs y RNAs-Regional

Los modelos regionales c-Medias-RNAs y ML-RNAs así como RNAs tradicional fueron utilizados para estimar escurrimientos mensuales en las regiones suroeste y sureste de Canadá. Se eligieron períodos de entrenamiento y validación para la aplicación. El 90% de los datos de toda la serie de escurrimientos utilizada para el entrenamiento y el 10% restante de los datos de toda la serie para validación.

Se ha observado en los principales resultados que los modelos regionales basados en redes neuronales siempre fueron mejor que el modelo tradicional RNAs estos resultados se pueden comprobar en la Tabla 6.11. El RMSE obtenido fue menor en los modelos regionales que en el modelo tradicional RNAs ya que los modelos regionales c-Medias-RNAs y ML-RNAs presentaron mejoras en un 29% y 53% respectivamente de todas las estaciones analizadas mientras que el modelo tradicional RNAs solo en un 19%. En el MAE los mejores resultados se consiguieron en el modelo ML-RNAs ya que fue mejor en un 47% del total de las estaciones mientras que el modelo c-Medias-RNAs y RNAs consiguieron mejoría en el 28% y 26% de las estaciones estudiadas. Por último en el índice de concordancia d favoreció al modelo ML-RNAs en un 49%, en un 32% para c-Medias-RNAs y en 19% para el modelo RNAs. Los modelos regionales basados en RNAs reducen el RMSE, MAE y el índice de concordancia d . También es claro que la regionalización tiene un efecto positivo sobre el rendimiento del modelo y más aún el modelo ML-RNAs demostró ser superior a los demás modelos ya que en la mayoría de las estaciones demostró ser más preciso en el pronóstico mensual de escurrimiento.

Tabla 6.11. RMSE, MAE e índice de concordancia d de los modelos RNAs y Regionales-RNAs para las regiones estudiadas (periodo de prueba).

Grupo	Estación	RMSE			MAE			d		
		RNAs	ML-RNAs	c-RNAs	RNAs	ML-RNAs	c-RNAs	RNAs	ML-RNAs	c-RNAs
1	05BC001	358	356	298	228	248	235	0.9401	0.9201	0.9488
	05BE004	739	630	635	472	414	464	0.8437	0.8845	0.8857
	05BJ001	901	265	287	568	112	133	0.7996	0.4986	0.4413
	05BJ004	281	186	217	125	95	118	0.4745	0.644	0.5574
	05BL007	217	52	54	118	20	21	0.557	0.2456	0.214
	05CC002	57	1268	1340	22	678	695	0.1842	0.5408	0.4905
	05DC001	59	1576	1561	30	1010	1065	0.0184	0.629	0.6475
	05DF001	1249	2716	2976	735	1747	1789	0.553	0.6774	0.6148
	05EA001	2479	82	78	1728	60	57	0.6948	0.4454	0.459
2	05AA004	53	43	46	28	20	23	0.4168	0.6085	0.5334

	05AA008	106	85	98	66	50	60	0.728	0.792	0.743
	05AC003	52	49	48	22	24	22	0.408	0.2874	0.3052
	05AD003	440	290	302	253	179	179	0.7352	0.8791	0.8655
	05AD005	183	136	137	114	88	83	0.7637	0.8603	0.8541
	05AE002	2662	68	71	1190	28	30	0.4076	0.5839	0.5181
	05AE005	83	12	13	41	6	6	0.5191	0.2865	0.2259
	05AE006	13	342	344	6	195	197	0.1577	0.5337	0.5424
	05AE016	428	20	19	195	15	16	0.3495	0.177	0.1229
	05AF010	18	13	12	14	10	8	0.0253	0.2181	0.3008
	05AH005	137	22	23	101	17	18	0.7072	-0.0731	-0.0062
3	05GG001	74	3084	3163	54	2131	2075	0.396	0.6904	0.6952
	05HA003	3209	10	10	2373	7	6	0.6498	0.336	0.3729
	05HG001	12	2935	3140	8	2007	2215	0.2556	0.5187	0.6129
	05JB001	3289	91	100	2352	37	51	0.5135	0.1276	-0.0552
	05JF001	57	244	241	33	128	141	0.1907	0.5008	0.4058
4	05KJ001	208	6784	5462	111	5339	3796	0.2768	0.7216	0.797
	05LJ007	6437	127	128	4838	65	64	0.7507	0.1977	0.2396
	05ME001	138	412	350	68	304	249	0.2236	0.3633	0.5378
	05MJ001	334	1296	1317	214	794	750	0.6265	0.5882	0.5962
	05NB001	1199	32	29	677	27	22	0.6426	0.2855	0.3144
	05ND004	37	43	41	24	33	27	0.2542	0.1334	0.1778
	05ND007	167	240	240	84	145	116	0.4406	0.165	0.181
	05NF002	45	74	74	23	36	29	0.1503	-0.0961	-0.0977
	05NG001	614	1015	1032	306	466	452	0.6248	0.285	0.3032
	05OC001	8509	8792	8971	4215	4617	4553	0.3996	0.2811	0.2949
	05OC004	988	656	661	368	294	275	0.3029	0.2962	0.3294
05OD001	682	633	648	374	388	366	0.5355	0.5618	0.5552	
05OE001	306	275	282	149	144	139	0.2384	0.2585	0.2417	
5	05PA006	1724	1324	1437	1044	855	970	0.727	0.8492	0.8244
	05PA012	730	711	767	459	478	532	0.6714	0.7243	0.6841
	05PB014	962	874	856	626	553	518	0.5403	0.6221	0.6251
	05PC018	6350	5683	5626	4499	3833	3874	0.6181	0.686	0.6909
	05PC019	5206	4283	4407	3577	2956	3076	0.6406	0.7401	0.7271
	02AB006	988	841	814	645	568	545	0.3823	0.4289	0.456
6	02BD002	816	783	1041	610	554	742	0.4875	0.5506	0.4148
	02BE002	728	506	699	528	395	503	0.3465	0.4694	0.2728
	02CE002	406	354	480	255	239	318	0.5364	0.6484	0.5515
	04LD001	3976	2682	3140	2868	1809	2258	0.4174	0.6715	0.6657
	04LF001	2228	1592	1778	1596	1086	1292	0.4625	0.6725	0.6776
	04LJ001	3038	2805	2614	2169	1840	1951	0.4493	0.6044	0.6245
7	02EA005	156	121	125	108	81	98	0.3368	0.6784	0.6242
	02EB004	551	420	423	384	283	329	0.4469	0.701	0.6667
	02EB006	1314	1165	1144	1019	859	931	0.5796	0.6777	0.6629
	02EC002	504	440	438	314	294	339	0.6279	0.7253	0.7155
	02FB007	74	59	63	49	42	46	0.4334	0.6813	0.613
	02FC001	1380	1170	1222	933	816	942	0.4538	0.6691	0.6178
	02FC002	724	586	616	466	408	479	0.4469	0.6973	0.6395
8	02GA003	813	788	801	573	592	585	0.4173	0.5338	0.559
	02GA010	289	258	259	197	193	189	0.3106	0.4903	0.5261
	02GB001	1175	1094	1095	854	828	805	0.3978	0.5125	0.546
	02GD001	379	357	353	270	271	259	0.411	0.5137	0.549
	02GD003	522	484	485	354	364	353	0.3252	0.4833	0.5075
	02GD004	105	97	98	73	77	74	0.3948	0.4686	0.513
	02GD005	428	382	382	294	277	274	0.2711	0.4579	0.479
	02GE002	1039	948	936	738	704	679	0.3309	0.4977	0.5367
	02HA003	9798	9355	9059	7633	7365	7195	0.6136	0.656	0.6895
	02HB001	24	25	25	17	18	19	0.5203	0.5912	0.6541

9	02HL001	826	637	726	554	454	538	0.5387	0.7234	0.6484
	02KB001	1061	883	883	633	581	577	0.6048	0.7312	0.7197
	02KC009	401	375	367	257	255	242	0.6929	0.7325	0.7226
	02KD004	1237	1116	1090	802	727	778	0.6085	0.6881	0.7048
	02KF005	17337	13193	13241	11318	9469	9554	0.5687	0.6877	0.6889
	02KF006	705	575	569	484	380	398	0.6059	0.7279	0.7232
	02LA004	1012	928	931	667	656	665	0.5062	0.6642	0.6229
	02LB005	1687	1523	1608	1078	1101	1212	0.3684	0.5865	0.5285
	02OJ007	3687	3334	3552	2759	2510	2847	0.7772	0.8407	0.8122
10	01AF002	11903	9420	8766	7617	6803	6047	0.4532	0.7296	0.7033
	01AK001	134	115	145	100	86	114	0.3684	0.6921	0.5841
	01AQ001	160	144	191	124	112	151	0.3965	0.6039	0.4617
	01AQ002	778	674	844	590	531	643	0.4456	0.6651	0.5649
	01AR004	380	311	352	276	231	253	0.2798	0.5419	0.4811
	01BE001	1040	883	820	679	599	575	0.5338	0.7805	0.7528
	01DG003	69	69	87	53	51	64	0.3835	0.4764	0.3784
	01EC001	335	290	363	252	209	268	0.4229	0.5808	0.492
01EF001	815	758	933	585	549	691	0.4632	0.5732	0.4975	
11	01EO001	837	751	1080	647	557	804	0.4093	0.6024	0.4376
	01FB001	336	281	273	252	209	214	0.4557	0.6453	0.6551
	01FB003	148	129	127	111	96	93	0.706	0.8015	0.8036
	02YK001	407	596	627	308	416	415	0.5041	0.5099	0.3446
	02YL001	1592	1309	1206	1284	980	961	0.5312	0.7275	0.7888

Las figuras de la 6.18 a la 6.19 muestran el ajuste de los modelos RNAs y los modelos regionales ML-RNAs y c-Medias-RNAs en dos estaciones de la región *H* en el periodo de validación y de prueba. Los resultados obtenidos en las los diferentes periodos se muestran en la tabla 6. Los valores de las pruebas mediante RMSE, MAE y el índice de concordancia *d* de los modelos RNAs y regionales permitieron ver qué ML-RNAs fue mejor que RNAs y c-Medias-RNAs. En el periodo de prueba se observó que la estación Spray river at banff (05BC001) obtuvo mejores resultados en el RMSE con el modelo ML-RNA con un valor de 298 que con los modelos RNAs y c-Medias-RNAs con 356 y 358 respectivamente. El MAE fue menor con el modelo RNAs con un valor de 228 mientras que los modelos ML-RNAs con 235 y c-Medias-RNAs con 248 finalmente el índice de concordancia del modelo que presentó el mejor valor fue el ML-RNAs con 0.9448 mientras que los modelos RNAs y c-Medias-RNAs con 0.9201 y 0.9401. La estación Bow river near seebe (05BE004) en las pruebas de bondad de ajuste RMSE y MAE los mejores resultados se obtuvieron con el modelo regional propuesto ML-RNAs con valores de 630 y 414 mientras que los modelos RNAs y c-Medias-RNAs presentaron valores de 739, 635 y 472, 464 respectivamente. Finalmente en el índice de concordancia *d* los mejores resultados fueron obtenidos con c-Medias-RNAs con 0.8857 ya que los modelos ML-RNAs y RNAs obtuvieron 0.8845 y 0.8437 sin embargo cabe destacar que las diferencias son mínimas en esta última prueba de bondad de ajuste. Se halló que los resultados de las pruebas de bondad de ajuste en la validación y prueba (pronóstico) del modelo son variados y con diferencias pequeñas. Los resultados obtenidos en este estudio apoyan la premisa de que los modelos de redes neuronales regionales (ML-RNAs y c-Medias-RNAs) tienen un gran potencial y prometen ser utilizados como método alternativo para pronóstico de escurrimientos en ríos de países en desarrollo. Los resultados también muestran que la selección de las estaciones apropiadas en la delimitación de regiones homogéneas a través de Momentos- L en combinación con redes neuronales (ML-RNAs) proporciona resultados más precisos que el modelo regional c-Medias-RNAs y que los modelos de redes neuronales artificiales tradicionales (RNAs).

Tabla 6.12. RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs- Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).

Modelo	05BC001			05BE004		
	RMSE	MAE	d	RMSE	MAE	d
RNAs-Validación	386	263	0.9471	655	446	0.9065
RNAs-Prueba	358	228	0.9401	739	472	0.8437
ML-RNAs-Validación	376	249	0.9502	641	432	0.9102
ML-RNAs-Prueba	356	248	0.9201	630	414	0.8845
c-Medias-RNAs-Validación	367	247	0.9534	617	434	0.9191
c-Medias-RNAs-Prueba	298	235	0.9488	635	464	0.8857

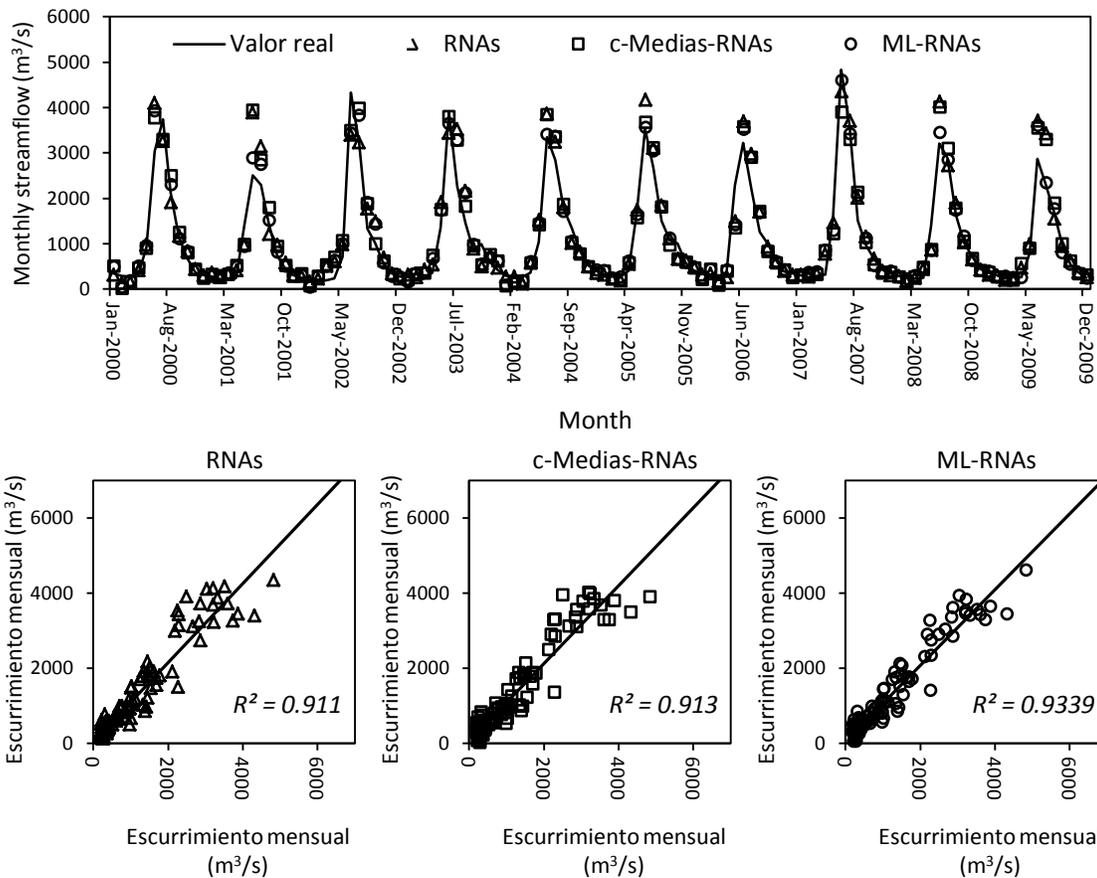


Figura 6.18. Estimación mensual de escurrimientos mediante el modelo RNAs y RNAs-Regional en el periodo de prueba—Estación Spray river at banff (2BC001).

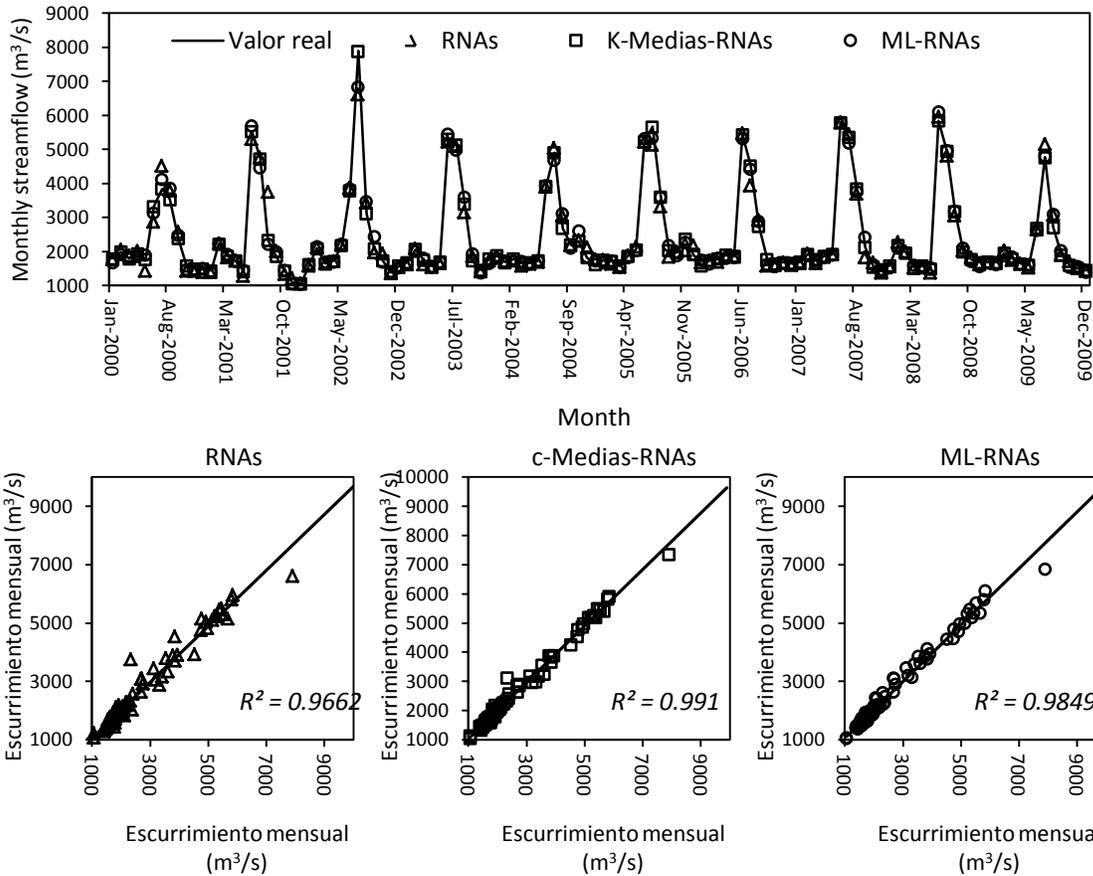


Figura 6.19. Estimación mensual de escurrimientos mediante el modelo RNAs y RNAs-Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).

6.16 Comparación de los modelos Ensamblado y Ensamblado-Regional

Los modelos ensamblados han demostrado buen desempeño en recientes años. En este análisis se han utilizado datos de escurrimientos. Los escurrimientos han sido modelados mediante el enfoque tradicional y el enfoque regional propuesto, donde se utilizaron dos tipos de modelos ensamblados. El modelo Ensamblado basado en RNAs es el enfoque tradicional actual y se proponen dos modelos regionales. Los cuales estan basados en c-Medias-Ensamblado y ML-Ensamblado. Estos tres modelos fueron evaluados en el pronóstico de escurrimientos mensuales para las regiones suroeste y sureste de Canadá. El periodo de entrenamiento correspondió al 90% de la serie utilizada el siguiente 10% corresponde a la validación y aplicación. Los resultados indican que los modelos regionales siempre fueron superiores al modelo tradicional. Esto se puede comprobar en la Tabla 6.13. El RMSE obtenido por el modelo Ensamblado fue mejor en solo un 32% de las estaciones analizadas mientras que en el restante 68% de las estaciones los mejores resultados fueron obtenidos por los modelos regionales c-Medias-Ensamblado y ML-Ensamblado con un 17% y 51% respectivamente. En el MAE el mejor ajuste lo obtuvo el modelo ML-Ensamblado con el 48% de las estaciones, después el modelo Ensamblado con el 29% finalmente el modelo c-Medias-Ensamblado con 23%, para esta prueba de bondad de ajuste es notable que el modelo ML-Ensamblado ha

conseguido el mejor ajuste y en el índice de concordancia, el modelo ML-Ensamblado logró el mejor ajuste en el 40% de las estaciones analizadas, el modelo Ensamblado en un 39% y c-Medias-Ensamblado en 21%. El mejor ajuste en las tres pruebas de bondad de ajuste se obtuvo con el modelo ML-Ensamblado, en segundo lugar el modelo Ensamblado y finalmente el modelo Ensamblado regional con c-Medias. Se concluye que una adecuada regionalización brinda un adecuado desempeño en el pronóstico de escurrimientos ya que el modelo ML-Ensamblado obtuvo el mejor ajuste.

Tabla 6.13. RMSE, MAE e índice de concordancia *d* de los modelos Ensamblado y Ensamblado Regional para las regiones estudiadas (periodo de prueba).

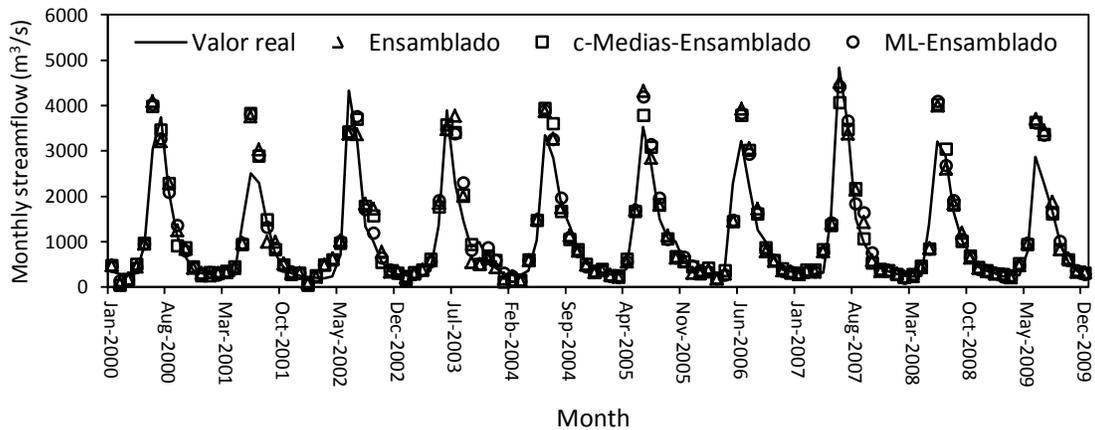
Grupo	Estación	RMSE			MAE			<i>d</i>		
		Ens.	ML-Ens	c-Ens.	Ens.	ML-Ens	c-Ens.	Ens.	ML-Ens	c-Ens.
1	05BC001	364	353	301	237	250	239	0.9921	0.9801	0.9188
	05BE004	746	624	626	478	414	454	0.7937	0.8545	0.8457
	05BJ001	909	271	294	577	113	131	0.8996	0.5186	0.4713
	05BJ004	279	179	207	129	92	120	0.3745	0.574	0.6474
	05BL007	207	42	59	122	11	24	0.637	0.2756	0.254
	05CC002	59	1276	1350	14	684	700	0.2242	0.5008	0.5805
	05DC001	63	1571	1561	27	1010	1061	0.0784	0.719	0.6875
	05DF001	1247	2721	2981	733	1743	1782	0.553	0.6374	0.7148
	05EA001	2469	80	79	1721	50	64	0.6448	0.4754	0.419
2	05AA004	45	42	41	31	22	19	0.4768	0.6585	0.5934
	05AA008	112	75	90	72	60	50	0.768	0.862	0.733
	05AC003	60	56	58	29	14	17	0.328	0.3074	0.3252
	05AD003	435	294	309	250	170	172	0.7352	0.8291	0.7855
	05AD005	187	146	144	117	95	87	0.7337	0.9303	0.7641
	05AE002	2669	69	63	1181	22	40	0.4976	0.6839	0.5281
	05AE005	89	6	5	48	-3	16	0.5791	0.1865	0.2459
	05AE006	17	336	353	-4	205	205	0.2277	0.6037	0.6124
	05AE016	430	10	9	204	11	24	0.3695	0.107	0.1329
05AF010	21	15	18	22	20	4	0.0653	0.1281	0.2008	
05AH005	136	19	15	109	12	21	0.7172	0.1731	0.0062	
3	05GG001	82	3076	3157	55	2136	2073	0.496	0.6704	0.7852
	05HA003	3205	7	6	2380	9	1	0.6398	0.306	0.3929
	05HG001	21	2935	3150	8	2005	2219	0.3356	0.5787	0.6329
	05JB001	3280	91	100	2342	32	55	0.5935	0.0276	0.0248
	05JF001	51	241	232	31	122	147	0.2507	0.6008	0.4758
4	05KJ001	203	6775	5452	118	5332	3805	0.2568	0.6416	0.737
	05LJ007	6440	127	118	4848	64	72	0.8007	0.2577	0.3296
	05ME001	138	416	353	66	295	253	0.1636	0.3933	0.5778
	05MJ001	329	1288	1325	217	804	751	0.5465	0.5882	0.5062
	05NB001	1191	29	37	685	29	13	0.6026	0.2955	0.3544
	05ND004	32	48	31	17	29	31	0.2442	0.0334	0.0778
	05ND007	174	240	236	84	141	111	0.4606	0.235	0.261
	05NF002	49	77	70	13	44	30	0.2403	0.1361	0.1777
	05NG001	610	1019	1024	308	458	444	0.6848	0.275	0.2932
	05OC001	8510	8794	8975	4222	4621	4554	0.3896	0.3211	0.2049
	05OC004	989	658	664	377	291	280	0.2929	0.2962	0.4194
	05OD001	675	628	639	382	383	369	0.4355	0.5218	0.4852
05OE001	306	269	276	153	146	135	0.2784	0.1885	0.3217	
5	05PA006	1728	1316	1446	1041	849	963	0.707	0.7792	0.7744
	05PA012	733	702	757	468	482	527	0.7014	0.7243	0.5841
	05PB014	968	878	851	635	562	514	0.5203	0.5521	0.6851
	05PC018	6352	5688	5620	4496	3833	3872	0.6781	0.776	0.6109

	05PC019	5212	4283	4412	3587	2955	3072	0.6706	0.7501	0.6671
	02AB006	991	841	805	655	561	553	0.3123	0.3889	0.436
6	02BD002	811	789	1047	619	544	738	0.4675	0.5006	0.3548
	02BE002	732	497	708	523	386	496	0.4165	0.5194	0.3128
	02CE002	401	346	473	259	238	312	0.4864	0.6284	0.6315
	04LD001	3979	2687	3133	2861	1816	2265	0.4574	0.5815	0.6457
	04LF001	2228	1583	1769	1593	1085	1282	0.4425	0.7725	0.7576
	04LJ001	3034	2799	2618	2171	1843	1947	0.4893	0.6444	0.6145
	7	02EA005	151	113	132	114	73	101	0.4268	0.6784
02EB004		544	414	419	381	282	324	0.4769	0.671	0.7067
02EB006		1318	1174	1142	1017	858	938	0.5196	0.5777	0.7529
02EC002		495	436	438	317	304	333	0.6679	0.6953	0.7955
02FB007		69	51	56	46	33	50	0.3434	0.6713	0.573
02FC001		1387	1168	1215	931	820	942	0.4038	0.6891	0.6278
02FC002		729	594	612	457	398	486	0.3769	0.5973	0.6895
8	02GA003	819	797	794	575	589	593	0.4173	0.4338	0.589
	02GA010	286	258	261	205	200	196	0.3706	0.5803	0.5461
	02GB001	1174	1097	1085	857	823	807	0.4378	0.4925	0.546
	02GD001	378	350	349	266	276	269	0.381	0.6137	0.619
	02GD003	530	485	478	354	354	356	0.3252	0.4333	0.5975
	02GD004	107	94	102	66	71	80	0.4448	0.5086	0.533
	02GD005	418	373	389	293	272	273	0.3711	0.3979	0.399
	02GE002	1045	958	928	740	707	683	0.3709	0.5277	0.5667
	02HA003	9793	9361	9054	7636	7360	7199	0.6236	0.696	0.6595
	02HB001	26	16	33	17	24	29	0.4203	0.5912	0.5741
9	02HL001	828	644	733	549	461	537	0.6287	0.7634	0.7084
	02KB001	1071	876	893	643	577	571	0.5648	0.6312	0.7497
	02KC009	408	369	375	260	261	240	0.7129	0.7325	0.6526
	02KD004	1232	1119	1090	807	734	771	0.6085	0.6581	0.7948
	02KF005	17332	13193	13247	11318	9477	9558	0.5987	0.6377	0.6389
	02KF006	714	578	563	490	382	405	0.5759	0.6679	0.6932
	02LA004	1014	926	939	657	664	662	0.4162	0.6142	0.6629
	02LB005	1687	1516	1616	1074	1109	1220	0.3484	0.5965	0.6285
	02OJ007	3683	3342	3549	2759	2502	2837	0.7272	0.8807	0.7322
10	01AF002	11906	9421	8763	7622	6812	6046	0.5532	0.8196	0.6333
	01AK001	127	118	145	109	94	122	0.4284	0.6021	0.6241
	01AQ001	157	143	198	134	114	150	0.3065	0.5939	0.3917
	01AQ002	772	668	844	581	532	650	0.5356	0.7451	0.5649
	01AR004	370	306	352	272	232	256	0.3698	0.6119	0.4011
	01BE001	1043	889	827	670	590	576	0.6138	0.7505	0.7128
	01DG003	65	72	87	57	46	71	0.3935	0.5564	0.3684
	01EC001	342	289	368	260	218	278	0.4529	0.5908	0.552
	01EF001	807	756	939	584	555	694	0.5232	0.5032	0.4875
11	01EO001	844	753	1070	649	563	800	0.4593	0.5524	0.3776
	01FB001	342	290	271	253	204	224	0.3857	0.5853	0.7151
	01FB003	149	129	136	109	98	88	0.796	0.8215	0.7036
	02YK001	406	606	634	311	425	418	0.4941	0.4799	0.3046
	02YL001	1589	1299	1212	1278	985	967	0.4512	0.6975	0.8888

En las figuras 6.20 y 6.21 se muestran los ajustes de los modelos Ensamblado y los modelos regionales ML-Ensamblado y c-Medias-Ensamblado en dos estaciones de la región 1 y en el periodo de validación y de prueba. Los resultados obtenidos en los diferentes periodos se muestran en la tabla 6.14. En el periodo de prueba se observó que la estación Spray river at banff (05BC001) el modelo c-Medias-Ensamblado obtuvo el mejor ajuste en los tres índices, esto es en el RMSE, MAE y d con valores de 301, 239 y 0.9188 respectivamente. Los modelos ML-Ensamblado y Ensamblado obtuvieron resultados similares en RMSE de 353 y 364 mientras que en el MAE y del mejor ajuste lo obtuvo el modelo ML-Ensamblado con 250 y 0.9801 mejores que los del modelo RNAs con 237 y 0.9921. La estación Bow river near seebe (05BE004) en las pruebas de bondad de ajuste RMSE, MAE y d los mejores resultados los presentó el modelo regional propuesto ML-RNAs con valores de 624, 414 y 0.8545 mientras que los modelos Ensamblado y c-Medias-Ensamblado presentaron valores de 746, 478, 0.7937 y 626, 454, 0.8457 respectivamente. Los modelos Ensamblados regionales presentan un adecuado desempeño para el pronóstico de escurrimientos, esto es mejorado mediante la combinación de Momentos-L y c-Medias y el modelo Ensamblado (ML-Ensamblado y c-Medias-Ensamblado), esto permite que dichos modelos sean herramientas alternativas para pronóstico de escurrimientos.

Tabla 6.14. RMSE, MAE e índice de concordancia d de los modelos Ensamblado y Ensamblado Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).

Modelo	05BC001			05BE004		
	RMSE	MAE	d	RMSE	MAE	d
Ensamblado-Validación	388	252	0.9418	700	453	0.8544
Ensamblado-Prueba	364	237	0.9921	746	478	0.7937
ML-Ensamblado-Validación	365	237	0.9480	705	468	0.8529
ML-Ensamblado-Prueba	353	250	0.9801	624	414	0.8545
c-Medias-Ensamblado-Validación	377	249	0.9443	712	474	0.8518
c-Medias-Ensamblado-Prueba	301	239	0.9188	626	454	0.8457



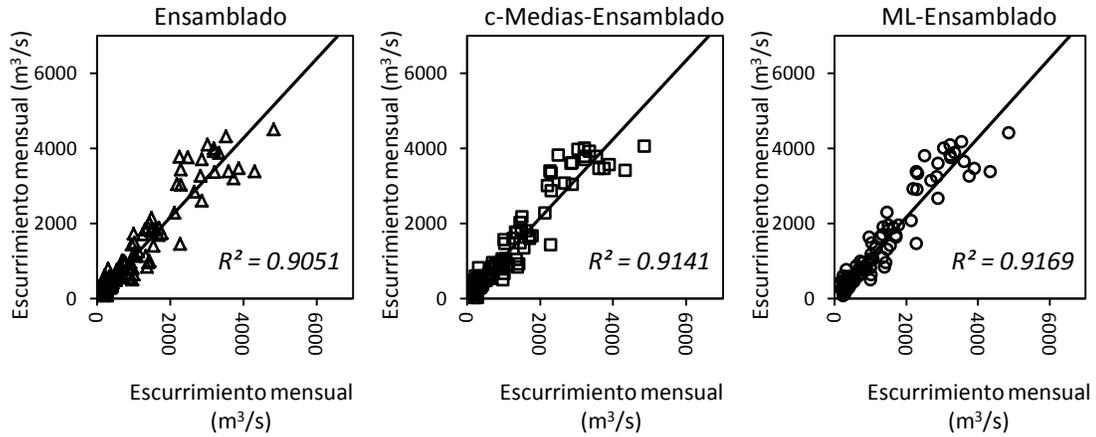


Figura 6.20. Estimación mensual de escurrimientos mediante el modelo Ensamblado y Ensamblado Regional en el periodo de prueba—Estación Spray river at banff (2BC001).

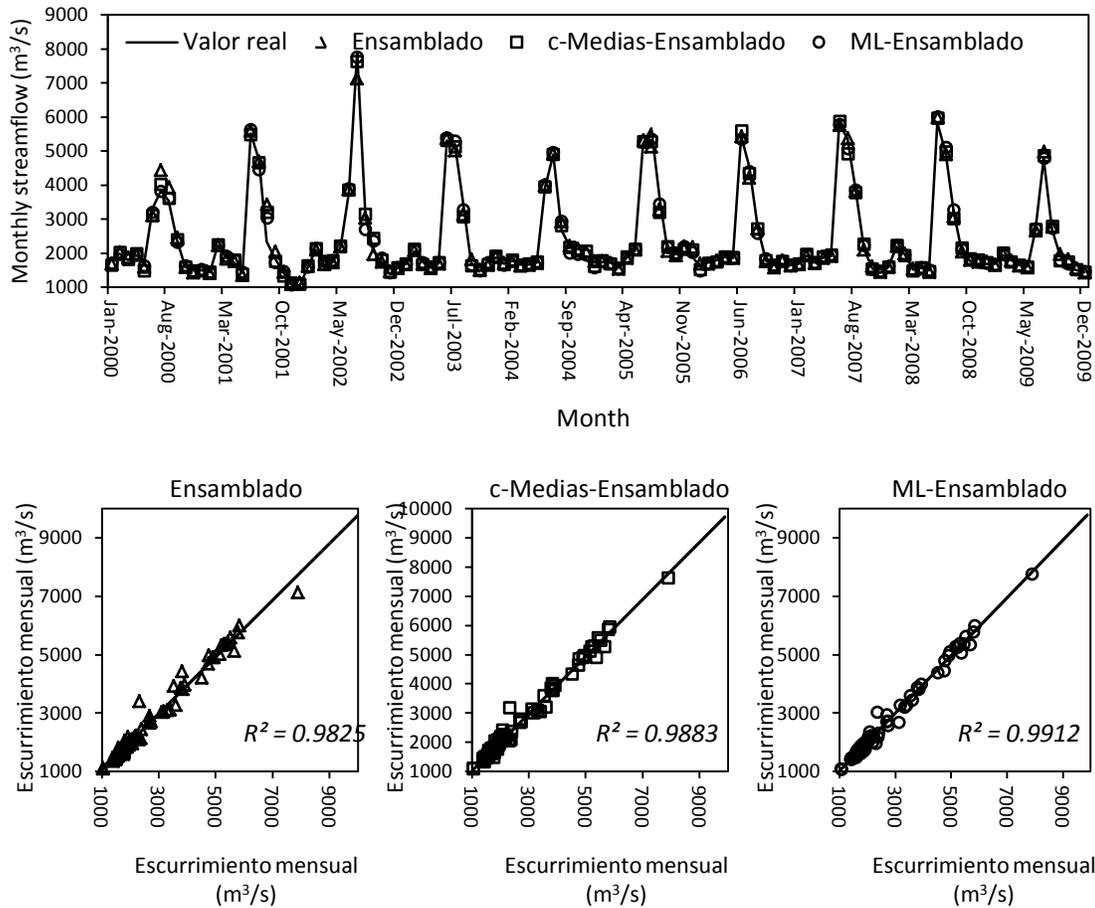


Figura 6.21. Estimación mensual de escurrimientos mediante el modelo Ensamblado y Ensamblado Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).

6.17 Comparación de los modelos Modular y Modular Regional

Los modelos modulares son altamente probabilísticos de allí su gran éxito en las diferentes áreas. Los modelos regionales c-Medias-Modular y ML-Modular así como Modular tradicional han sido utilizados para estimar escurrimientos mensuales en las regiones suroeste y sureste de Canadá. Se ha elegido igual que en las anteriores pruebas períodos de entrenamiento y validación para la aplicación. El 90% de los datos de toda la serie de escurrimientos utilizada para el entrenamiento y el 10% restante de los datos de toda la serie para validación.

Se observó en los resultados que los modelos regionales basados en el enfoque Modular siempre fueron mejor que el modelo tradicional Modular, estos resultados se pueden comprobar en la Tabla 6.15. El RMSE obtenido fue menor en el modelo Modular Regional ML-Modular, después le sigue el modelo Modular y por último el modelo c-Medias-Modular en orden de mejor ajuste respectivamente obtuvieron los ajustes en los siguientes porcentajes con un 53%, 28% y 19% respectivamente de todas las estaciones analizadas. En el MAE también el mejor ajuste lo obtuvo el modelo ML-Modular, después los modelos Modular y c-Medias con mínimas diferencias entre ellos. Los porcentajes fueron 46%, 28% y 27% respectivamente de las estaciones estudiadas. Por último en el índice de concordancia *d* favoreció al modelo ML-Modular en un 40%, en un 39% para el modelo Modular y 21% para para el modelo c-Medias-Modular. Los resultados demuestran que el mejor desempeño se consiguió con el modelo ML-Modular. Los modelos regionales pueden obtener un adecuado desempeño siempre y cuando se tenga una adecuada delimitación de las regiones homogéneas puesto que el modelo ML-Modular demostró ser superior a los demás modelos y de ahí que es una adecuada herramienta en el pronóstico mensual de escurrimiento.

Tabla 6.15. RMSE, MAE e índice de concordancia *d* de los modelos Modular y Modular Regional para las regiones estudiadas (periodo de prueba).

Grupo	Estación	RMSE			MAE			<i>d</i>		
		Modular	ML-M	c-M	Modular	ML-M	c-M	Modular	ML-M	c-M
1	05BC001	352	353	301	228	244	232	0.8701	0.9501	0.9088
	05BE004	732	623	637	475	404	466	0.8837	0.9445	0.8057
	05BJ001	896	272	296	576	108	141	0.7596	0.5086	0.4913
	05BJ004	279	189	212	126	93	123	0.4045	0.584	0.5074
	05BL007	225	45	63	121	25	28	0.637	0.2756	0.274
	05CC002	59	1274	1350	30	688	702	0.2542	0.4608	0.5705
	05DC001	55	1566	1555	27	1020	1066	0.0684	0.669	0.5475
	05DF001	1258	2726	2966	739	1754	1793	0.573	0.7674	0.5648
	05EA001	2475	89	83	1725	65	67	0.7948	0.5154	0.399
2	05AA004	48	40	43	29	20	18	0.4768	0.5985	0.5434
	05AA008	111	86	96	63	47	60	0.688	0.842	0.703
	05AC003	59	44	54	32	14	14	0.468	0.2174	0.3752
	05AD003	434	281	294	247	177	174	0.6552	0.9691	0.7955
	05AD005	189	134	139	104	87	83	0.8237	0.7703	0.7641
	05AE002	2667	68	68	1200	32	26	0.3076	0.5939	0.5881
	05AE005	88	4	20	39	4	-3	0.5591	0.3765	0.3059
	05AE006	11	332	335	13	199	207	0.2377	0.4437	0.5124
	05AE016	437	23	11	199	6	17	0.2695	0.097	0.0229

	05AF010	8	10	18	13	2	9	0.0053	0.2081	0.3308	
	05AH005	137	16	13	98	26	20	0.7772	0.1431	0.0062	
3	05GG001	77	3075	3171	46	2138	2072	0.456	0.6404	0.6352	
	05HA003	3213	19	11	2364	2	13	0.6898	0.246	0.3329	
	05HG001	2	2926	3146	16	2012	2205	0.2956	0.5587	0.6929	
	05JB001	3299	94	103	2356	46	55	0.4135	0.0376	0.0052	
	05JF001	55	245	249	39	119	144	0.2407	0.5708	0.4758	
		05KJ001	216	6786	5467	105	5337	3790	0.1768	0.7816	0.807
4	05LJ007	6447	120	130	4841	56	57	0.6907	0.0977	0.1796	
	05ME001	136	403	349	76	310	255	0.1636	0.3633	0.5678	
	05MJ001	339	1303	1324	215	784	758	0.5365	0.4882	0.6362	
	05NB001	1197	31	25	671	36	32	0.6026	0.2855	0.2144	
	05ND004	41	52	32	21	35	22	0.2742	0.1834	0.2578	
	05ND007	157	236	240	83	139	122	0.4706	0.265	0.211	
	05NF002	53	83	74	27	44	32	0.0503	0.1361	0.1277	
	05NG001	623	1010	1026	305	469	453	0.7048	0.255	0.3632	
	05OC001	8515	8792	8968	4219	4621	4556	0.3496	0.3411	0.2149	
	05OC004	989	663	663	361	295	277	0.3929	0.2862	0.2294	
	05OD001	691	623	644	364	390	363	0.4355	0.6418	0.5152	
	05OE001	312	267	290	148	142	131	0.3284	0.2585	0.1817	
	5	05PA006	1718	1329	1430	1044	857	966	0.757	0.9092	0.8344
		05PA012	737	716	767	464	470	535	0.5814	0.6743	0.7641
05PB014		954	884	858	633	554	511	0.5803	0.6621	0.7251	
05PC018		6348	5690	5625	4502	3840	3872	0.5181	0.706	0.7609	
05PC019		5204	4287	4402	3577	2951	3078	0.6606	0.7501	0.7171	
02AB006		998	838	817	638	570	541	0.4423	0.5289	0.386	
6	02BD002	813	788	1046	617	559	736	0.4075	0.4606	0.3948	
	02BE002	725	496	699	525	390	502	0.3965	0.3894	0.2128	
	02CE002	412	362	479	265	249	316	0.6064	0.5784	0.6215	
	04LD001	3978	2689	3132	2865	1814	2260	0.3674	0.6515	0.6057	
	04LF001	2224	1597	1773	1603	1095	1294	0.3825	0.7125	0.7676	
	04LJ001	3030	2806	2623	2178	1840	1945	0.3893	0.5744	0.6545	
7	02EA005	163	128	119	98	83	88	0.3268	0.6884	0.5942	
	02EB004	552	424	418	390	288	338	0.3769	0.801	0.5967	
	02EB006	1317	1167	1153	1020	863	936	0.4996	0.6677	0.7029	
	02EC002	504	440	433	321	294	339	0.6879	0.6853	0.6555	
	02FB007	69	67	53	41	42	54	0.3934	0.6913	0.703	
	02FC001	1373	1165	1231	942	808	941	0.3838	0.6291	0.5578	
	02FC002	734	588	615	476	404	471	0.5269	0.5973	0.6295	
8	02GA003	815	780	795	564	587	576	0.4373	0.4538	0.559	
	02GA010	279	260	252	200	189	189	0.4106	0.4603	0.5361	
	02GB001	1182	1101	1090	863	834	796	0.2978	0.4725	0.466	
	02GD001	372	352	343	268	264	262	0.411	0.5837	0.449	
	02GD003	519	485	483	345	358	359	0.3952	0.5633	0.4475	
	02GD004	113	89	93	64	67	71	0.4248	0.4386	0.583	
	02GD005	423	376	388	287	273	276	0.3211	0.5479	0.509	
	02GE002	1045	944	936	741	710	677	0.3409	0.4777	0.5267	
	02HA003	9791	9359	9055	7642	7371	7187	0.6036	0.596	0.5995	
	02HB001	27	35	25	23	22	29	0.5003	0.6212	0.6241	
9	02HL001	824	635	725	550	449	538	0.5487	0.6534	0.7184	
	02KB001	1066	882	876	641	576	572	0.5348	0.8212	0.7397	
	02KC009	405	384	360	264	260	250	0.7529	0.7025	0.7926	
	02KD004	1246	1110	1091	810	726	784	0.5585	0.5881	0.7848	
	02KF005	17330	13196	13232	11324	9479	9555	0.5487	0.7177	0.7789	
	02KF006	700	583	564	490	387	401	0.5059	0.8079	0.7632	
	02LA004	1011	922	925	662	657	674	0.5662	0.7542	0.6129	
	02LB005	1682	1520	1603	1079	1095	1217	0.3884	0.6765	0.6085	

	02OJ007	3697	3330	3553	2768	2510	2845	0.6872	0.7507	0.8022
10	01AF002	11907	9418	8773	7610	6810	6049	0.4632	0.6496	0.7533
	01AK001	135	114	142	104	96	114	0.4084	0.7321	0.5941
	01AQ001	159	144	201	126	106	147	0.2965	0.5039	0.4917
	01AQ002	781	665	842	593	530	634	0.3556	0.6751	0.4649
	01AR004	383	321	353	275	240	246	0.3398	0.4419	0.5311
	01BE001	1033	889	824	681	605	570	0.5638	0.7505	0.8528
	01DG003	78	68	87	52	47	68	0.4135	0.3764	0.4484
	01EC001	332	281	361	254	217	270	0.4229	0.5208	0.572
	01EF001	810	755	941	577	542	686	0.4032	0.5232	0.4075
11	01EO001	838	757	1074	644	559	809	0.3693	0.5424	0.5076
	01FB001	345	273	271	253	219	207	0.4357	0.6553	0.6751
	01FB003	147	136	124	111	90	89	0.756	0.7715	0.8736
	02YK001	397	595	617	303	425	412	0.5441	0.4899	0.2846
	02YL001	1588	1301	1213	1280	973	969	0.4812	0.6375	0.7088

Las figuras de la 6.22 a la 6.23 muestran el ajuste de los modelos Modulares y los modelos regionales ML-Modular y c-Medias-Modular en dos estaciones de la región 1 en el periodo de validación y de prueba. Los resultados obtenidos en las los diferentes periodos se muestran en la tabla 6.16. Los valores de las pruebas mediante RMSE, MAE y el índice de concordancia d de los modelos Modular y regionales c-Medias-Modular y ML-Modular demostraron que para esta estación el modelo ML-Modular fue mejor que los modelos Modular y c-Medias-Modular. En el periodo de prueba se observó que la estación Spray river at banff (05BC001) obtuvo mejores resultados en el RMSE con el modelo c-Medias-Modular con un valor de 301 que con los modelos Modular y ML-Modular con 352 y 353 respectivamente. El MAE fue menor con el modelo Modular con un valor de 228 mientras que los modelos ML-Modular con 244 y c-Medias-Modular con 232 finalmente el índice de concordancia del modelo que presentó el mejor valor fue el ML-Modular con 0.9501 mientras que los modelos Modular y c-Medias-Modular con 0.8701 y 0.9088. La estación Bow river near seebe (05BE004) en las pruebas de bondad de ajuste RMSE los mejores resultados los presentó el modelo regional propuesto ML-Modular con un valor de 623 mientras que los modelos Modular y c-Medias-Modular presentaron valores de 732 y 637 respectivamente. En el MAE también el mejor ajuste lo obtuvo el modelo ML-Modular con 404, Modular y c-Medias-MOular con 475 y 466. Finalmente en el índice de concordancia d los mejores resultados fueron obtenidos con ML-Modular con 0.9445 ya que los modelos c-Medias-Modular y Modular obtuvieron 0.8837 y 0.8057 sin embargo cabe destacar que las diferencias son mínimas en esta última prueba de bondad de ajuste. Se encontró que los resultados de las pruebas de bondad de ajuste en la validación y prueba (pronostico) del modelo son variados y con diferencias pequeñas.

Los resultados obtenidos en este estudio apoyan la premisa de que los modelos de redes neuronales regionales (ML-Modular y c-Medias-Modular) tienen un gran potencial y prometen ser utilizados como método alternativo para pronóstico de escurrimientos en ríos de países en desarrollo. Los resultados también muestran que la selección de las estaciones apropiadas en la delimitación de regiones homogéneas a través de Momentos- L en combinación con redes neuronales (ML-Modular) proporciona resultados más precisos que el modelo regional c-Medias-Modular y que los modelos Modulares tradicionales.

Tabla 6.16. RMSE, MAE e índice de concordancia d de los modelos Modular y Modular Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).

Modelo	05BC001			05BE004		
	RMSE	MAE	d	RMSE	MAE	d
Modular-Validación	399	254	0.9388	686	445	0.8632
Modular-Prueba	352	228	0.8701	732	475	0.8837
ML-Modular-Validación	368	253	0.9479	689	457	0.8589
ML-Modular-Prueba	353	244	0.9501	623	404	0.9445
c-Medias-Modular-Validación	358	230	0.9499	676	453	0.8638
c-Medias-Modular-Prueba	301	232	0.9088	637	466	0.8057

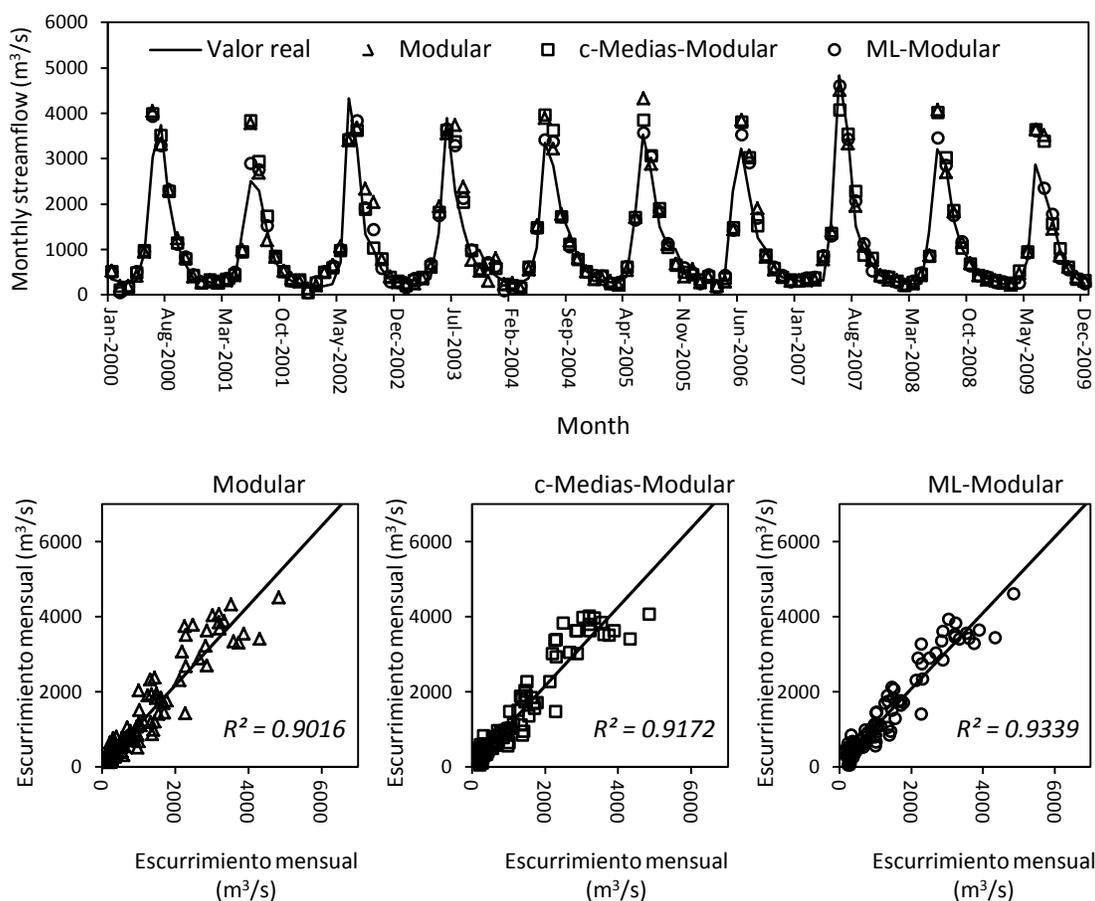


Figura 6.22. Estimación mensual de escurrimientos mediante el modelo Modular y Modular Regional en el periodo de prueba—Estación Spray river at banff (2BC001).

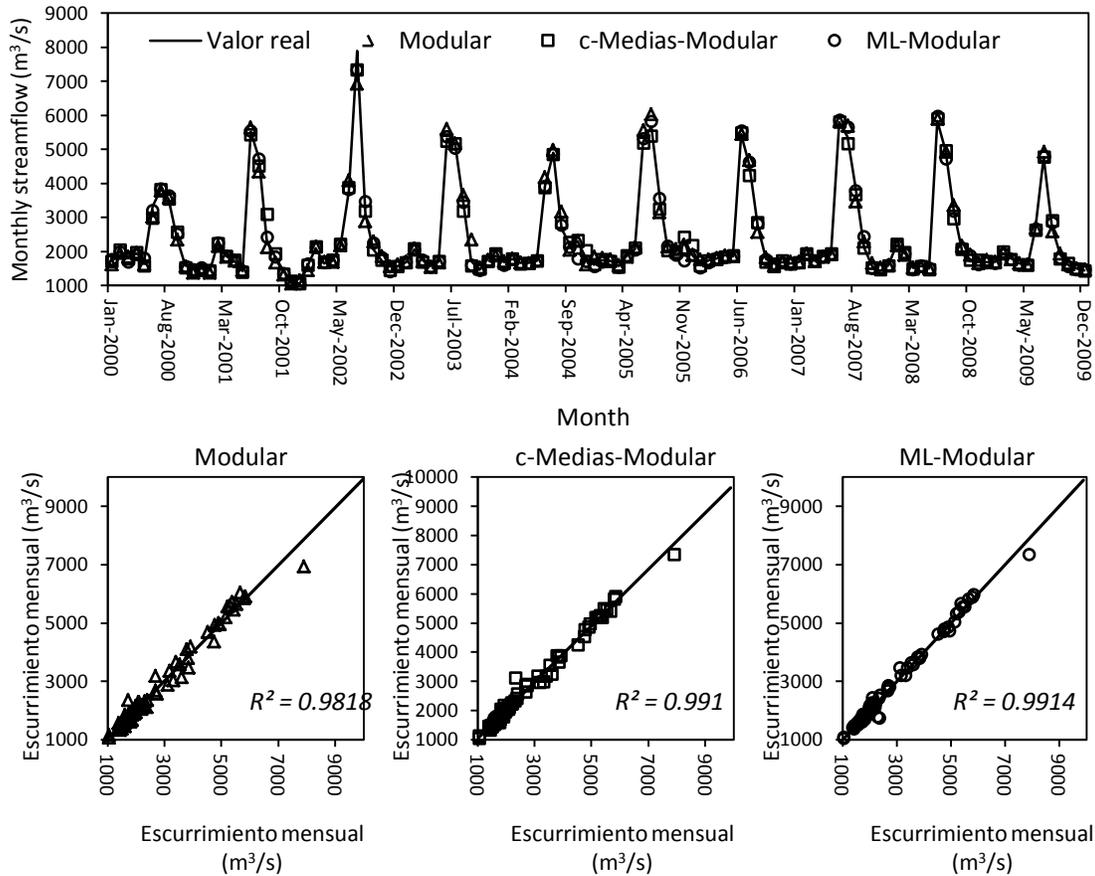


Figura 6.23. Estimación mensual de escurrimientos mediante el modelo Modular y Modular Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).

6.18 Comparación de escurrimientos en las escalas mensual trimestral y semestral.

En esta parte se realiza la comparación del modelo de predicción tradicional Ensamblado via RNAs y el modelo Ensamblado mediante el enfoque regional propuesto. La variable estudiada es el escurrimiento en las escalas de tiempo mensual, trimestral y semestral de las diferentes series de tiempo tomadas de las estaciones hidrométricas dentro de la cuenca del Amazonas.

La cuenca del Amazonas tiene más de seis millones de kilómetros cuadrados, repartida entre los países de Brasil, Perú, Ecuador, Bolivia y Venezuela. Ésta es una zona de selva tropical muy poco habitada por el humano. El río Amazonas contiene un quinto del agua dulce del mundo. El Amazona nace principalmente de dos ríos: el Ucayali y el Marañón. Ambos ríos nacen en los glaciares de los Andes peruanos. El río Amazonas aporta mucho sedimento a la zona.

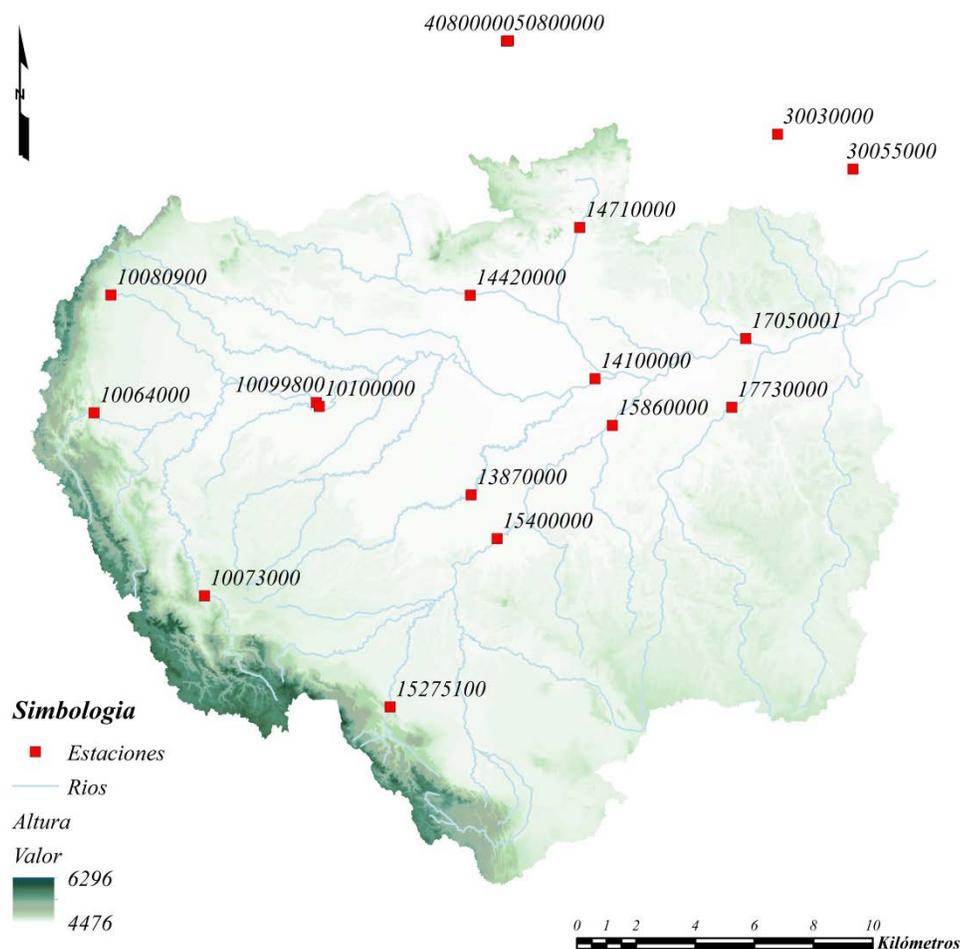


Figura 6.24. Distribución de las estaciones climatológicas en la cuenca del Amazonas.

La cuenca del Amazonas tiene aproximadamente 3.500 km en su extensión Este-Oeste y su ancho aproximadamente oscila entre 300 y 1.000 km. Una gran parte de esta cuenca está conformada por la selva pluvial tropical y es surcada por el río Amazonas. La inclinación del relieve dentro de la cuenca es relativamente baja, sólo en el corte occidental se alcanzan elevaciones sobre los 200 metros. En un recorrido de más de 3.500 km, desde la ciudad peruana de Iquitos hasta la desembocadura, el Amazonas desciende sólo 100 metros.

La cuenca puede dividirse en tres secciones:

- La parte superior de la cuenca está limitada (al Oeste) por Los Andes y llega hasta la confluencia de los ríos Negro y Solimões en Manaus. Esa zona sólo perteneció, transitoriamente, en el carbónico superior, a un espacio de sedimentación paleozóica de la cuenca amazónica.
- La sección media de la cuenca comprende desde Manaus hasta la desembocadura del río Xingú. Ahí la llanura disminuye y es acosada por series paleozóicas en el Norte y el Sur. En esta zona todos los afluentes descargan sus caudalosos torrentes desde

elevadas alturas con relativa rapidez en la cuenca del Amazonas, generalmente como raudales o cascadas.

- La sección baja de la cuenca es la misma desembocadura del Amazonas. La cuenca se abre hacia un inmenso estuario y el Amazonas se divide en varios ramales que bordean la Isla de Marajó.

Desde un punto de vista estructural, la cuenca se divide en tres arcos (o elevaciones) dispuestos transversalmente y surgidos en el Paleozoico:

- El Arco de Iquitos pasa por la ciudad de Iquitos en el Este y separa la subcuenca Acre del curso superior.
- Cerca de Manaus el Arco Perú es el límite de la sección superior y media de la cuenca del Amazonas y se halla cerca de la desembocadura del río Perú en el Solimões.
- El Arco Gurupá que pasa al Este de la desembocadura del río Xingú separa la sección media de la zona de la ría.

La zona de estudio correspondiente a la cuenca del Amazonas misma que se ha delimitado en regiones homogéneas mediante la técnica de Momentos-L. Para este estudio se utilizaron las series de acumulados mensuales de escurrimientos tomados de 18 estaciones hidrométricas de Observatory (ORE) HYBAM (Geodynamical, hydrological and biogeochemical control of erosion/alteration and material transport in the Amazon basin). La región en estudio denominada Cuenca del Amazonas se encuentra entre los 5° y 20° de latitud Sur y 50° y 80° Longitud Oeste. Se utilizaron series de escurrimientos acumulados de un mes, de dos meses y de seis meses de 18 estaciones hidrométricas (Figura 6.24) con longitud de registro de 1951 a 2012.

6.19 Preparación de datos

Los datos utilizados en este estudio corresponden al escurrimiento en la escala del acumulado mensual, acumulado de tres meses y acumulado de 6 meses. Para la detección y comprobación de la calidad de los datos se eliminaron los valores falsos asociados a su medición y a errores de transcripción, debido a la ausencia de datos en algunas estaciones se procedió a interpolar mediante la técnica regresión lineal, sin embargo la mayoría de las estaciones están completas logrando con ello resultados confiables. Las estaciones utilizadas en este estudio se presentan en la siguiente tabla.

Tabla 6.17. Estaciones hidrométricas en la Cuenca del Amazonas.

Altitud	Clave	Nombre	Latitud	Longitud
200	10064000	Rio Maranon at Borja	-4.4704	-77.5483
0	10073000	Rio Ucayali at Atalaya Aval	-10.6782	-73.8179
262	10080900	Rio Napo at Francisco de Orellana (Coca)	-0.4733	-76.9825
0	10099800	Rio Amazonas (Peru) at Nazareth	-4.1208	-70.0357
60	10100000	Rio Solimoes at Tabatinga	-4.2500	-69.9333
60	13870000	Rio Purus at Labrea	-7.2522	-64.8000
20	14100000	Rio Solimoes at Manacapuru	-3.3083	-60.6094
30	14420000	Rio Negro at Serrinha	-0.4819	-64.8289
0	14710000	Rio Branco at Caracarai	1.8214	-61.1236

210	15275100	Rio Beni at Rurrenabaque	-14.4453	-67.5343
55	15400000	Madeira at Porto Velho	-8.7367	-63.9203
20	15860000	Madeira at Fazenda Vista Alegre	-4.8972	-60.0253
0	17050001	Rio Amazonas at Obidos	-1.9472	-55.5111
0	17730000	Rio Tapajos at Itaituba	-4.2833	-55.9833
30	30055000	Rio Oyapock at Saut Maripa	3.8017	-51.8847
0	40800000	Rio Orinoco at Ciudad Bolivar	8.1432	-63.6074
270	50800000	Rio Congo at Congo à Beach Brazzaville	8.1536	-63.5361
25	30030000	Rio Maroni at Langa Tabiki	4.9862	-54.4368

6.20 Metodología

En los análisis anteriores se compararon los modelos RNAs, Ensamblado y Modular tradicionales y regionales mediante el enfoque propuesto. Los resultados han demostrado que los modelos Ensamblados regionales han sido superiores. En esta parte de la tesis se han realizado comparaciones en el pronóstico de series de escurrimientos con los modelos Ensamblado tradicional y Ensamblado mediante el enfoque regional propuesto, el cual está basado en la delimitación de regiones homogéneas mediante la técnica de Momentos-L. Los resultados de la comparación de modelos Ensamblados en Australia y Canadá demostraron que el algoritmo de clasificación AdaBoost.M2 es mejor predictor en lluvia y escurrimiento. De ahí que en esta parte de la tesis se comparan los modelos Ensamblados mediante RNAs (Ensamblado) y el modelo Ensamblado mediante el enfoque regional propuesto (Ensamblado-Regional). Los resultados demuestran que el pronóstico se puede mejorar mediante la inclusión de más estaciones que pertenecen a una región potencialmente homogénea. Se encontraron resultados prometedores al igual que en el modelo Ensamblado mediante el enfoque regional propuesto. Una de las ventajas de este enfoque es que al no contar con suficiente información en el tiempo es difícil extrapolar basándose en el comportamiento de la serie con la que se cuenta pues la incertidumbre es alta dependiendo de la cantidad de datos disponibles (Ashkar, et al.1993). Sin embargo este problema se puede resolver “paliando la carencia en el tiempo con la abundancia en el espacio” de acuerdo con Hosking (1997).

Los modelos ensamblados analizados tienen como entradas a los escurrimientos mensuales de uno, tres y seis meses de las estaciones que se encuentran dentro de una región potencialmente homogénea.

6.21 Delimitación de regiones homogéneas

La delimitación de las regiones homogéneas se llevó a cabo mediante la técnica de Momentos-L, este procedimiento consiste básicamente en definir regiones estadísticamente homogéneas en una zona determinada. Para determinar que un grupo de estaciones conforman una región, las estaciones deben cumplir el criterio de homogeneidad. La identificación de regiones homogéneas se realizó mediante análisis clúster c-medias de los estadísticos coeficiente de variación L (L-CV) y el coeficiente de asimetría L (Skewness-L), los grupos entonces son medidos a través de un proceso iterativo hasta que finalmente se obtiene regiones aceptablemente homogéneas. Una vez teniendo una región propuesta con N sitios donde cada

sitio i tiene una longitud de registro n_i y sus radios de momentos L t^i , t_3^i y t_4^i y se denotan como t^R , t_3^R y t_4^R siendo los estadísticos media regional L , coeficiente de asimetría L y a la curtosis L siendo proporcionales a los estadísticos de cada uno de los sitios de la región. Después se encontraron los radios de momentos L regionales y con la distribución kappa se simularon 500 muestras y se calculó V de las simulaciones asimismo la media y la desviación estándar μv y σv . de acuerdo a Hosking & Wallis (1997) si H es lo suficientemente grande en la literatura se sugiere $H \geq 2$ en caso contrario se tiene una región homogénea.

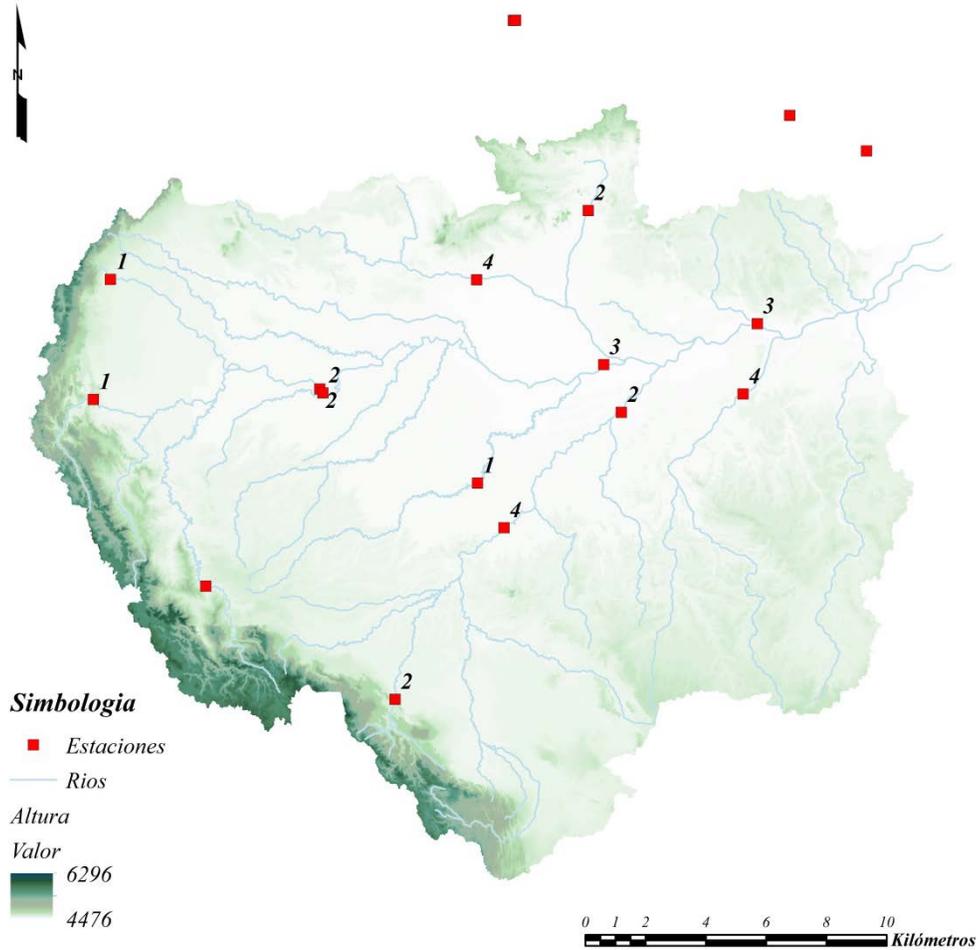


Figura 6.25. Regiones homogéneas en la cuenca del Amazonas.

6.22 Análisis clúster

De la región en estudio cuenca del Amazonas se escogieron 13 estaciones de las 18. De las estaciones utilizadas se formaron 4 grupos. Estas formaciones iniciales se hicieron mediante la técnica de agrupamiento de c-medias y dendograma mediante la distancia euclídea (Tabla. 6.18).

Tabla 6.18. Resultados por clase

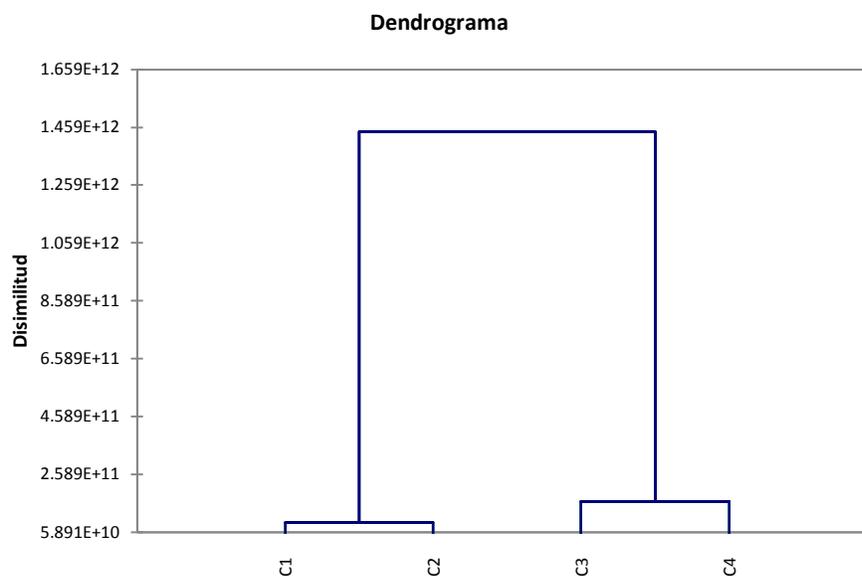
Clase	1	2	3	4
Objetos	5	3	2	3
Suma de los pesos	5	3	2	3
Varianza intraclase	613620296.888	1053672384.282	164992092125.440	1917009645.802
Distancia mínima al centroide	11520.156	16488.723	287221.249	18203.135
Distancia media al centroide	20099.943	24979.225	287221.249	33536.719
Distancia máxima al centroide	37739.925	37203.412	287221.249	48524.934
	EST-10064000	EST-10099800	EST-14100000	EST-14420000
	EST-10080900	EST-10100000	EST-17050001	EST-15400000
	EST-13870000	EST-15860000		EST-17730000
	EST-14710000			
	EST-15275100			

La distancia entre los objetos centrales es como sigue.

Tabla 6.19. Distancias entre los objetos centrales:

Estaciones	1 (EST-14710000)	2 (EST-10100000)	3 (EST-14100000)	4 (EST-14420000)
1 (EST-14710000)	0	238784.151	738384.531	108061.625
2 (EST-10100000)	238784.151	0	501011.526	133029.354
3 (EST-14100000)	738384.531	501011.526	0	631592.285
4 (EST-14420000)	108061.625	133029.354	631592.285	0

Quedando representado en el dendrograma de la siguiente manera.

**Figura 6.26.** Dendrograma de los grupos formados en el análisis clúster

De las regiones agrupadas en el análisis clúster se ha procedido a determinar mediante la prueba de discordancia si realmente son homogéneas, esto es que todos los grupos presenten valores menores a 1. Posteriormente se determinan la medida de heterogeneidad mediante el estadístico H .

Tabla 6.20. Grupos de estaciones utilizadas

grupo	Estación	t1	t3	t4
1	EST-10064000	-0.048	0.066	0.677
	EST-10080900	-0.255	-0.361	0.009
	EST-13870000	0.003	-0.914	0.478
	EST-14710000	-0.052	0.250	0.118
	EST-15275100	-0.037	-1.111	-0.442
2	EST-15860000	0.021	0.134	-0.001
	EST-10099800	-0.015	0.621	0.154
	EST-10100000	-0.012	0.644	0.468
3	EST-14100000	-0.022	-0.036	0.187
	EST-17050001	-0.022	-0.073	0.008
	EST-14420000	-0.001	0.568	-0.276
4	EST-15400000	0.021	0.461	0.316
	EST-17730000	-0.004	-0.576	0.535

En este estudio se han analizado los grupos 1 al 4 (Cuadro 1). Para el grupo 1 los resultados obtenidos del ajuste de la distribución Kappa a sus parámetros regionales son $t^R = -0.0777$, $t_3^R = -0.8138$ y $t_4^R = 0.3679$ con una $V = 0.0906$ con ello se obtuvo una $H = 1.1139$ aunque el valor de H es mayor que 1 está dentro del 10% de tolerancia ya que el valor de H no es estrictamente de 1.

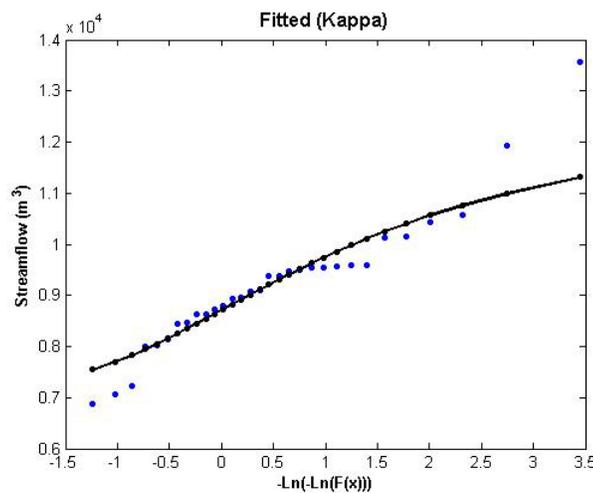


Figura 6.27. Ajuste de la distribución Kappa al grupo homogéneo 1.

Del grupo 2 se han obtenido del ajuste de la distribución Kappa a sus parámetros regionales los siguientes resultados $t^R = -0.0190$, $t_3^R = 0.1077$ y $t_4^R = -0.1406$ con una $V = 0.02500$ de lo cual se obtuvo una $H = 0.6264$. Las dos regiones que se han tomado para el análisis regional han demostrado ser homogéneas de acuerdo con Hosking y Wallis (1997) pues H es menor que el valor propuesto de 1.

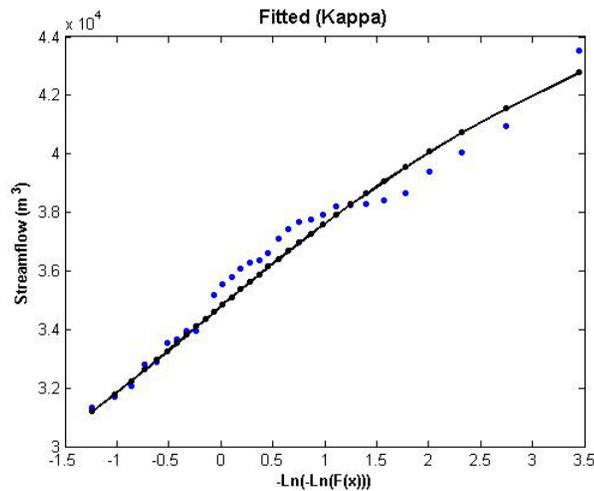


Figura 6.28. Ajuste de la distribución Kappa al grupo homogéneo 2.

Para el grupo 3 los resultados obtenidos del ajuste de la distribución Kappa a sus parámetros regionales son $t^R = -0.0223$, $t_3^R = -0.0548$ y $t_4^R = 0.0975$ con una $V = 0.001487$ con ello se obtuvo una $H = 0.0067$ demostrando ser homogénea.

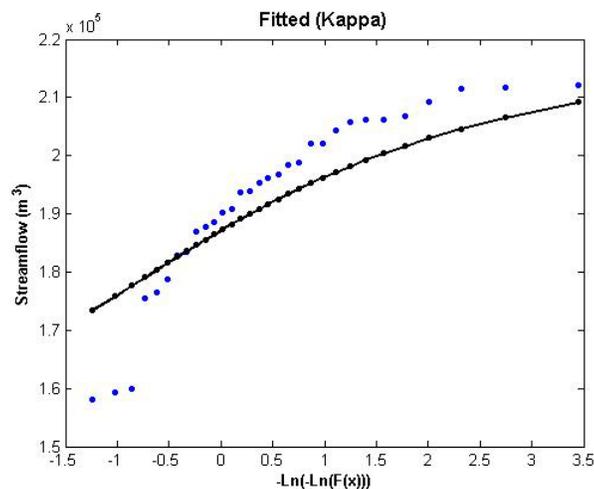


Figura 6.29. Ajuste de la distribución Kappa al grupo homogéneo 3.

Finalmente para el grupo 4 los resultados obtenidos del ajuste de la distribución Kappa a sus parámetros regionales son $t^R = 0.0056$, $t_3^R = 0.1489$ y $t_4^R = 0.8073$ con una $V = 0.0112$ con ello se obtuvo una $H = 0.7318$ demostrando ser homogénea.

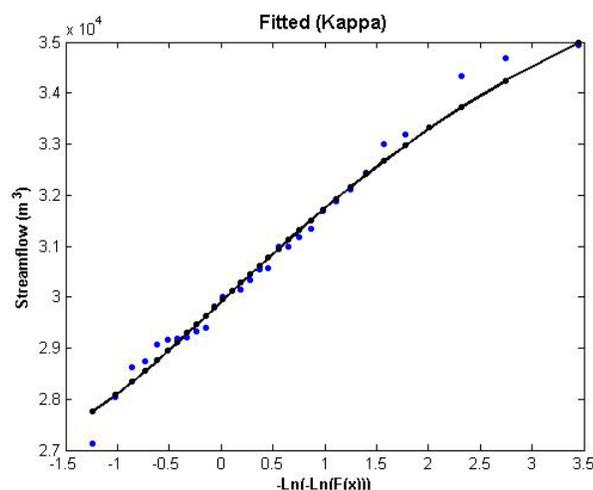


Figura 6.30. Ajuste de la distribución Kappa al grupo homogéneo 4.

6.23 Análisis de series mensuales

En el análisis anterior se realizó la comparación entre el modelo Ensamblado de predicción tradicional basado en RNAs y Ensamblado mediante el enfoque regional propuesto. Esto demostró que el enfoque regional propuesto es altamente competitivo. De esta forma los resultados encontrados en la comparación del modelo ensamblado tradicional y el modelo ensamblado mediante el enfoque regional han sido satisfactorios con los objetivos de esta tesis.

En la tabla 6.21 se muestran los resultados de la pruebas de bondad de ajuste obtenidos en el modelo Ensamblado regional propuesto y el modelo Ensamblado tradicional aplicados a la serie de acumulados mensuales de escurrimientos de la región 2, esto es en las diferentes regiones homogéneas de la cuenca del Amazonas. Los resultados indican que el modelo Ensamblado regional es superior al modelo tradicional en un 90%. Puesto que en todas las estaciones se obtuvieron de acuerdo a las pruebas de bondad de ajuste siempre los menores errores.

Tabla 6.21 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región 2

Estaciones	Ensamblado-Regional				Ensamblado tradicional			
	MSE	RMSE	MAE	<i>d</i>	MSE	RMSE	MAE	<i>d</i>
14710000	3619543.19	1902.51	1410.80	0.75	6329292.70	2515.81	1845.46	0.62
15275100	763612.27	873.85	533.81	0.84	1249813.49	1117.95	715.97	0.72
15860000	11397918.57	3376.08	2664.99	0.97	17514220.96	4185.00	3217.97	0.96
10099800	22890060.95	4784.36	3607.27	0.91	26235696.88	5122.08	4203.23	0.88
10100000	23772326.57	4875.69	3651.25	0.91	26946169.36	5190.97	4242.12	0.89

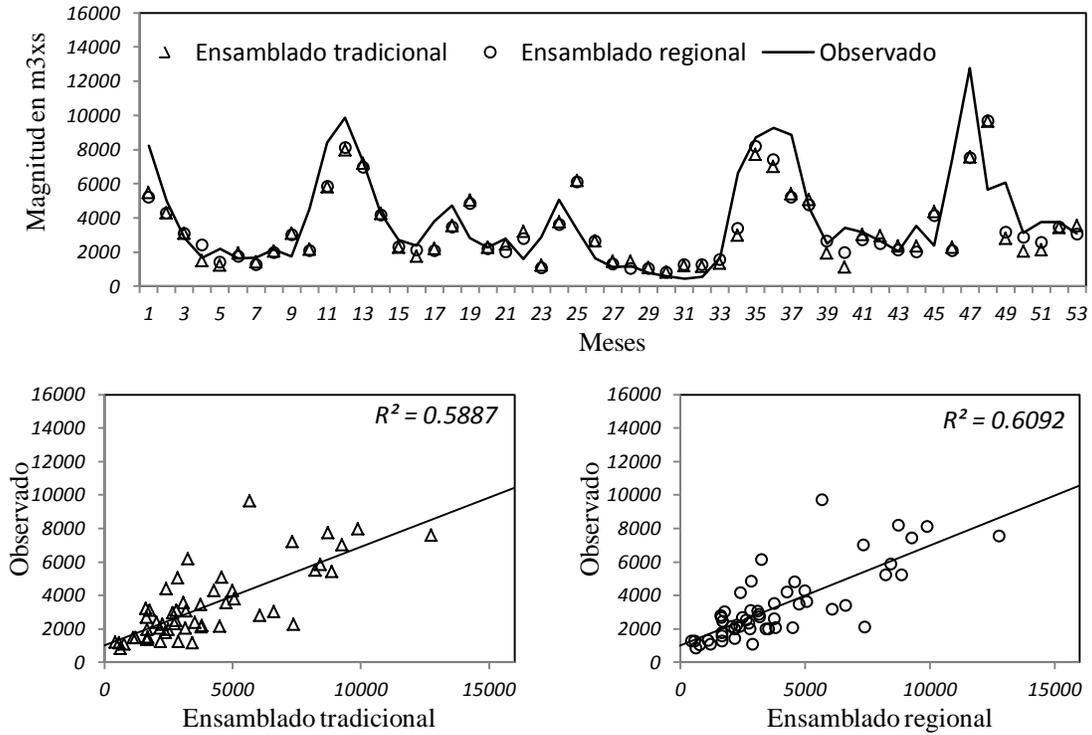


Figura 6.31. Pronostico mensual de escurrimiento en la estación 14710000

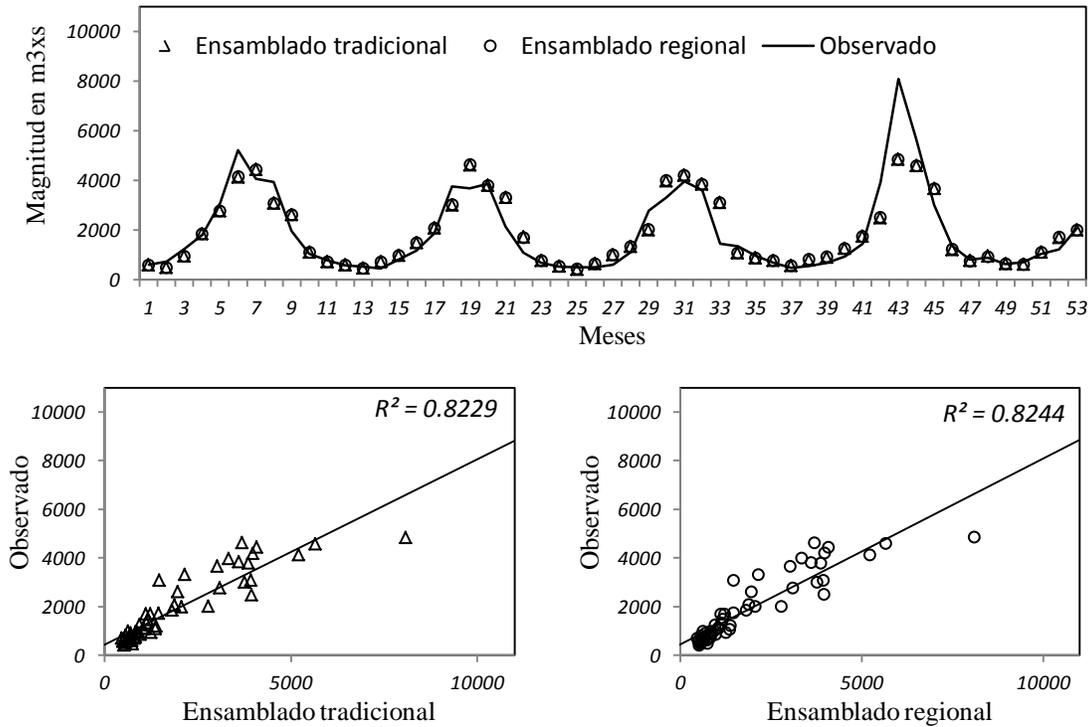


Figura 6.32. Pronostico mensual de escurrimiento en la estación 15275100

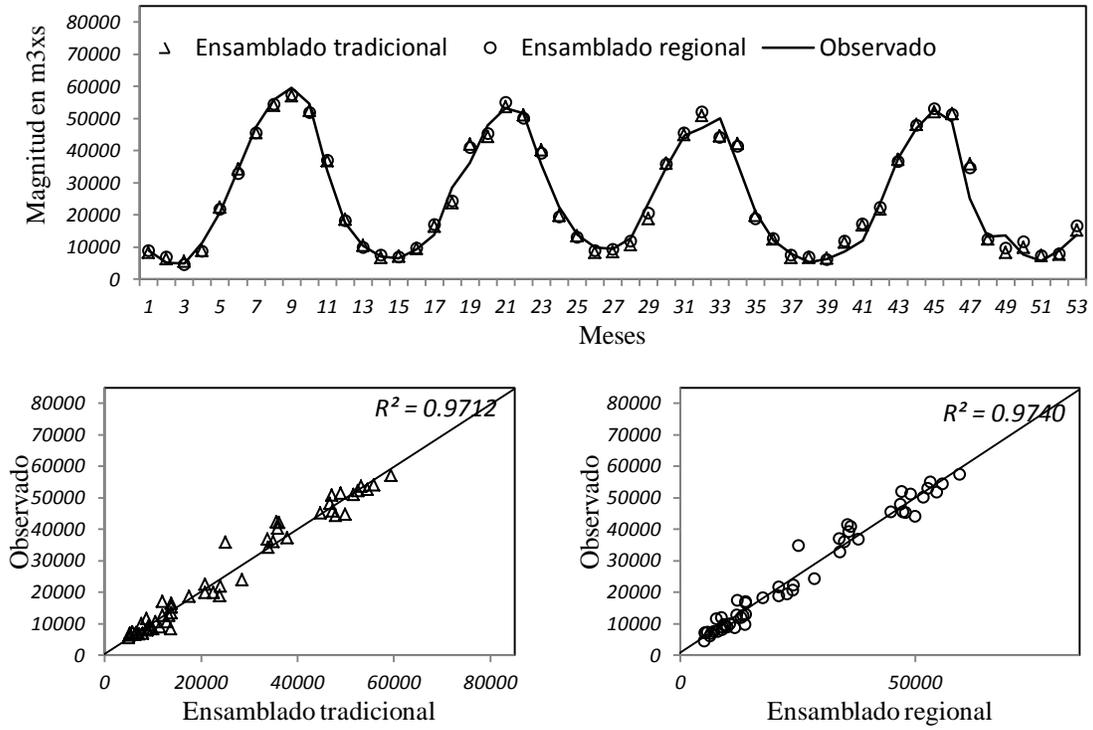


Figura 6.33. Pronostico mensual de escurrimiento en la estación 15860000

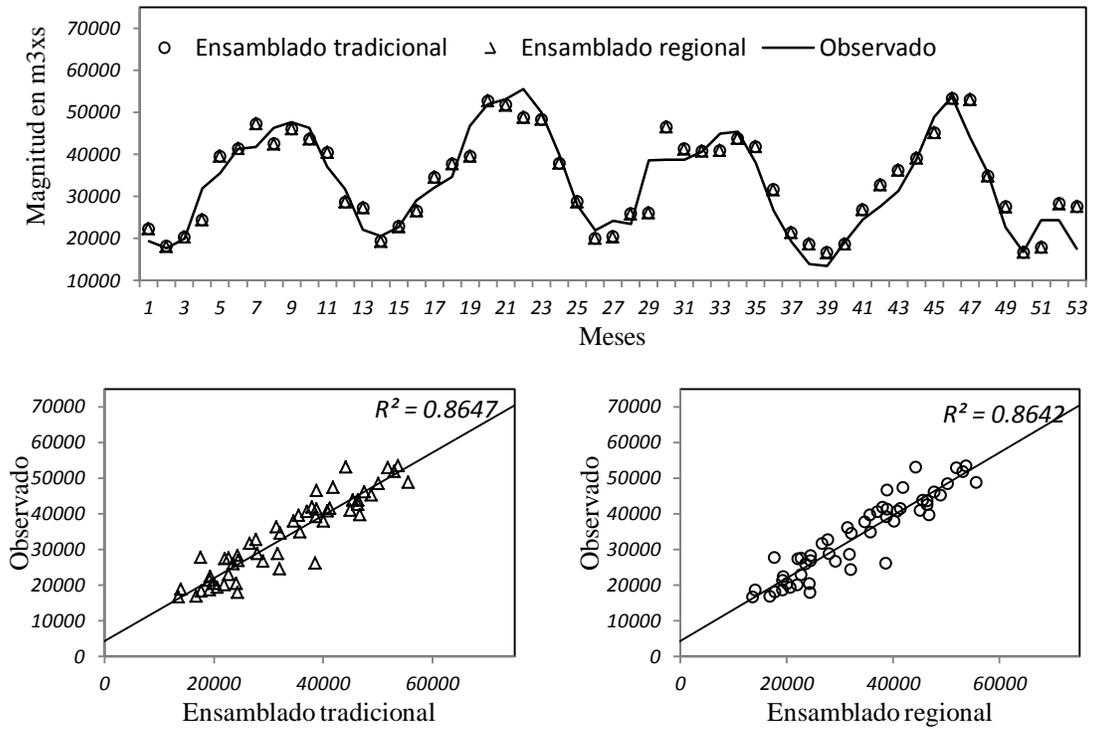


Figura 6.34. Pronostico mensual de escurrimiento en la estación 10099800

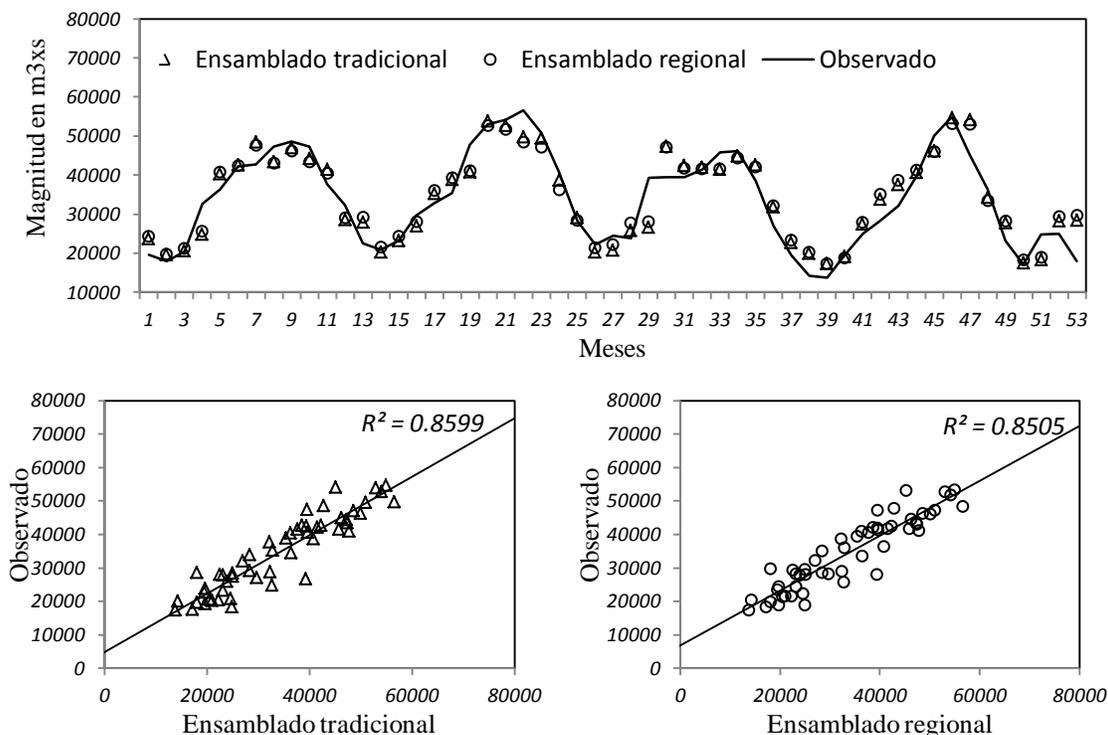


Figura 6.35. Pronostico mensual de escurrimiento en la estación 10100000

6.24 Análisis de series trimestrales

Los resultados obtenidos del análisis de escurrimientos acumulados trimestrales mediante los modelos Ensamblado tradicional y regional se muestran en la tabla 6.22. Este análisis se realizó para las diferentes regiones de la cuenca del Amazonas. Para la región dos los resultados indican, que en general el modelo ensamblado fue superior al modelo tradicional. Se observó que en más del 80% el modelo Ensamblado regional propuesto ha sido superior al modelo tradicional. Se concluye que en general el modelo Ensamblado regional propuesto es superior al modelo ensamblado tradicional.

Tabla 6.22. Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región 2

Estaciones	Ensamblado-Regional				Ensamblado			
	MSE	RMSE	MAE	d	MSE	RMSE	MAE	d
14710000	34088093.90	5838.50	4918.64	0.58	38651998.94	6217.07	5360.74	0.50
15275100	6595004.95	2568.07	1418.28	0.78	3250810.91	1803.00	1288.17	0.89
15860000	200156099.62	14147.65	11404.95	0.95	240698530.78	15514.46	11556.47	0.95
10099800	289312298.94	17009.18	13823.13	0.84	395936386.63	19898.15	15351.78	0.84
10100000	301072282.14	17351.43	14133.78	0.84	411298613.80	20280.50	15538.08	0.84

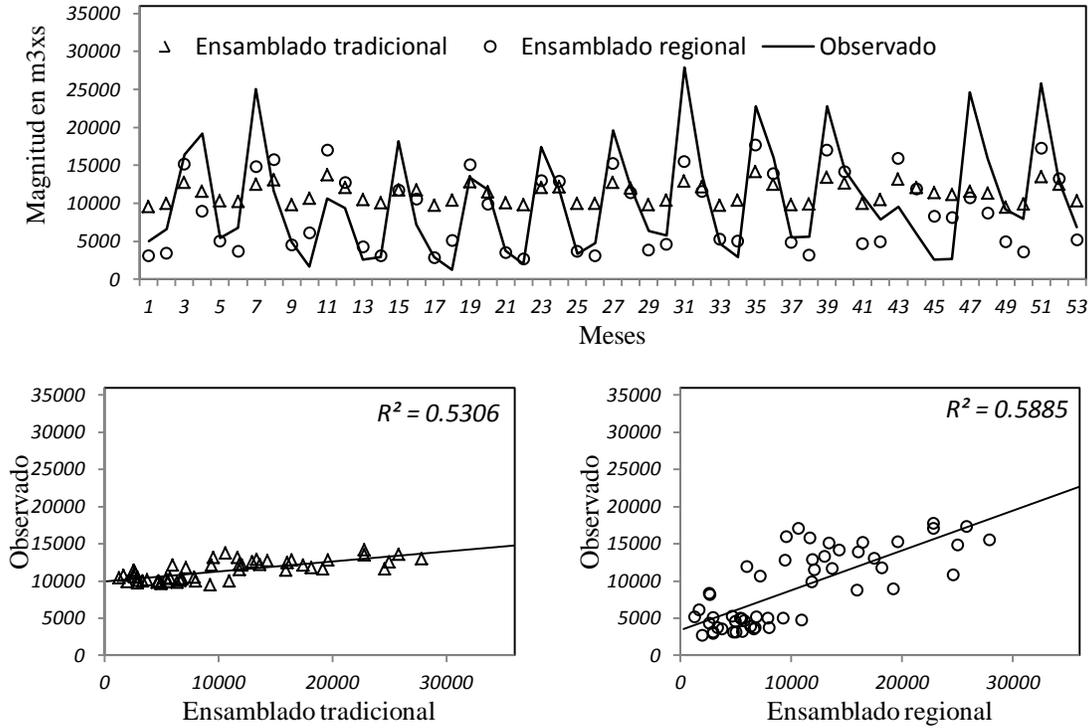


Figura 6.36. Pronostico trimestral de escurrimiento en la estación 14710000

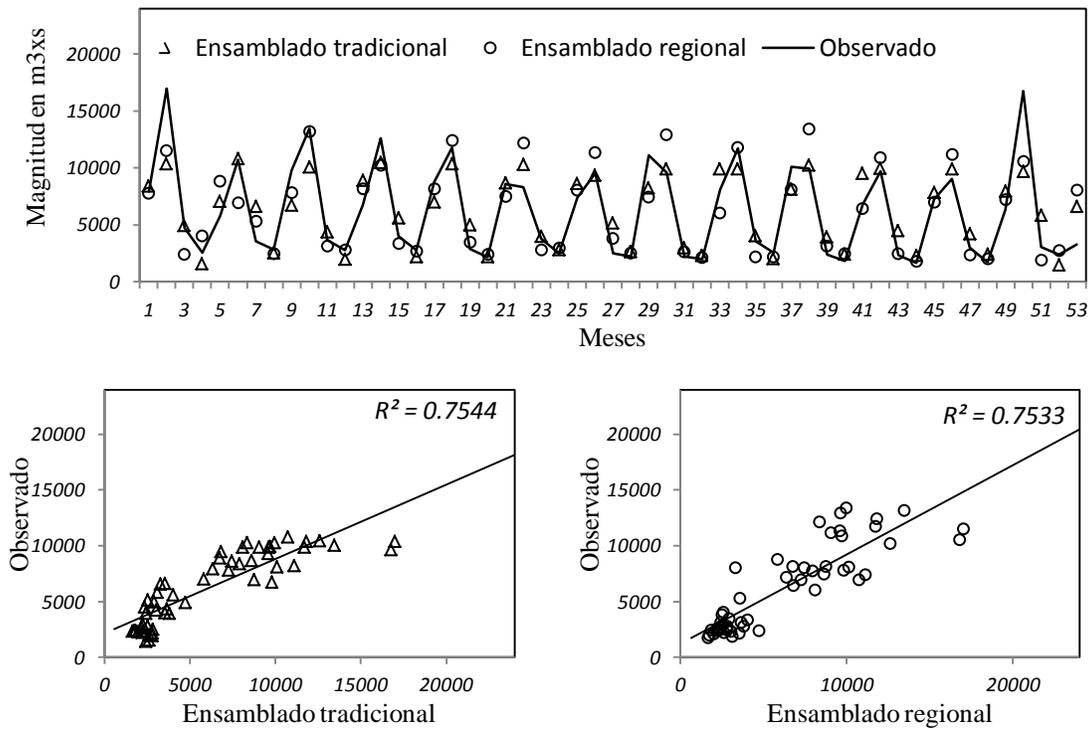


Figura 6.37. Pronostico trimestral de escurrimiento en la estación 15275100

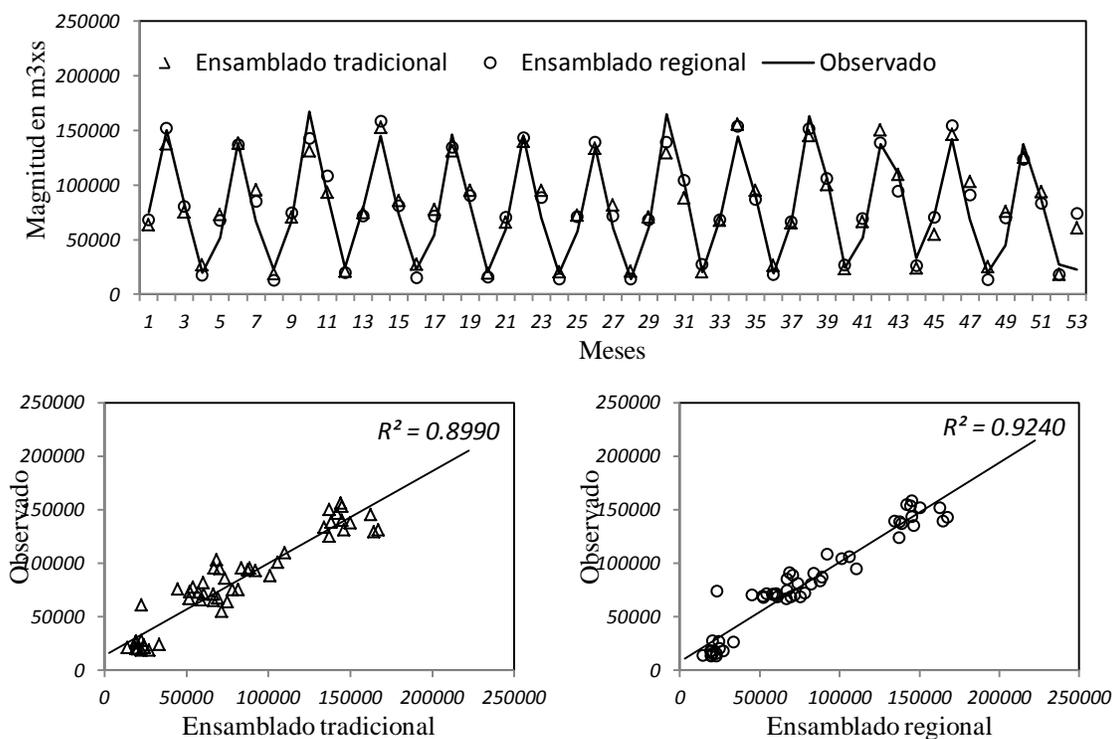


Figura 6.38. Pronostico trimestral de escurrimiento en la estación 15860000

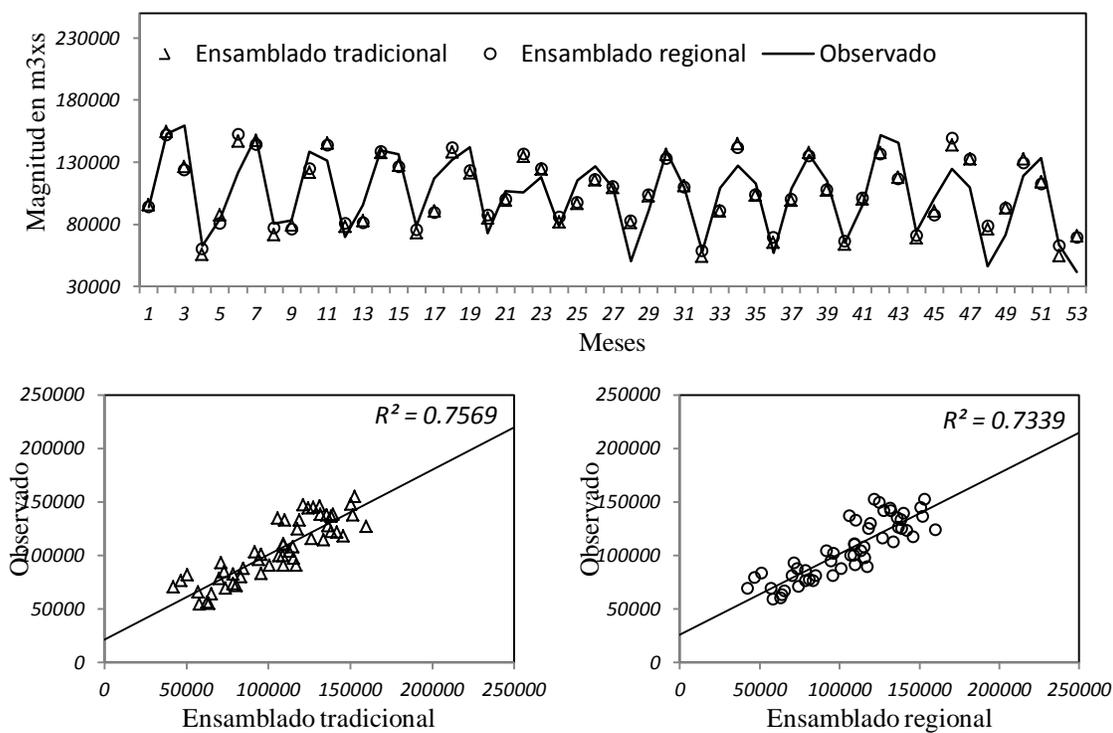


Figura 6.39. Pronostico trimestral de escurrimiento en la estación 10099800

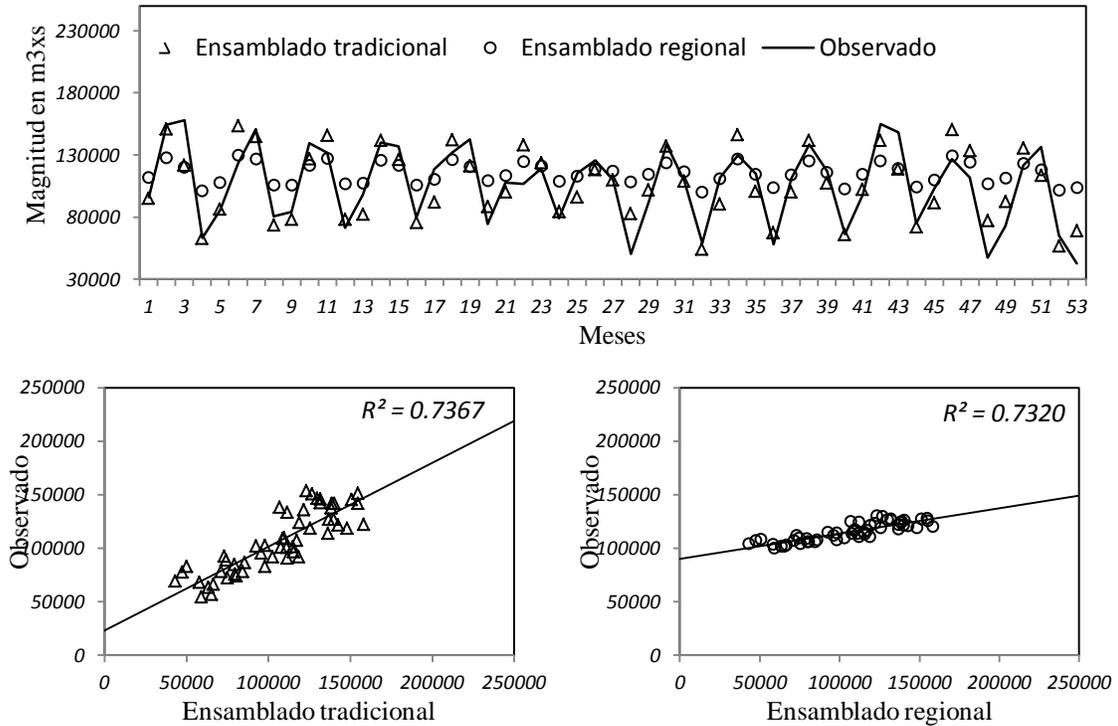


Figura 6.40. Pronostico trimestral de escurrimiento en la estación 10100000

6.25 Análisis de series semestrales

En la tabla 6.23 se muestran los resultados obtenidos en el modelo Ensamblado regional propuesto y el modelo tradicional de redes neuronales, esto es para los escurrimientos acumulados de seis meses de la región dos de la cuenca del Amazonas. En general el modelo Ensamblado regional ha sido más eficiente en el 83% de todas las estaciones analizadas que el modelo tradicional en el pronóstico de escurrimientos de acuerdo a las pruebas de bondad de ajuste usadas en esta tesis.

Tabla 6.23 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región 2

Estaciones	Ensamblado-Regional				Ensamblado			
	MSE	RMSE	MAE	d	MSE	RMSE	MAE	d
14710000	54127474.17	7357.14	6812.77	0.27	91504407.37	9565.79	8666.02	0.27
15275100	2092120.35	1446.42	1137.26	0.98	8172887.51	2858.83	2247.97	0.93
15860000	850814418.37	29168.72	21616.82	0.85	893085287.49	29884.53	24473.91	0.86
10099800	575306612.00	23985.55	19576.66	0.61	2331205883.91	48282.56	36698.35	0.50
10100000	502589054.17	22418.50	16768.79	0.72	2464346299.15	49642.18	38081.41	0.50

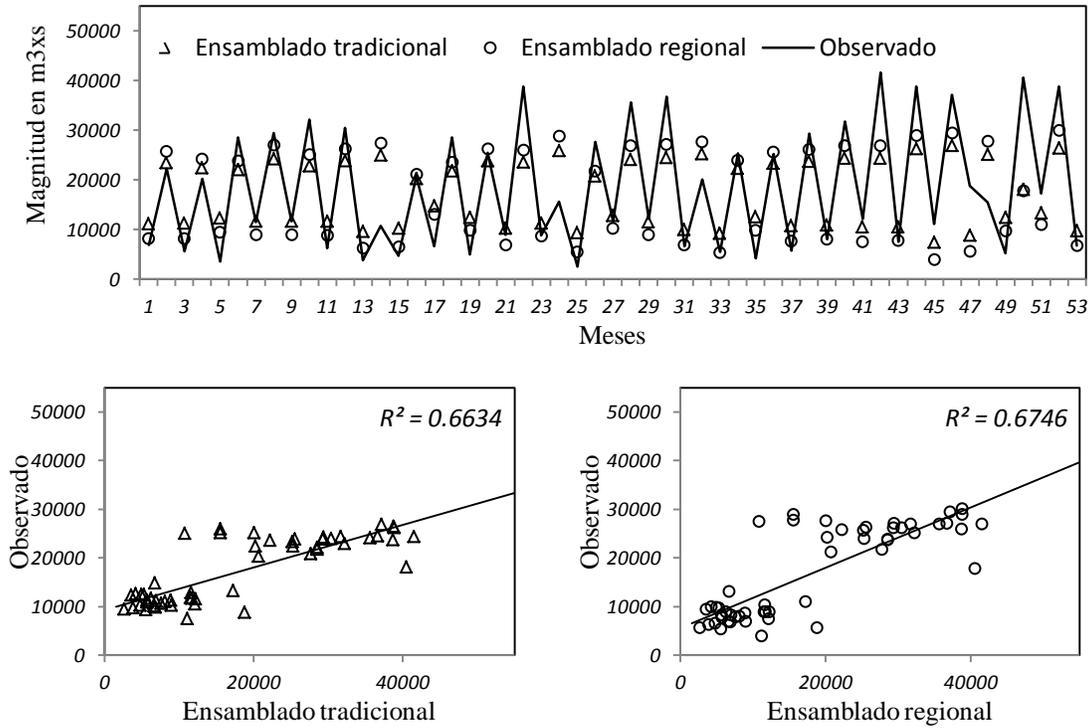


Figura 6.41. Pronostico semestral de escurrimiento en la estación 14710000

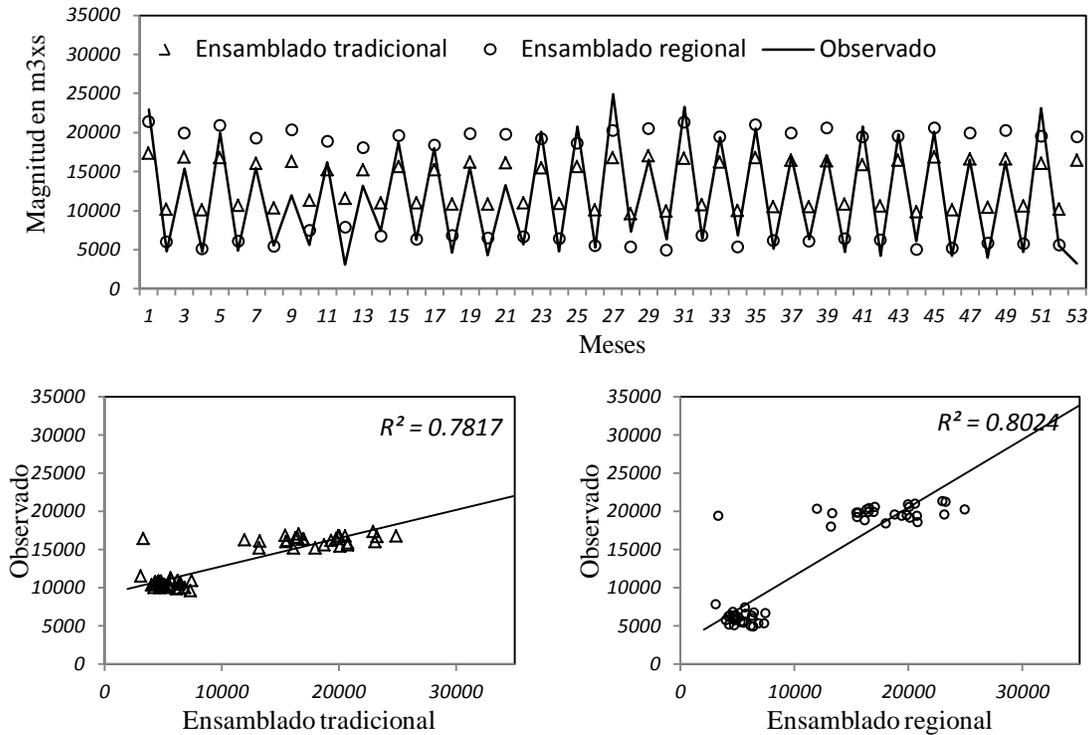


Figura 6.42. Pronostico semestral de escurrimiento en la estación 15275100

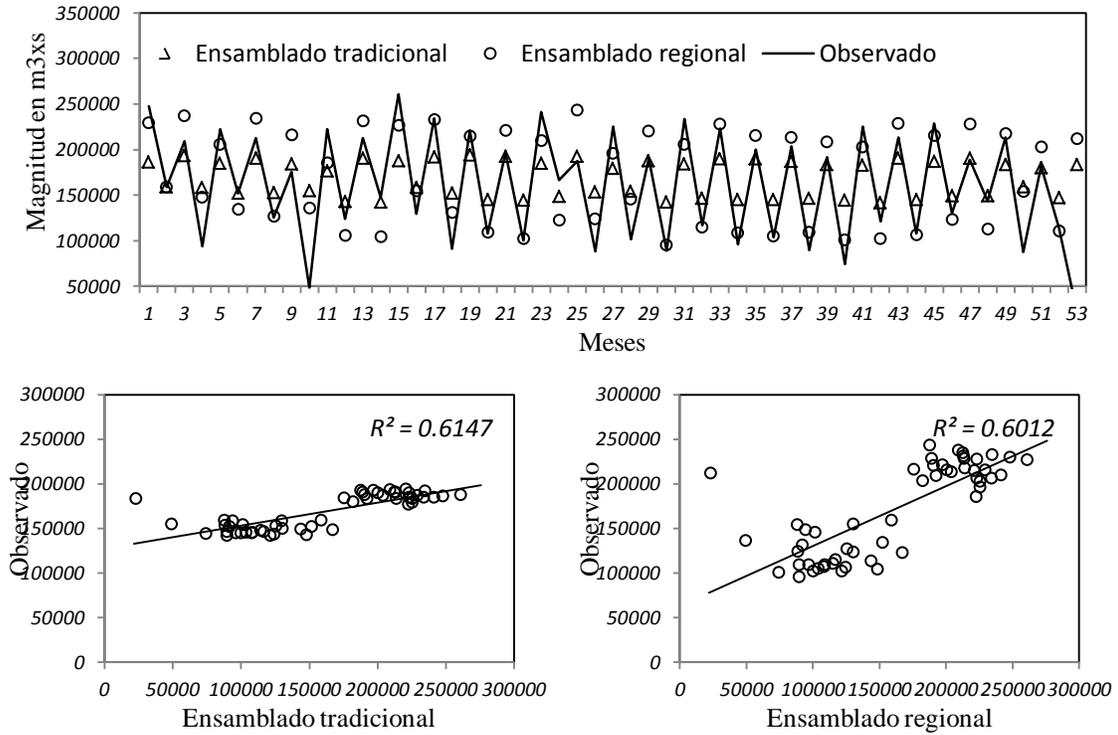


Figura 6.43. Pronostico semestral de escurrimiento en la estación 15860000

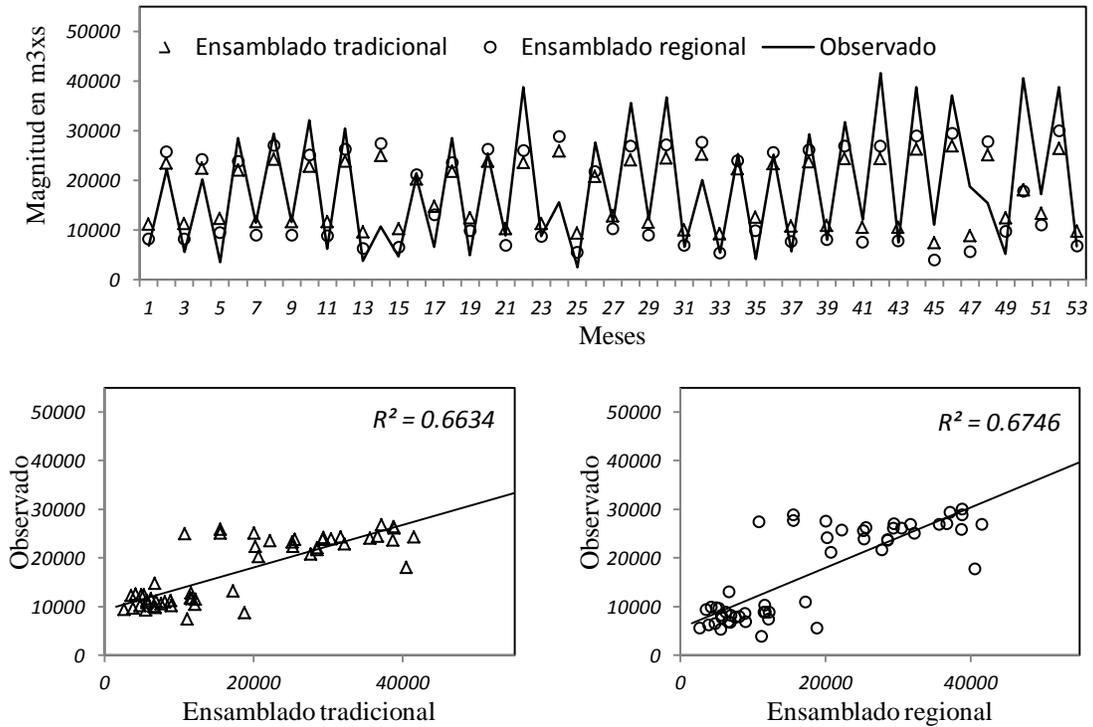


Figura 6.44. Pronostico semestral de escurrimiento en la estación 10099800

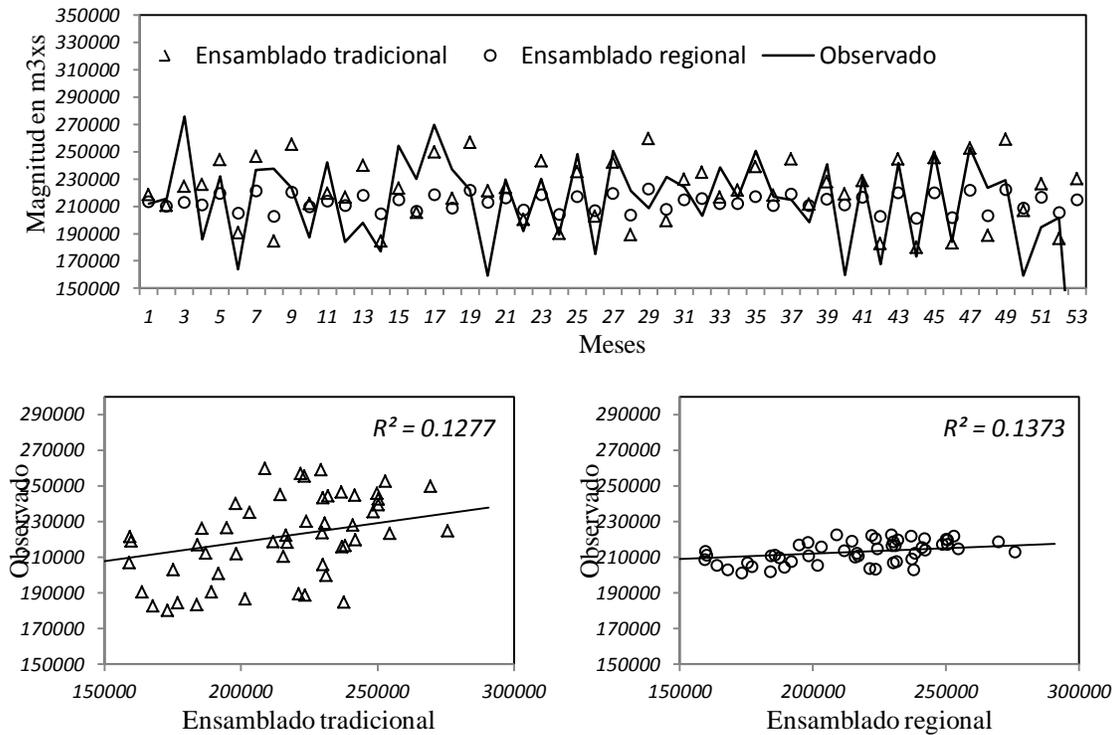


Figura 6.45. Pronostico semestral de escurrimiento en la estación 10100000

6.26 Comparación de lluvias en las escalas mensual trimestral y semestral.

La comparación de lluvias en la escala de tiempo mensual, trimestral y semestral se realizó mediante un modelo Ensamblado basado en el enfoque regional propuesto y el otro mediante el enfoque tradicional. La evaluación de los modelos fue mediante las medidas de eficiencia tradicionales. Las series de lluvia fueron tomadas de las estaciones climatológicas en la región denominada Costa Oeste de Estados Unidos.

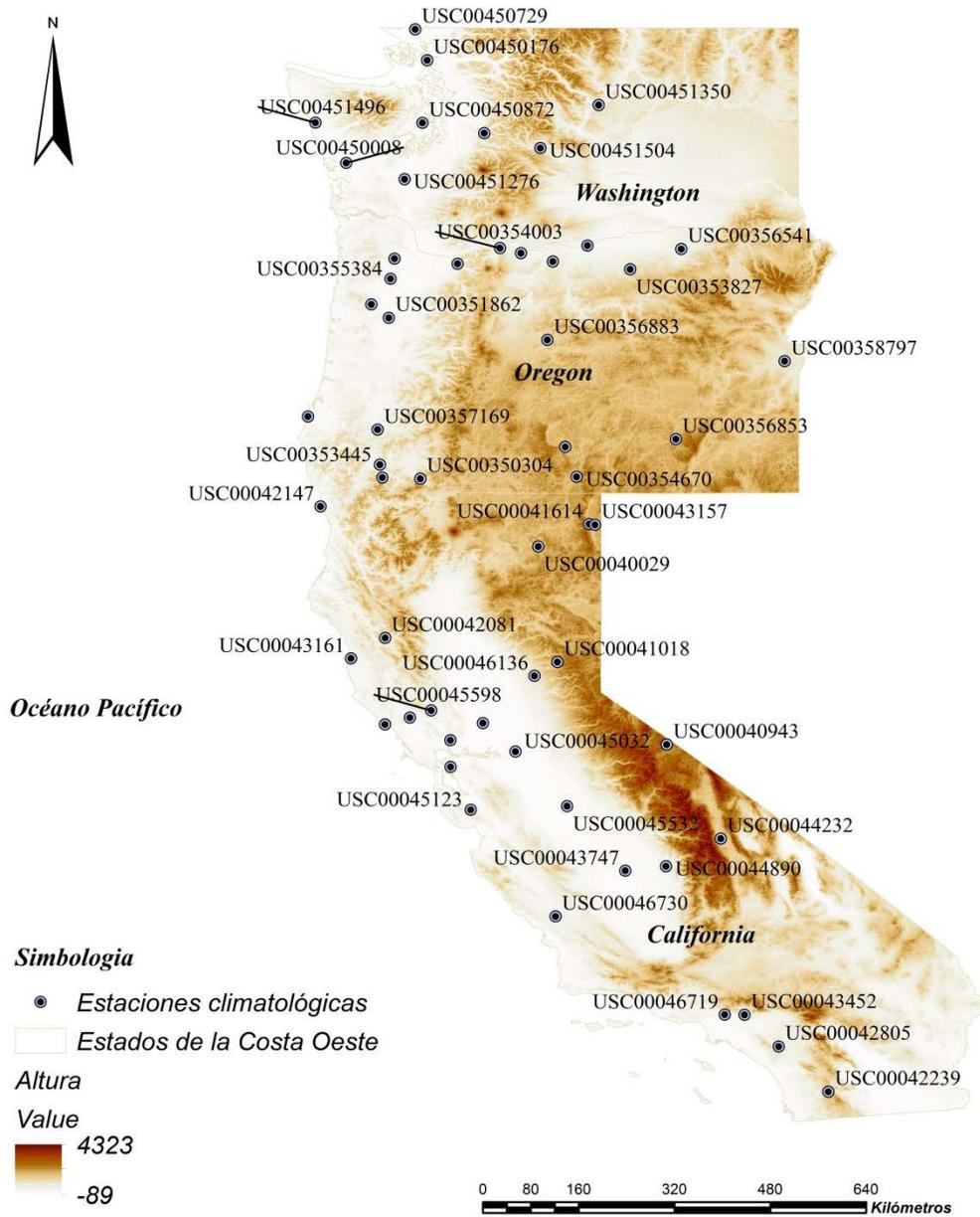


Figura 6.46. Estaciones climatológicas en la Costa Oeste de Estados Unidos.

La región de estudio corresponde a la Costa Oeste de Estados Unidos en la cual se han delimitados regiones homogéneas mediante la técnica de Momentos-L para propósitos de este análisis. En este estudio se utilizaron las series de acumulados mensuales de lluvia tomados de 56 estaciones climatológicas de la red de estaciones de la National Oceanic and Atmospheric Administration (NOOA) ubicadas en la Costa Oeste de Estados Unidos.

La región Costa Oeste está integrada por los estados de California, Oregón y Washington, aunque no forma parte de la Costa Oeste, Alaska y Hawái bordean el océano Pacífico y podrían incluirse en esta región. La población estimada de la Costa Oeste, en 2007, estaba entre los 50 y los 60 millones, incluyendo los tres primeros estados mencionados.

Existe un término más antiguo con el cual se denomina a la región en estudio que se cree que surgió en el Noreste de Estados Unidos y que alude a la Costa Oeste nombrándola simplemente como “La Costa”. Las principales ciudades costeras de la Costa Oeste son San Diego, Los Ángeles, San Francisco y Seattle. Entre las de interior se encuentran Portland (Oregón), Las Vegas y Sacramento (California).

La región en estudio denominada Costa Oeste de Estados Unidos se encuentra entre los 32° y 50° de latitud Norte y 114° y 126° Longitud Oeste (Figura 1). Se utilizaron series de lluvia acumulada de un mes, de dos meses y de seis meses de 56 estaciones climatológicas (Fig. 2.1) con longitud de registro de 1900 a 2012 ($n = 113$).

6.27 Preparación de datos

Los datos utilizados en este estudio corresponden a la variable precipitación de lluvia en la escala acumulado mensual, acumulado de tres meses y acumulado de seis meses. Para la detección y comprobación de la calidad de los datos se eliminaron los valores falsos asociados a su medición y a errores de transcripción, debido a la ausencia de datos en algunas estaciones se procedió a interpolar mediante la técnica interpolación inversa (IDW), sin embargo la mayoría de las estaciones están completas logrando con ello resultados bastante confiables.

Tabla 6.24. Estaciones climatológicas en la Costa Oeste de EEUU

Clave	Nombre	Latitud	Longitud	Elevación
USC00040029	CA ADIN RS	41.1936	-120.9447	1278.6
USC00040693	CA BERKELEY	37.8744	-122.2606	94.5
USC00040943	CA BODIE CA ST HISTORIC PARK	38.2119	-119.0142	2551.2
USC00041018	CA BOWMAN DAM	39.4539	-120.6556	1641.3
USC00041614	CA CEDARVILLE	41.5306	-120.1803	1431
USC00042081	CA COVELO	39.8158	-123.2444	430.7
USC00042147	CA CRESCENT CITY NNW	41.7958	-124.2147	13.1
USC00042239	CA CUYAMACA	32.9897	-116.5872	1414.3
USC00042294	CA DAVIS WSW EXP FARM	38.535	-121.7761	18.3
USC00042805	CA ELSINORE	33.6692	-117.3319	391.7
USC00043157	CA FT BIDWELL	41.52	-120.09	1442.3
USC00043161	CA FT BRAGG N	39.5092	-123.7567	37.5
USC00043191	CA FT ROSS	38.515	-123.2447	34.1
USC00043452	CA GLENDORA FC 287B	34.1472	-117.8472	284.4
USC00043747	CA HANFORD S	36.3158	-119.6369	72.2
USC00043875	CA HEALDSBURG	38.6175	-122.8731	32.9
USC00044232	CA INDEPENDENCE	36.7981	-118.2036	1204
USC00044890	CA LEMON COVE	36.3817	-119.0264	156.4
USC00045032	CA LODI	38.1061	-121.2878	12.2
USC00045123	CA LOS GATOS	37.2319	-121.9592	111.3
USC00045532	CA MERCED	37.2858	-120.5117	46.6
USC00045598	CA MIDDLETOWN 4SE	38.725	-122.5514	340.8
USC00046074	CA NAPA STATE HOSPITAL	38.2778	-122.2647	10.7
USC00046136	CA NEVADA CITY	39.2467	-121.0008	847.6

USC00046719	CA PASADENA	34.1483	-118.1447	263.3
USC00046730	CA PASO ROBLES	35.6278	-120.6856	222.5
USC00350265	OR ARLINGTON	45.7211	-120.2064	84.4
USC00350304	OR ASHLAND	42.2128	-122.7144	532.2
USC00350471	OR BANDON NNE	43.1497	-124.4019	6.1
USC00351862	OR CORVALLIS STATE UNIV	44.6342	-123.19	68.6
USC00352800	OR FALLS CITY SSW	44.8358	-123.4525	210.3
USC00352997	OR FOREST GROVE	45.5247	-123.1028	54.9
USC00353445	OR GRANTS PASS	42.4244	-123.3236	283.5
USC00353770	OR HEADWORKS PORTLAND	45.4486	-122.1547	228
USC00353827	OR HEPPNER	45.3653	-119.5639	574.5
USC00354003	OR HOOD RIVER EXP STN	45.6847	-121.5175	152.4
USC00354670	OR LAKEVIEW NNW	42.2414	-120.3678	1490.5
USC00355384	OR MC MINNVILLE	45.2214	-123.1622	47.2
USC00355734	OR MORO	45.4825	-120.7236	570
USC00356426	OR PAISLEY	42.6922	-120.5403	1328.9
USC00356541	OR PENDLETON DWTN	45.6703	-118.7964	317
USC00356853	OR P-RANCH SOUTH PLACE	42.8081	-118.8778	1283.2
USC00356883	OR PRINEVILLE	44.3022	-120.8081	888.5
USC00357169	OR RIDDLE	42.9506	-123.3572	207.3
USC00358407	OR THE DALLES	45.6069	-121.2047	45.7
USC00358797	OR VALE	43.9814	-117.2439	682.8
USC00359390	OR WILLIAMS NW	42.2283	-123.2858	442
USC00450008	WA ABERDEEN	46.9658	-123.8292	3
USC00450176	WA ANACORTES	48.5119	-122.6136	6.1
USC00450729	WA BLAINE	48.9775	-122.7928	18.3
USC00450872	WA BREMERTON	47.5689	-122.6828	33.5
USC00451233	WA CEDAR LAKE	47.4144	-121.7561	475.5
USC00451276	WA CENTRALIA	46.72	-122.9528	56.4
USC00451350	WA CHELAN	47.8361	-120.0381	335
USC00451496	WA CLEARWATER	47.5711	-124.2922	24.4
USC00451504	WA CLE ELUM	47.1889	-120.9131	579.1

6.28 Metodología

En el análisis de escurrimientos se ensayó un modelo Ensamblado el cual brindó resultados satisfactorios en el pronóstico mensual, trimestral y semestral demostrando ser lo suficientemente eficiente. En los Modelos ensamblados que se presentan para este análisis de la variable lluvia, las estradas que alimentan al modelo corresponden a la lluvia acumulada mensual, trimestral y semestral de varias estaciones que integran a una misma región homogénea. El algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) fue utilizado para entrenar a los ensambles mediante RNAs ya que se ha demostrado en esta tesis que para el entrenamiento de las RNAs en comparación con diversos algoritmos de optimización no lineal. El algoritmo de Levenberg Marquardt ha sido el mejor.

En el Ensamblaje se ha aprovechado la bondad de los modelos no lineales basados en RNAs y mediante el enfoque regional se ha demostrado ser altamente eficiente en el pronóstico debido a una mayor gama de información espacial. El entrenamiento de los modelos es como sigue; la entrada de datos son las series acumuladas mensuales, trimestrales y semestrales, entonces una vez entrenada cada RNAs en el ensamble la asignación final de los pesos corresponden a un clasificador. Este puede ser AdaBoost.M2, LPBoost, TotalBoost y RusBoost, se ha tomado el que mejor resultado brinda logrando con ello un mejor pronóstico en variable estudiada.

6.29 Delimitación de regiones homogéneas

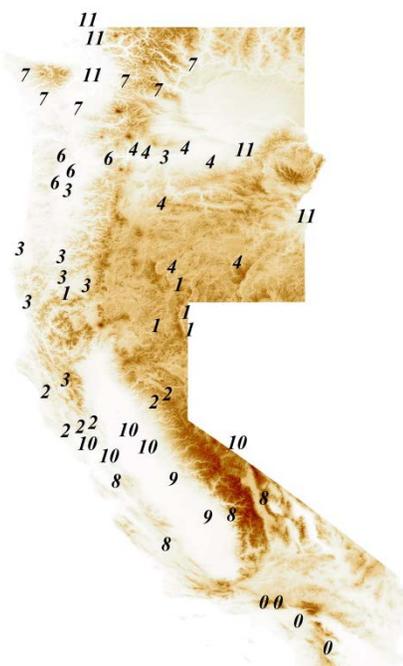
La delimitación de las regiones homogéneas en esta tesis se ha llevado a cabo mediante la técnica de Momentos-L, este procedimiento consiste básicamente en definir regiones estadísticamente homogéneas en una zona determinada. Para determinar que un grupo de estaciones conforman una región, las estaciones deben cumplir el criterio de homogeneidad.

La identificación de regiones homogéneas se realizó mediante análisis clúster c-medias y el parámetro de forma de la distribución GVE. Esta es una primera aproximación de la formación de regiones homogéneas. Esto es agrupando o encontrando similitudes en el parámetro de forma y en los estadísticos; coeficiente de variación L (L-CV) y el coeficiente de asimetría L (Skewness-L). Una vez formados los grupos entonces son medidos a través de un proceso iterativo hasta que finalmente se obtiene regiones aceptablemente homogéneas. El procedimiento que sigue ya ha sido descrito anteriormente, donde una vez teniendo una región propuesta con N sitios donde cada sitio i tiene una longitud de registro n_i y sus radios de momentos L t^i , t_3^i y t_4^i y se denotan como t^R , t_3^R y t_4^R siendo los estadísticos media regional L , coeficiente de asimetría L y a la curtosis L siendo proporcionales a los estadísticos de cada uno de los sitios de la región. Después se encontraron los radios de momentos L regionales y con la distribución kappa se simularon 500 muestras y se calculó V de las simulaciones asimismo la media y la desviación estándar μ_v y σ_v . de acuerdo a Hosking & Wallis (1997) si H es lo suficientemente grande en la literatura se sugiere $H \geq 2$ en caso contrario se tiene una región homogénea.

6.30 Resultados

De la región en estudio Costa Oeste de los Estados Unidos se escogieron 56 estaciones, cabe mencionar que para la demostración de la utilidad del modelo fue necesario tomar esta región pues se tiene la suficiente información confiable.

De las estaciones utilizadas se formaron 12 grupos de regiones homogéneas mediante la técnica de Momentos-L (Figura 4.1), los cuales son potencialmente homogéneos. En este estudio se han analizado los grupos A y B (Cuadro 1). Para el grupo A los resultados obtenidos del ajuste de la distribución Kappa a sus parámetros regionales son $t^R = 0.023$, $t_3^R = 0.647$ y $t_4^R = -1.099$ con una $V = 0.0193$ con ello se obtuvo una $H = -0.536$. Del grupo B se han obtenido del ajuste de la distribución Kappa a sus parámetros regionales los siguientes resultados $t^R = -0.027$, $t_3^R = 0.605$ y $t_4^R = -0.799$ con una $V = 0.0237$ de lo cual se obtuvo una $H = -0.538$. Las dos regiones que se han tomado para el análisis regional han demostrado ser homogéneas de acuerdo con Hosking y Wallis (1997) pues H es menor que 2.



Grupo	Estación	t^i	t_3^i	t_3^i
A	0 USC00042239	0.012	-0.04	-0.93
	0 USC00042805	-0.003	0.10	-0.07
	0 USC00043452	0.042	0.286	0.02
	0 USC00046719	0.041	-0.73	-0.42
	7 USC00450008	-0.058	-0.29	0.14
B	7 USC00451233	-0.058	-0.29	0.14
	7 USC00451276	0.002	0.72	-0.91
	7 USC00451350	-0.024	-0.80	0.17
	7 USC00451496	-0.026	0.94	-0.75
	7 USC00451504	-0.002	-0.65	0.41

Figura 6.47. Formación de grupos homogéneos la Costa Oeste de EEUU.

6.31 Análisis de series mensuales

En el análisis de escurrimientos se utilizó un modelo ensamblado mediante el enfoque regional propuesto obteniendo resultados exitosos, puesto que el modelo ensamblado propuesto siempre fue superior al modelo tradicional. Para este caso de estudio el mismo modelo ensamblado mediante el enfoque regional propuesto se comparará con el modelo tradicional ensamblado en el pronóstico de lluvia. En la tabla 6.25 se muestran los resultados obtenidos en la comparación del modelo Ensamblado propuesto y el modelo tradicional en el pronóstico de la serie de lluvia acumulada mensual, esto es para la región A de la Costa Oeste de Estados Unidos. Los resultados indican que el modelo propuesto es superior al modelo tradicional, ya que en todas las estaciones se obtuvieron de acuerdo a las pruebas de bondad de ajuste siempre los menores errores.

También se ha observado que el pronóstico no ha sido del todo eficiente en algunos casos, tal es el caso de la estación USC000042239 y en los casos en donde se tienen valores atípicos, ya que es complicado aproximarse a dichos eventos. Esto se observa en la mayoría de las estaciones que corresponden a la variable lluvia.

Tabla 6.25 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A

Estaciones	Ensamblado regional				Ensamblado tradicional			
	MSE	RMSE	MAE	d	MSE	RMSE	MAE	d
USC00042239	4510683	2123.84	1019.64	0.0280	9180088	3029.87	1465.96	-0.1023
USC00042805	469352	685.09	330.53	0.2464	641934	801.21	410.56	0.1336
USC00043452	853621	923.92	521.32	0.3850	2641548	1625.28	873.15	0.0455
USC00046719	625765	791.05	444.62	0.3686	19156365	4376.80	1653.99	0.0287

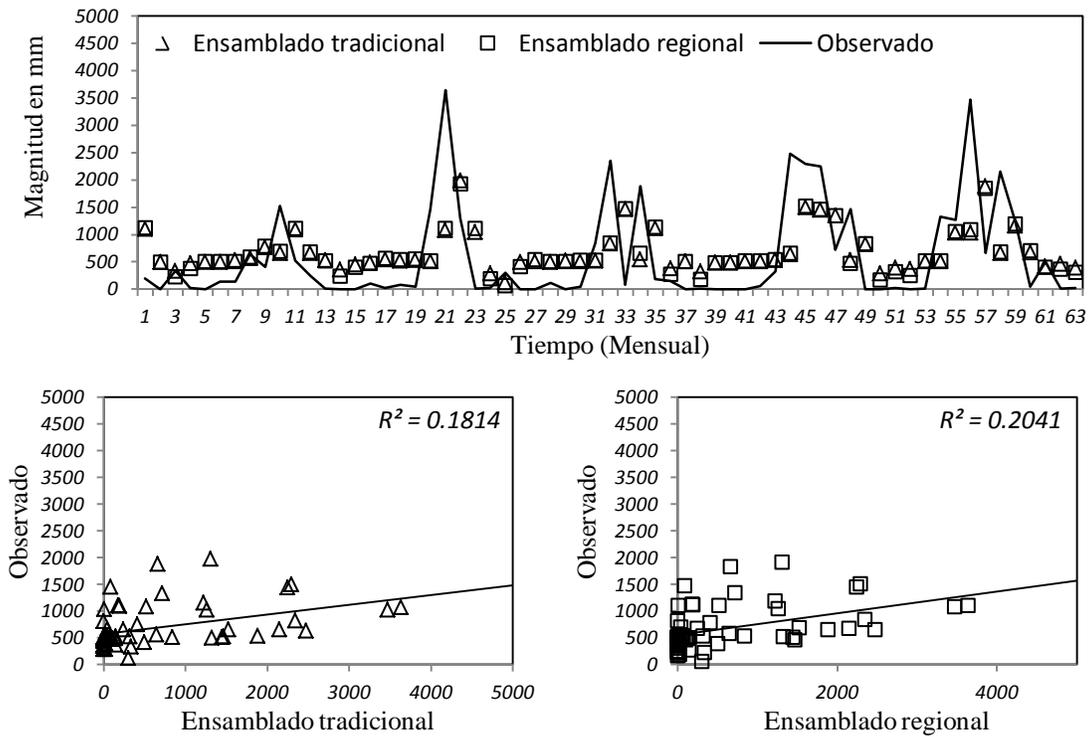


Figura 6.48. Pronostico mensual de lluvia acumulada en la estación USC00042239

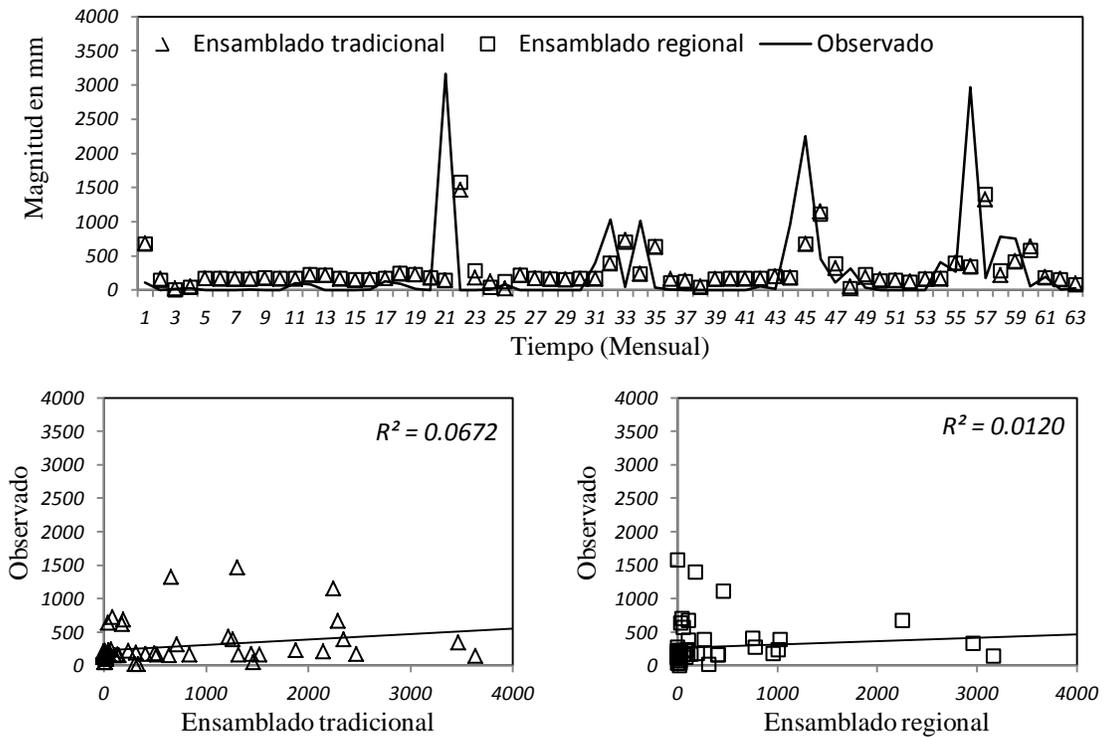


Figura 6.49. Pronostico mensual de lluvia acumulada en la estación USC00042805

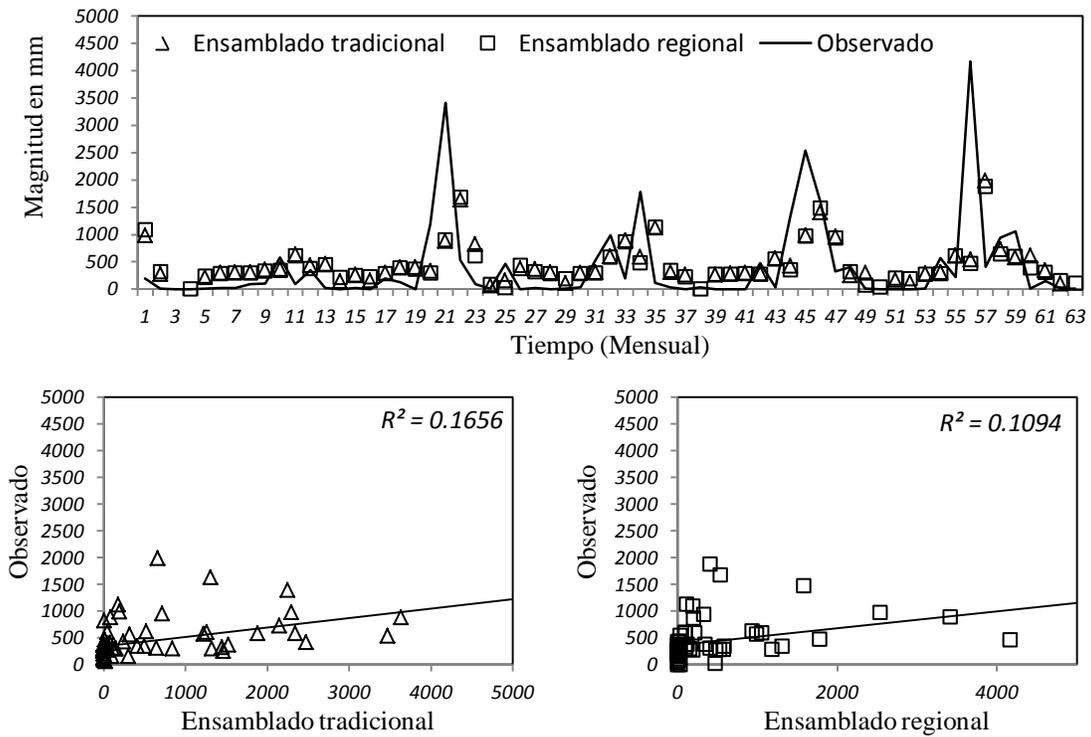


Figura 6.50. Pronostico mensual de lluvia acumulada en la estación USC00043452

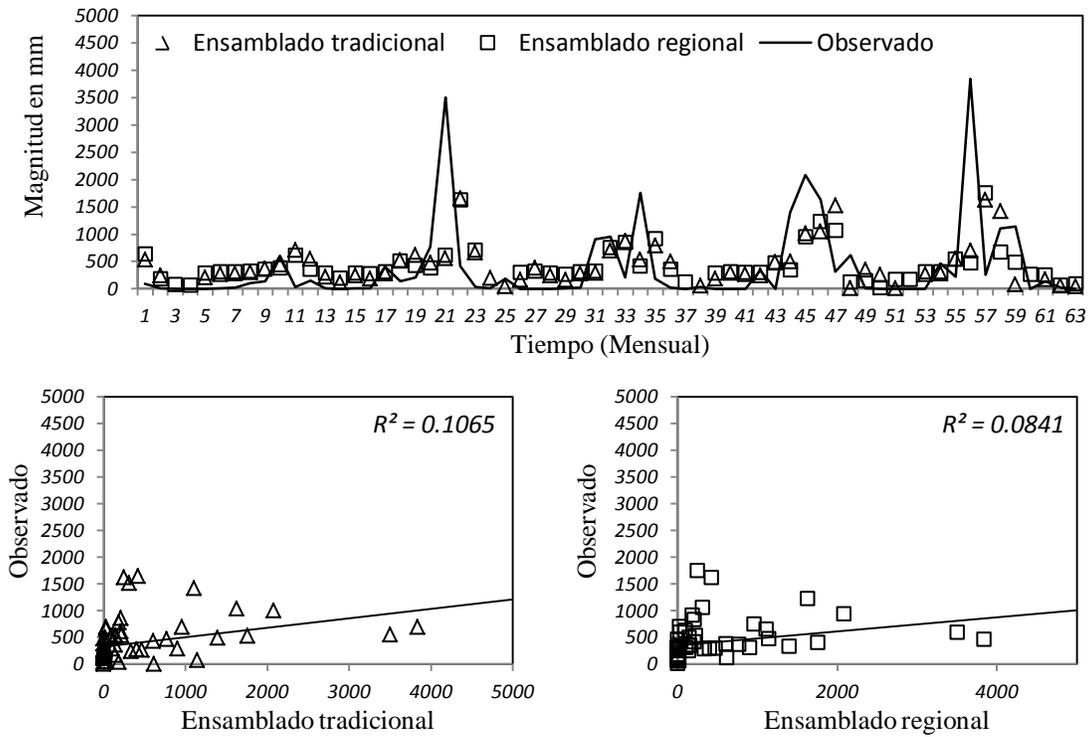


Figura 6.51. Pronostico mensual de lluvia acumulada en la estación USC00046719

En la tabla 6.26 se presentan los resultados obtenidos de la comparación del modelo Ensamblado propuesto y el modelo tradicional aplicado al pronóstico de la serie de acumulados mensual de la región B en la Costa Oeste de Estados Unidos. Los resultados demuestran que en más de un 60% el modelo ensamblado fue superior al modelo tradicional.

Tabla 6.26 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B

Estaciones	Ensamblado regional				Ensamblado tradicional			
	MSE	RMSE	MAE	d	MSE	RMSE	MAE	d
USC00450008	1053800	1026.55	651.60	0.6976	1107749	1052.50	786.59	0.6601
USC00451233	1601897	1265.66	885.25	0.4421	15537119	3941.72	1983.56	0.0503
USC00451276	239707	489.60	398.42	0.5735	342028	584.83	474.68	0.3598
USC00451350	188320	433.96	310.71	0.0114	276775	526.09	356.21	-0.3461
USC00451496	3867032	1966.48	1519.39	0.2779	10979057	3313.47	2083.70	-0.0711
USC00451504	95436	308.93	270.27	0.1798	81319	285.17	252.14	0.3374

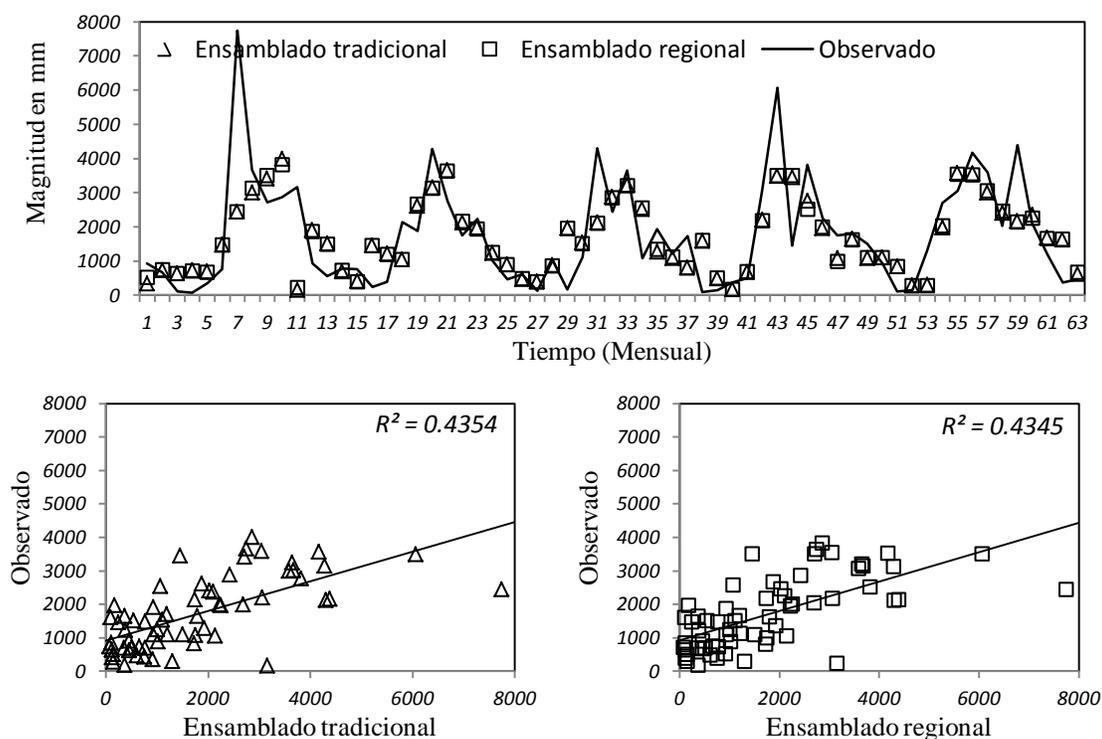


Figura 6.52. Pronostico mensual de lluvia acumulada en la estación USC00450008

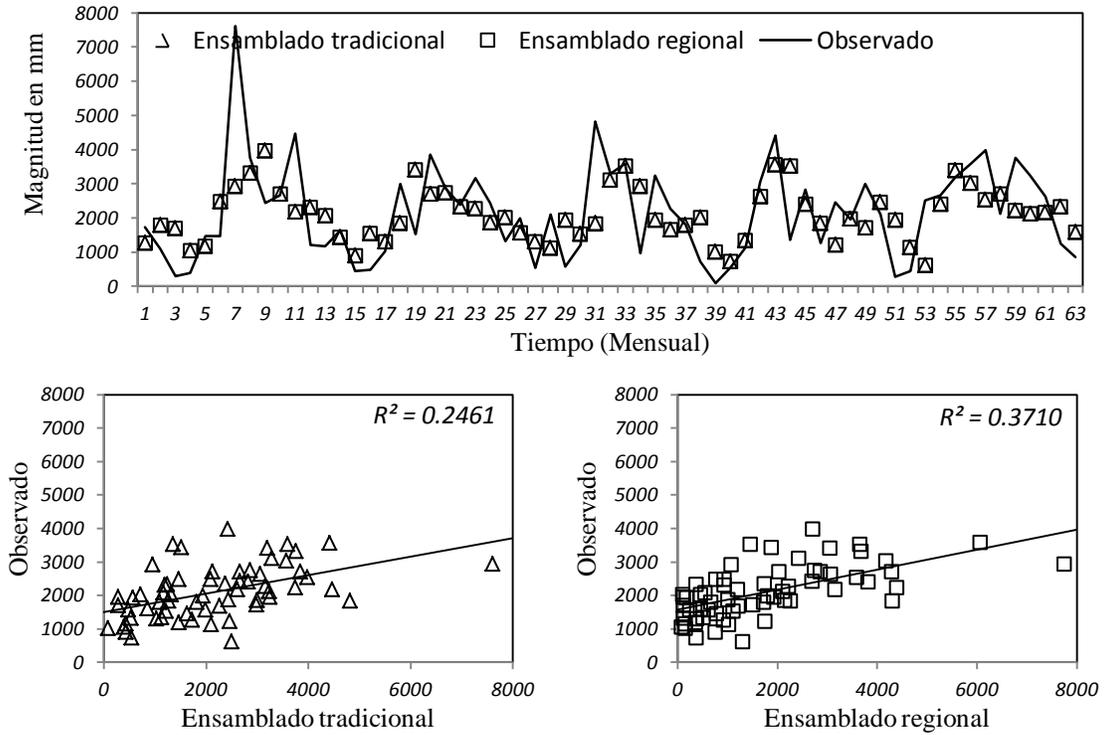


Figura 6.53. Pronostico mensual de lluvia acumulada en la estación USC00451233

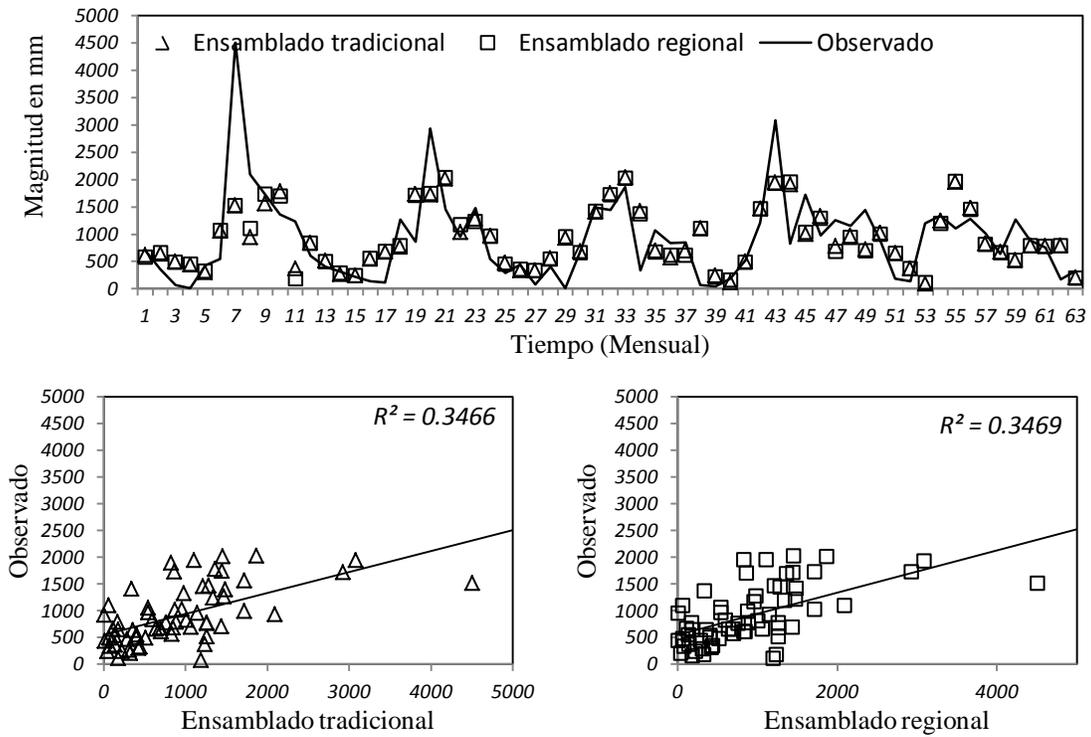


Figura 6.54. Pronostico mensual de lluvia acumulada en la estación USC00451276

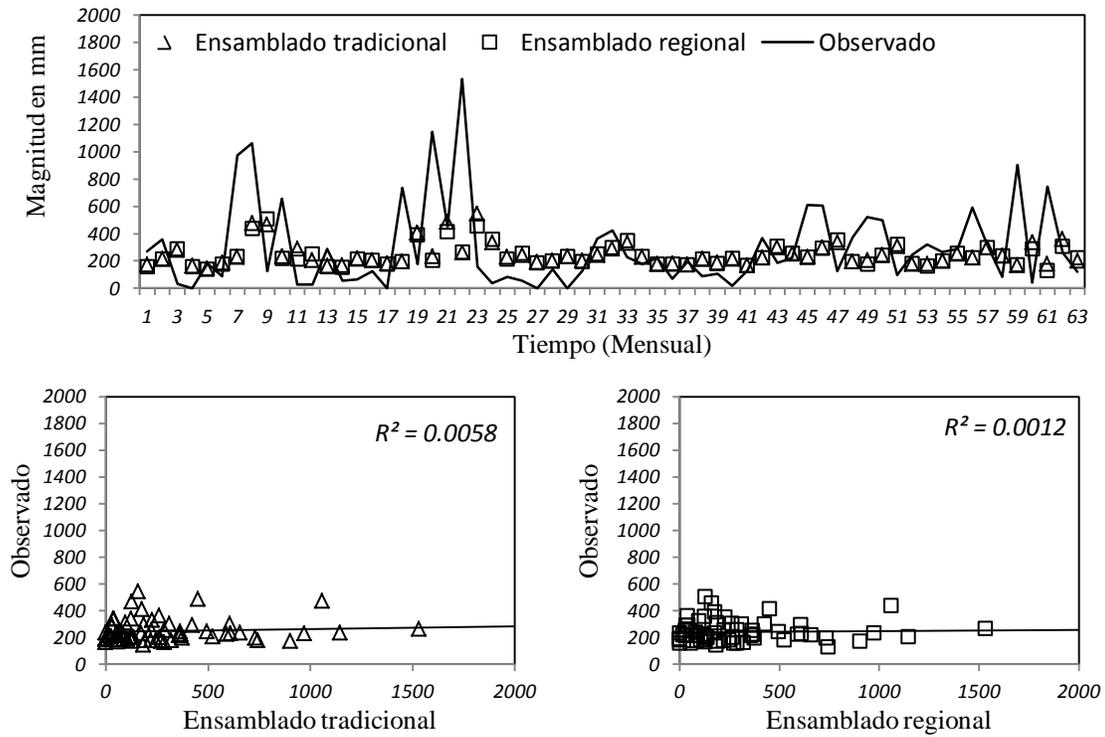


Figura 6.55. Pronostico mensual de lluvia acumulada en la estación USC00451350

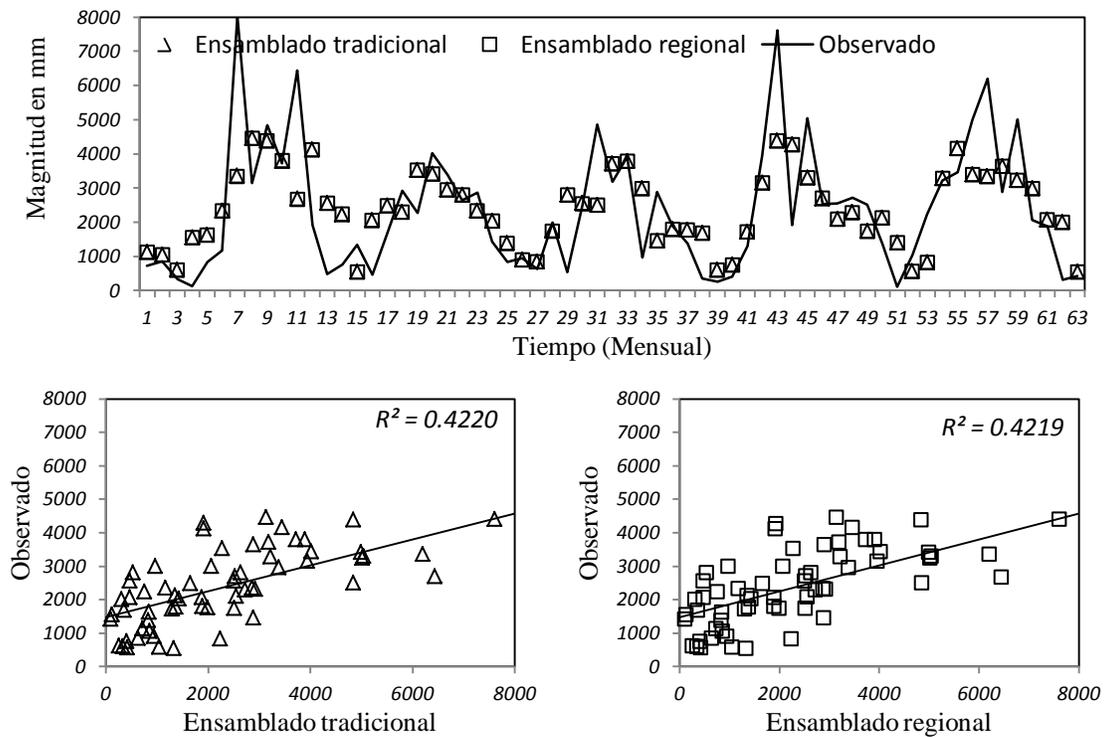


Figura 6.56. Pronostico mensual de lluvia acumulada en la estación USC00451496

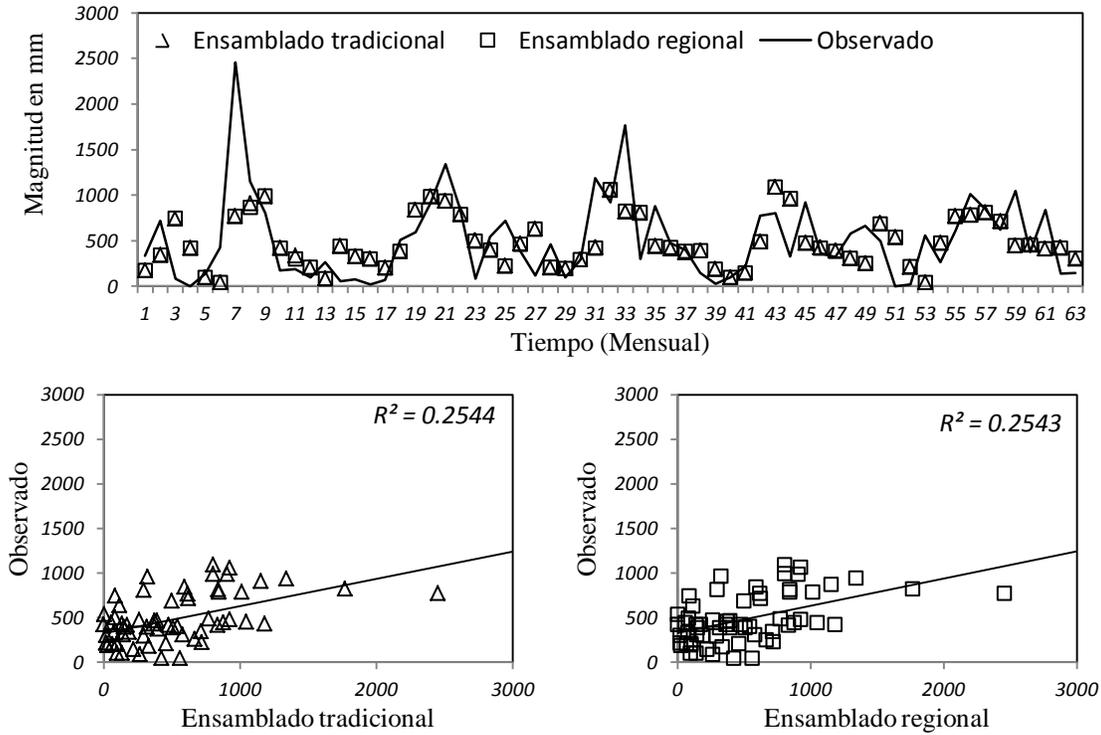


Figura 6.57. Pronostico mensual de lluvia acumulada en la estación USC00451504

6.32 Análisis de series trimestrales

En la tabla 6.27 se muestran los resultados obtenidos en el modelo ensamblado propuesto y el modelo tradicional en la serie trimestral, esto es para la región A. los resultados indican que en general el modelo ensamblado es superior al modelo tradicional.

Tabla 6.27 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A

Estaciones	Ensamblado regional				Ensamblado tradicional			
	MSE	RMSE	MAE	<i>d</i>	MSE	RMSE	MAE	<i>d</i>
USC00042239	4067408	2016.78	1281.08	0.3969	12068114	3473.92	2137.31	0.2752
USC00042805	2516728	1586.42	994.53	0.1932	1731175	1315.74	889.01	0.0563
USC00043452	7063591	2657.74	1495.21	0.1374	9967823	3157.19	1514.53	0.0836
USC00046719	4506286	2122.80	1438.23	0.3945	19529191	4419.18	1791.29	0.1159

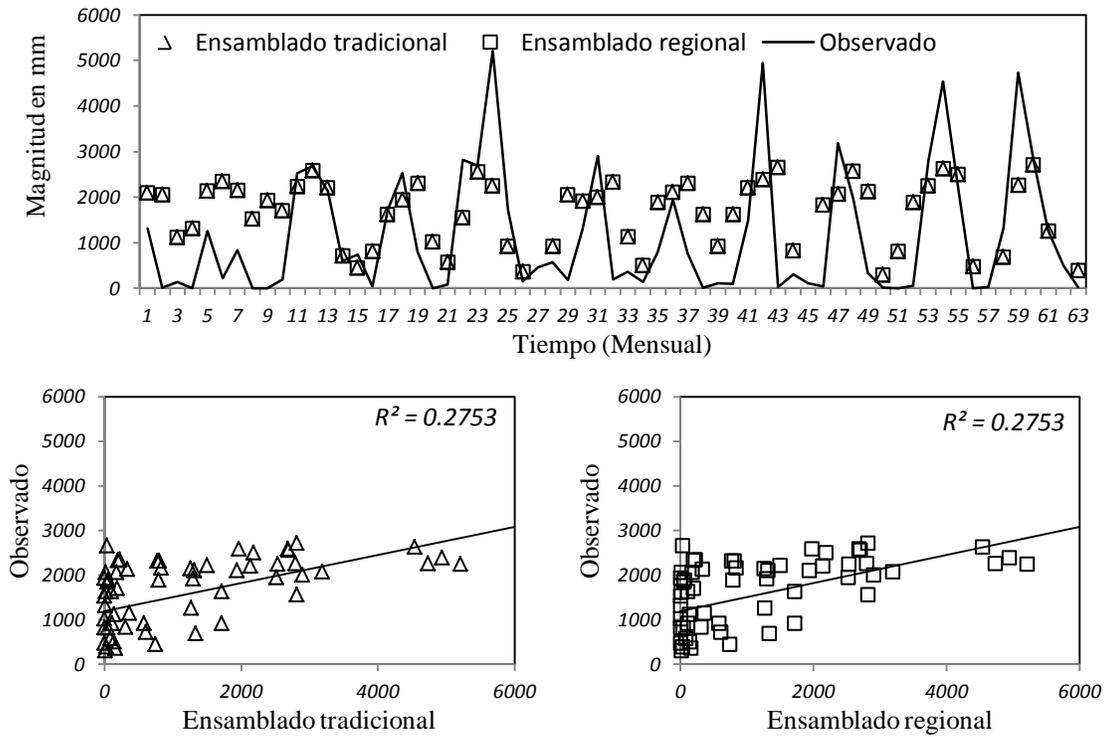


Figura 6.58. Pronostico trimestral de lluvia acumulada en la estación USC00042239

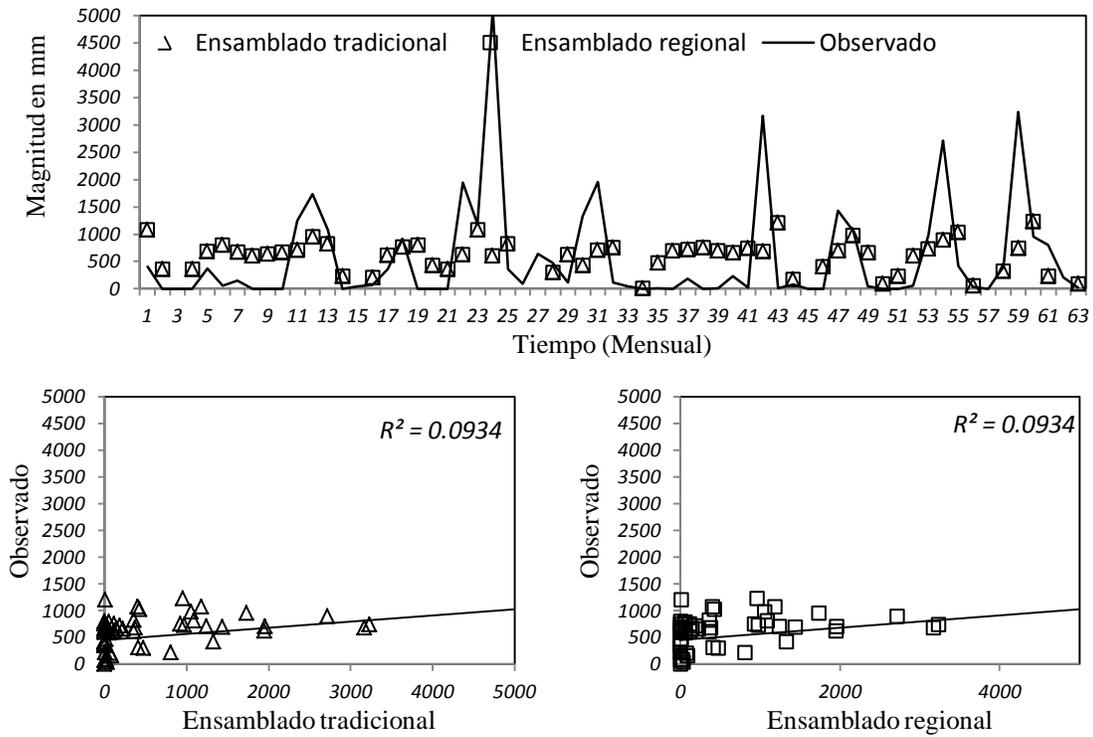


Figura 6.59. Pronostico trimestral de lluvia acumulada en la estación USC00042805

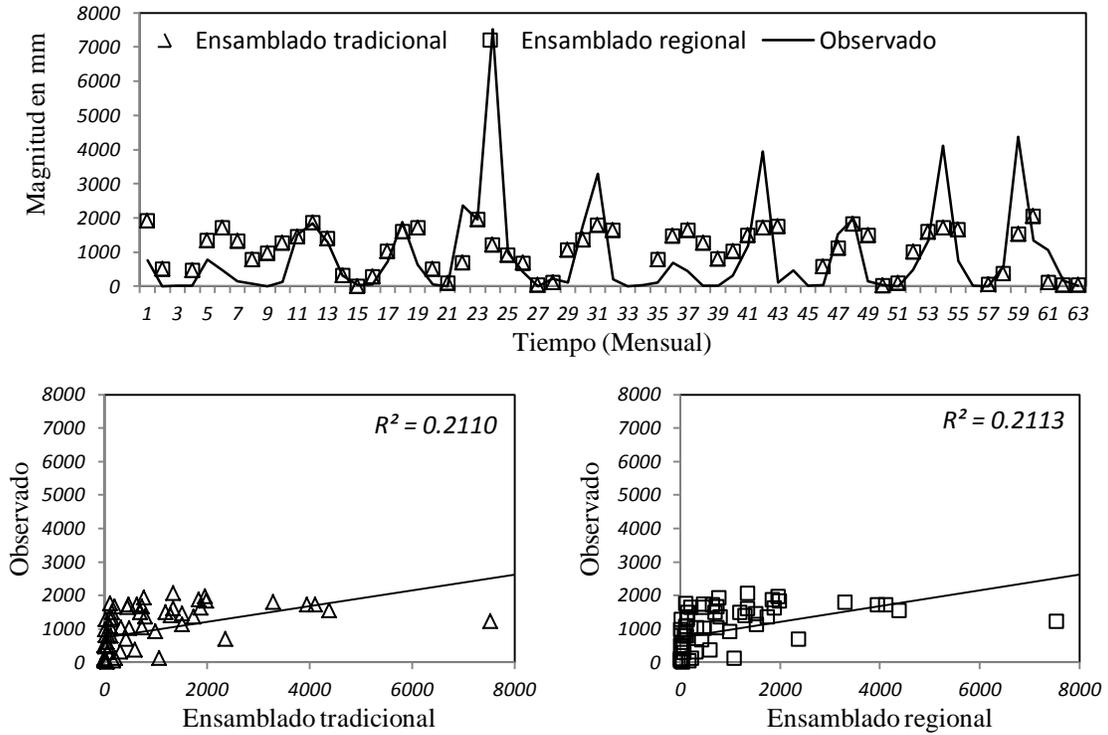


Figura 6.60. Pronostico trimestral de lluvia acumulada en la estación USC00043452

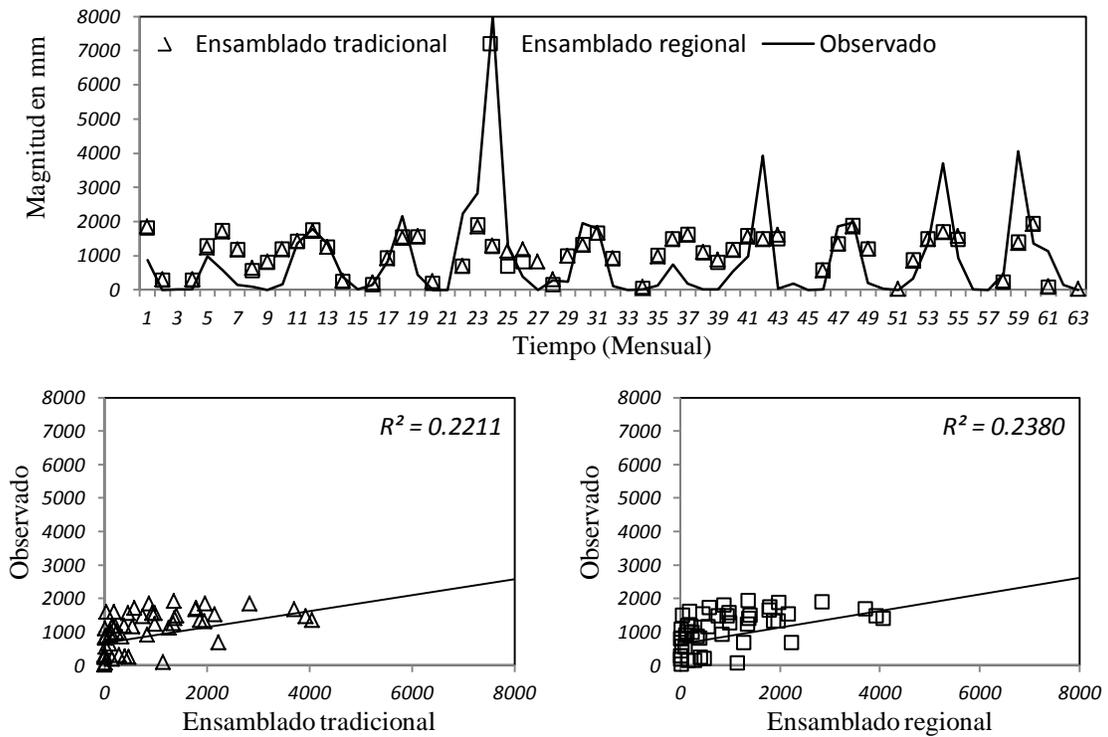


Figura 6.61. Pronostico trimestral de lluvia acumulada en la estación USC00046719

En la tabla 6.28 se observó que en más de un 80% los resultados demuestran que el modelo ensamblado propuesto ha sido superior en la aplicación a la serie trimestral, esto es para la región B. se concluye que en general el modelo ensamblado es superior al modelo tradicional ensamblado.

Tabla 6.28 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B

Estaciones	Modelo (ARNN-ER)				Tradicional (ARNN-E)			
	MSE	RMSE	MAE	<i>d</i>	MSE	RMSE	MAE	<i>d</i>
USC00450008	12563951	3544.57	2323.83	0.6894	83600774	9143.35	6159.72	0.3014
USC00451233	15816079	3976.94	2924.77	0.4190	42706316	6535.01	3507.40	0.2267
USC00451276	5480548	2341.06	1639.20	0.4731	64779663	8048.58	3890.04	0.0205
USC00451350	4294267	2072.26	1060.18	-0.0724	801761	895.41	672.85	-0.0229
USC00451496	9478824	3078.77	2334.63	0.7204	33445336	5783.19	3854.87	0.4394
USC00451504	6567782	2562.77	1424.70	-0.1776	7672813	2769.98	1661.94	0.1188

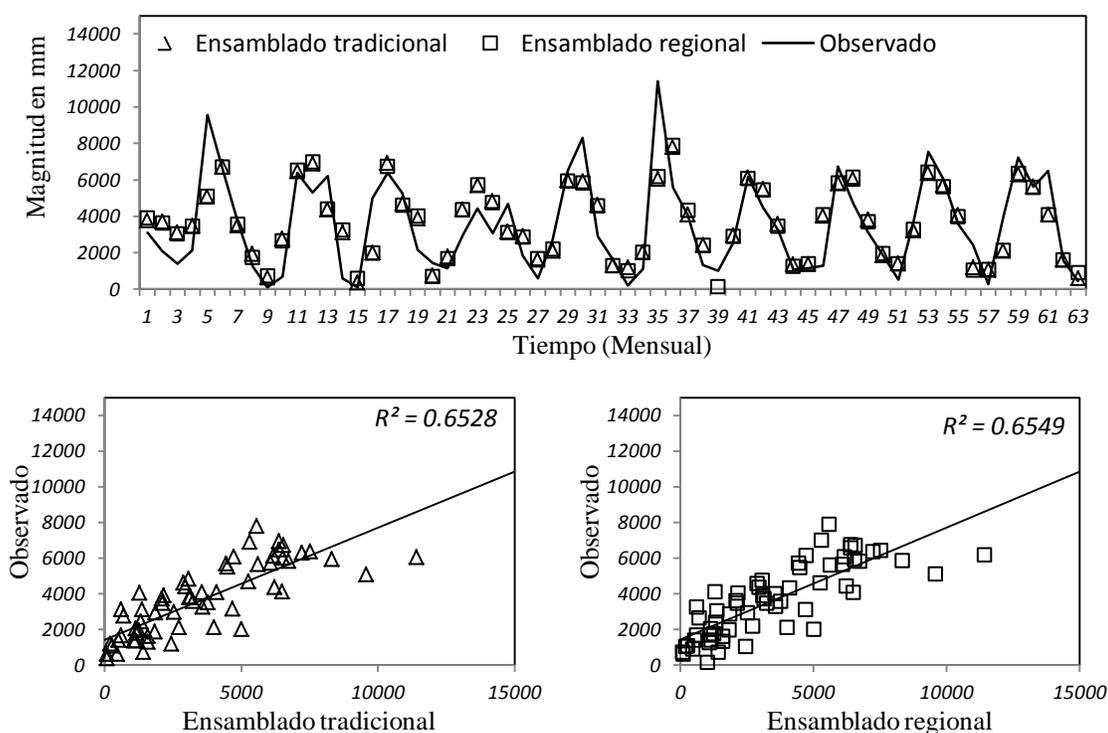


Figura 6.62. Pronostico trimestral de lluvia acumulada en la estación USC00450008

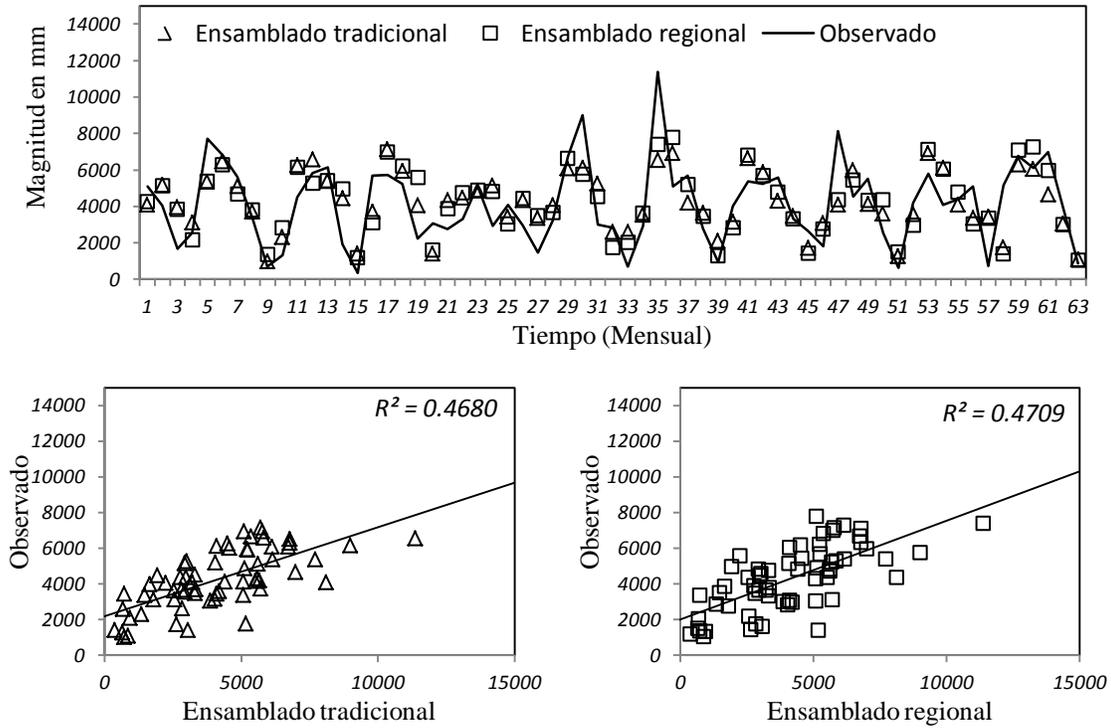


Figura 6.63. Pronostico trimestral de lluvia acumulada en la estación USC00451233

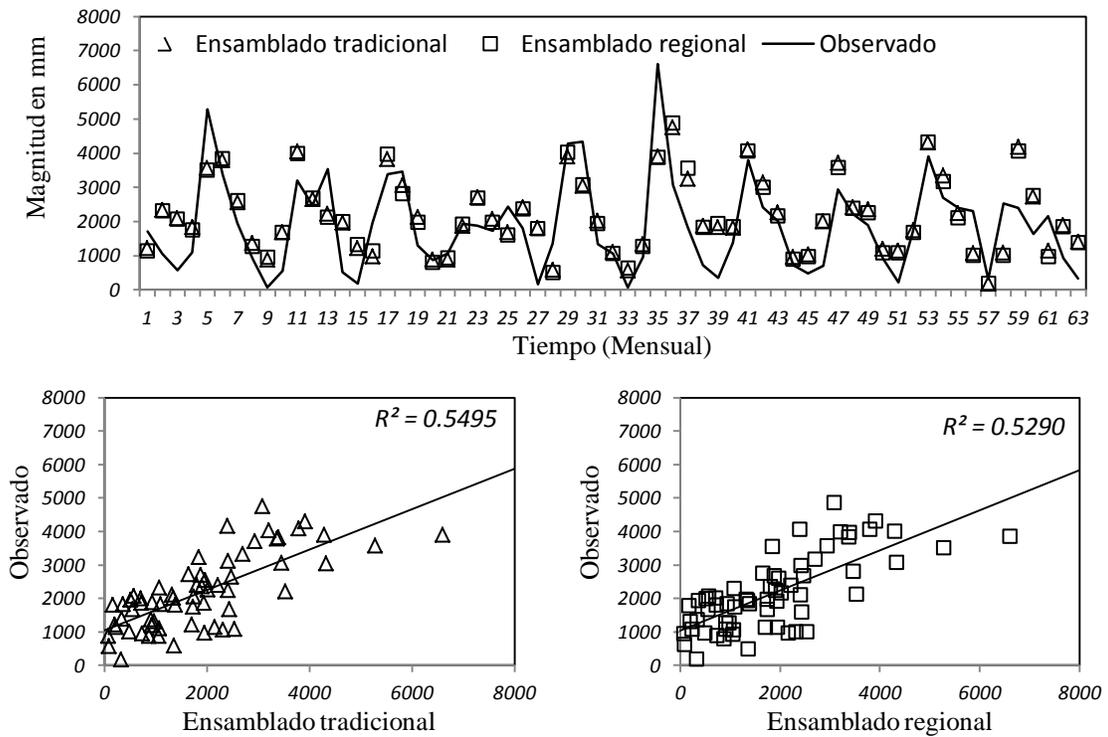


Figura 6.64. Pronostico trimestral de lluvia acumulada en la estación USC00451276

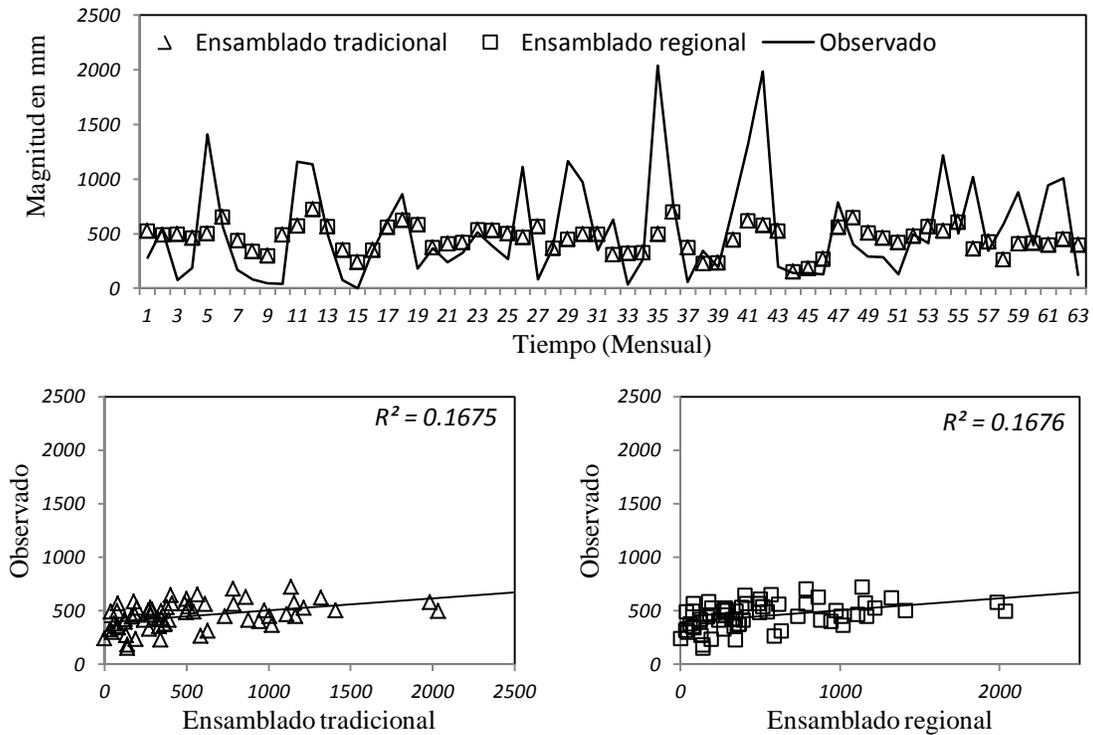


Figura 6.65. Pronostico trimestral de lluvia acumulada en la estación USC00451350

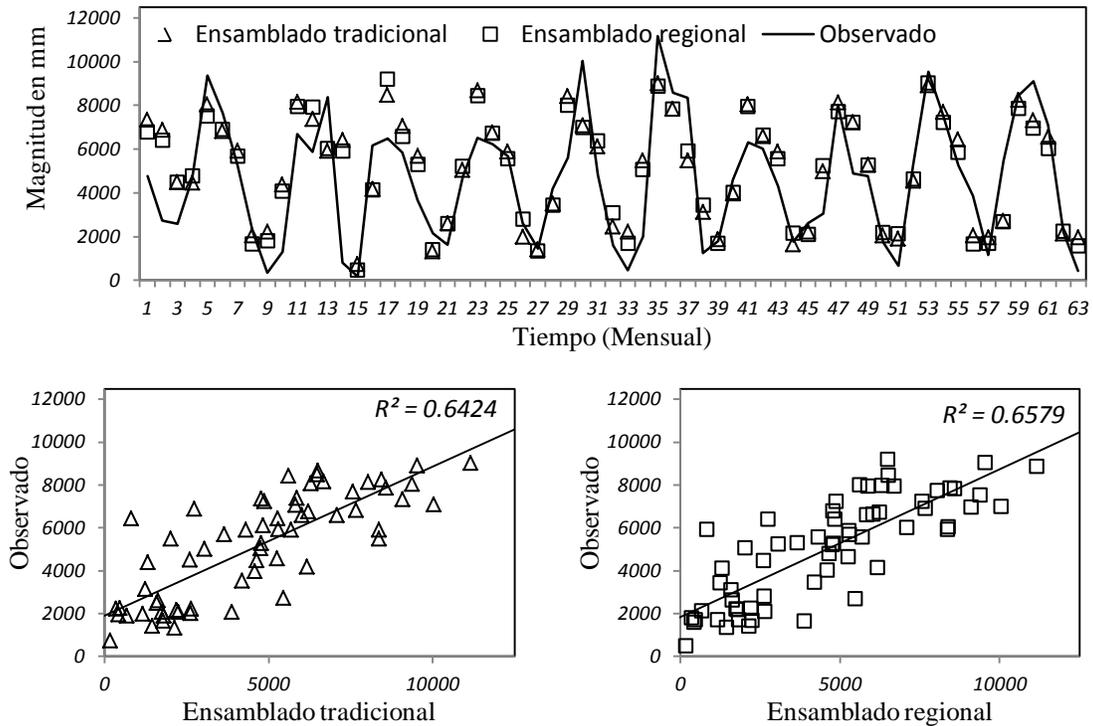


Figura 6.66. Pronostico trimestral de lluvia acumulada en la estación USC00451496

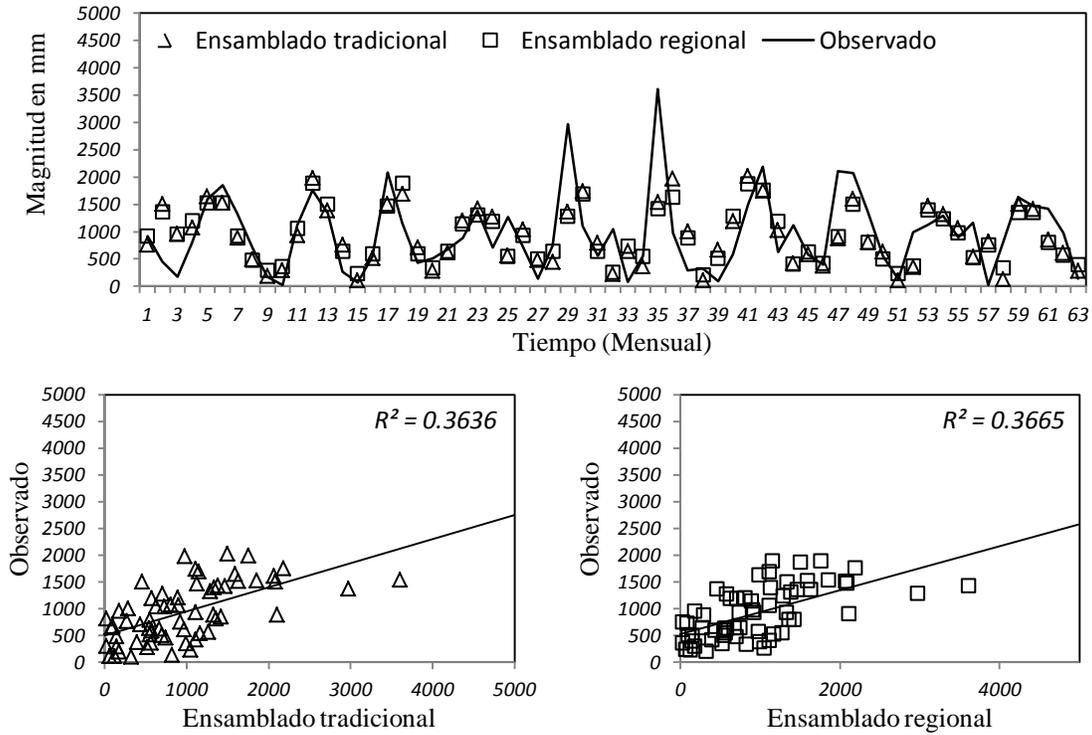


Figura 6.67. Pronostico trimestral de lluvia acumulada en la estación USC00451504

6.33 Análisis de series semestrales

En las tablas 6.29 y 6.30 se muestran los resultados obtenidos de la comparación del modelo Ensamblado regional y el modelo Ensamblado, esto es para la región A y B de la serie de lluvia acumulada de seis meses en la región Costa Oeste de Estados Unidos. Los resultados indican que en general el modelo ensamblado mediante un enfoque regional ha tenido un mejor desempeño en el pronóstico de la lluvia, pues de acuerdo a las pruebas de bondad de ajuste fue superior en un 85% de todas las estaciones analizadas.

Tabla 6.29 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A

Estaciones	Ensamblado regional				Ensamblado tradicional			
	MSE	RMSE	MAE	<i>d</i>	MSE	RMSE	MAE	<i>d</i>
USC00042239	8217847	2866.68	2239.08	0.0155	68167284	8256.35	3956.97	-0.0734
USC00042805	3298634	1816.21	1350.74	-0.1427	23173039	4813.84	3229.63	0.1842
USC00043452	56421913	7511.45	3672.45	-0.0789	135946666	11659.62	6466.95	-0.1084
USC00046719	10150100	3185.92	2328.92	-0.2128	214260063	14637.62	8463.68	0.0952

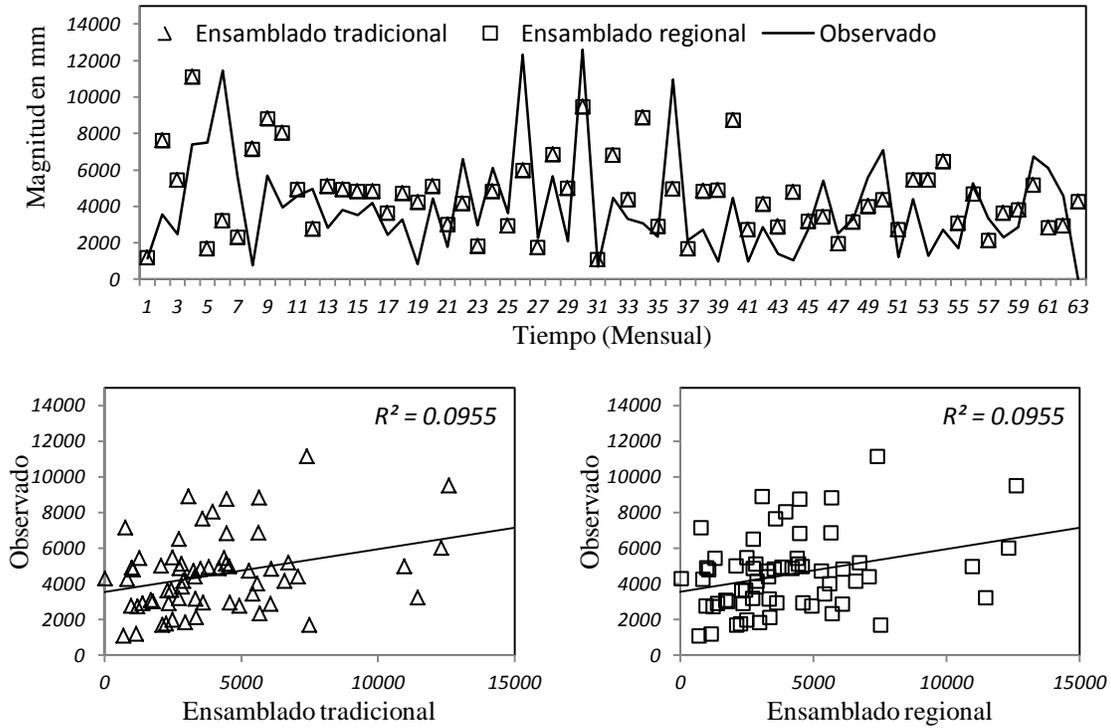


Figura 6.68. Pronostico semestral de lluvia acumulada en la estación USC00042239

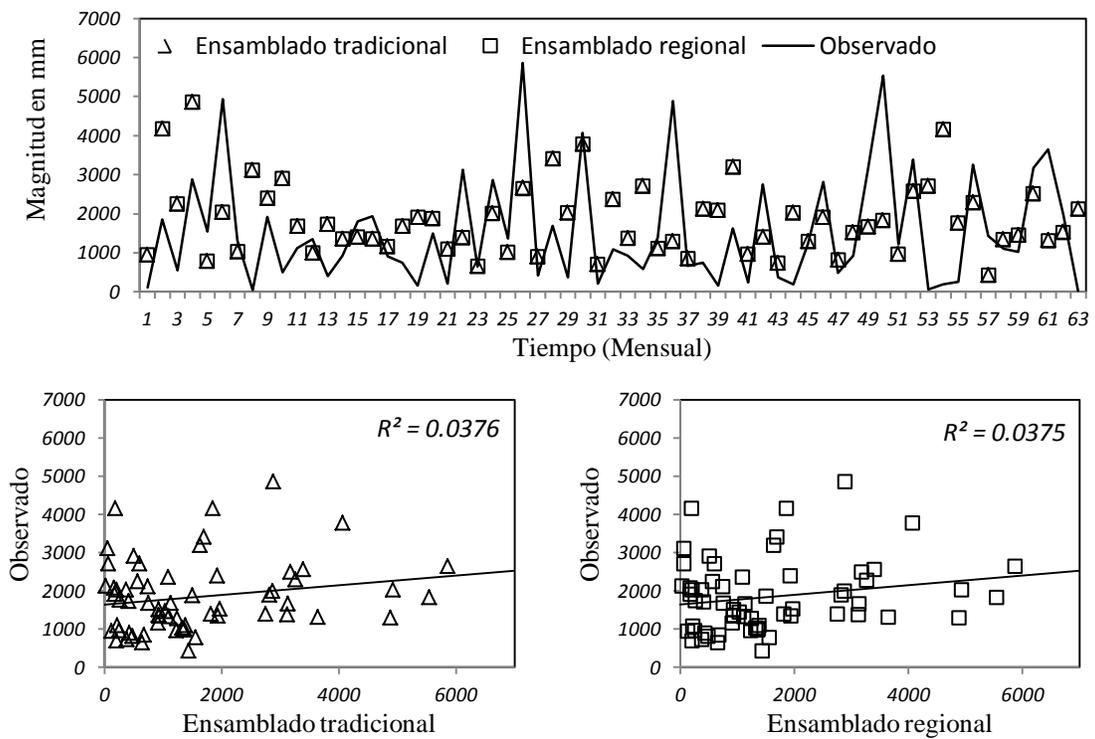


Figura 6.69. Pronostico semestral de lluvia acumulada en la estación USC00042805

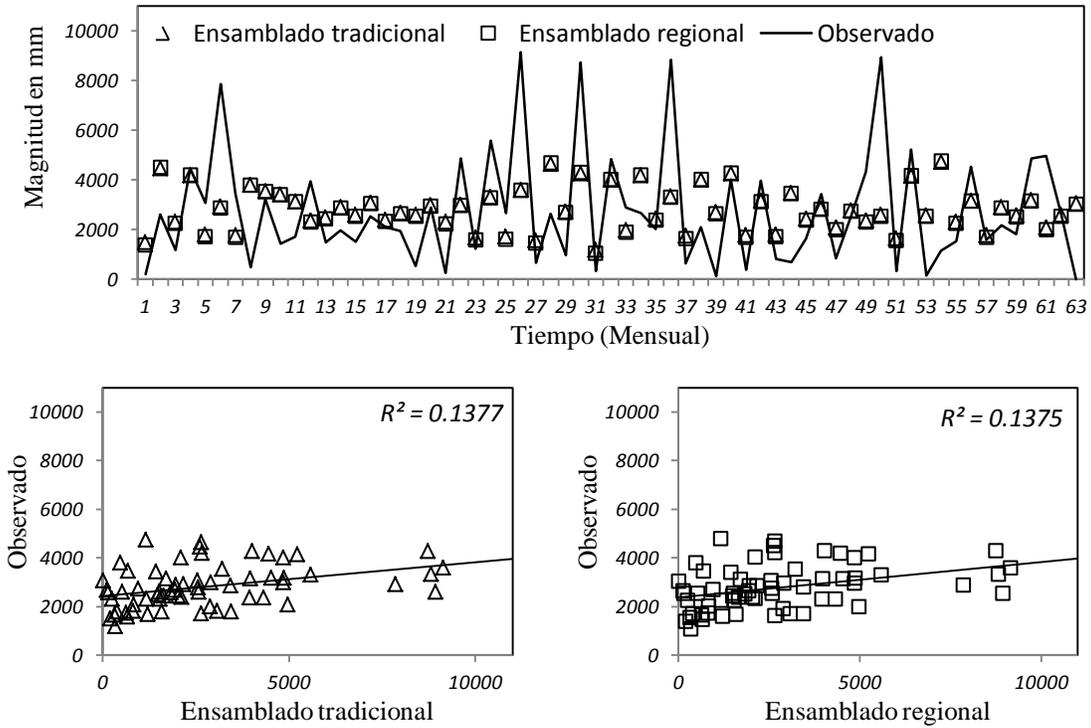


Figura 6.70. Pronostico semestral de lluvia acumulada en la estación USC00043452

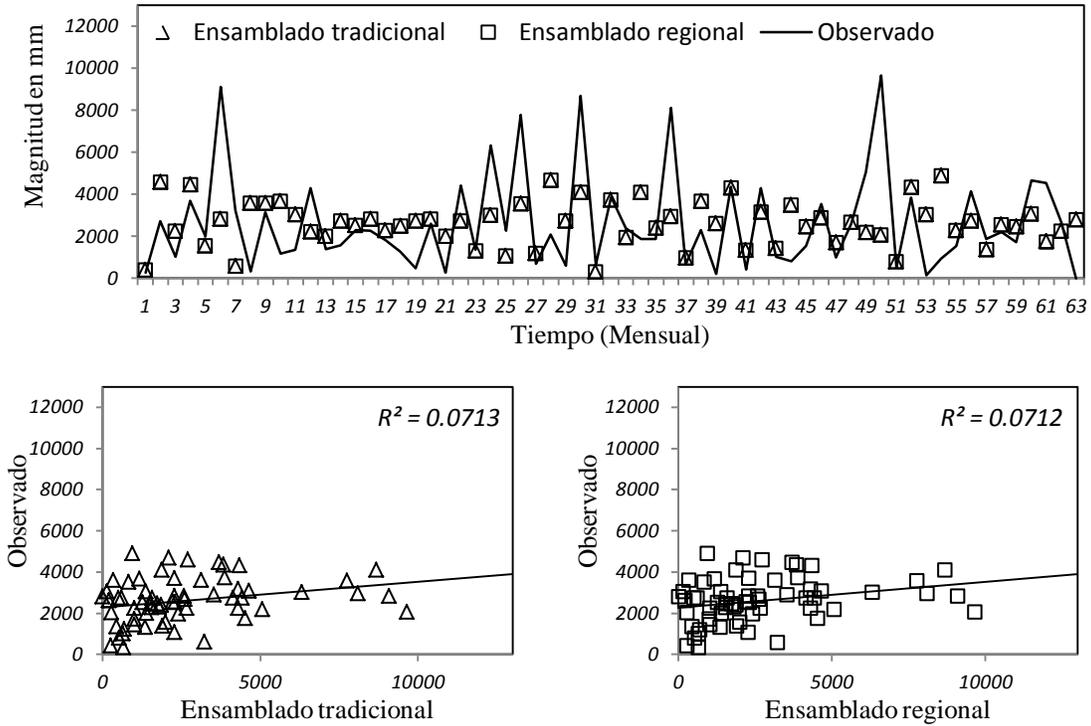


Figura 6.71. Pronostico semestral de lluvia acumulada en la estación USC00049719

Tabla 6.30 Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B

Estaciones	Ensamblado regional				Ensamblado tradicional			
	MSE	RMSE	MAE	d	MSE	RMSE	MAE	d
USC00450008	27866810	5278.90	3709.10	0.6789	150000226	12247.46	6770.49	0.1463
USC00451233	108082187	10396.26	7271.63	0.4053	36781130	6064.74	4160.86	0.5467
USC00451276	10015820	3164.78	2352.06	0.5913	30363490	5510.31	3852.07	0.3840
USC00451350	1348493	1161.25	927.70	0.2588	1640738	1280.91	996.12	-0.1313
USC00451496	27891617	5281.25	3788.00	0.7524	230733318	15189.91	9175.97	0.0116
USC00451504	10100218	3178.08	2255.94	0.3304	28251774	5315.24	2964.53	0.1007

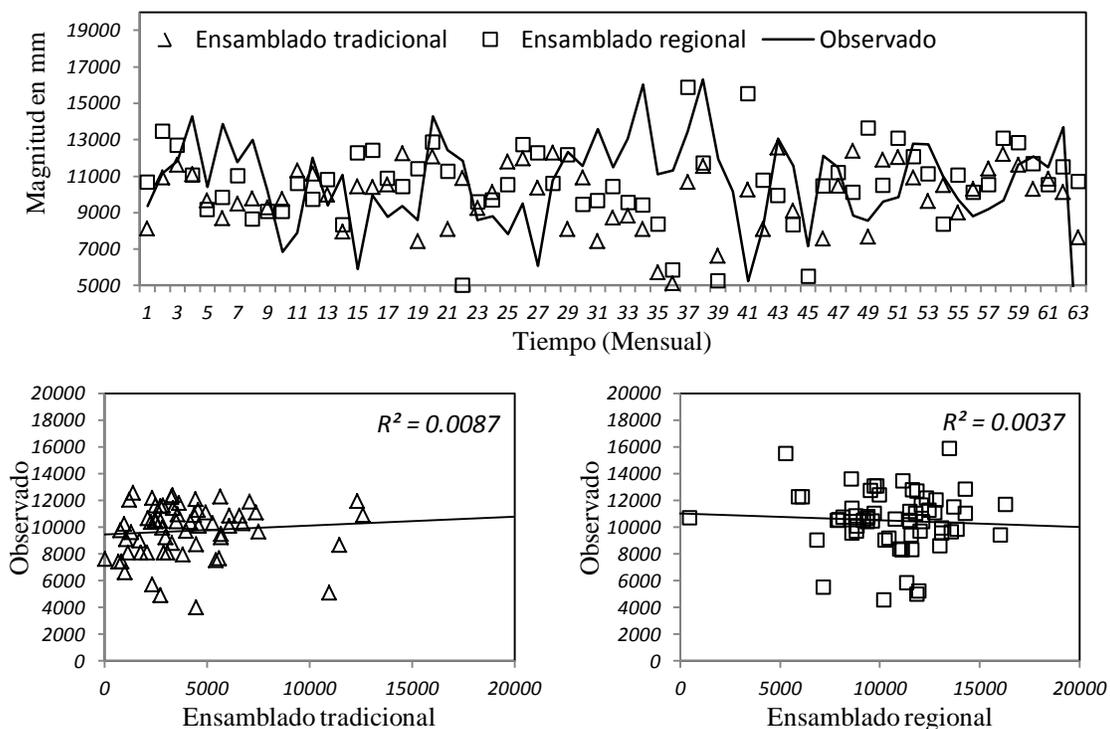


Figura 6.72. Pronostico semestral de lluvia acumulada en la estación USC00450008

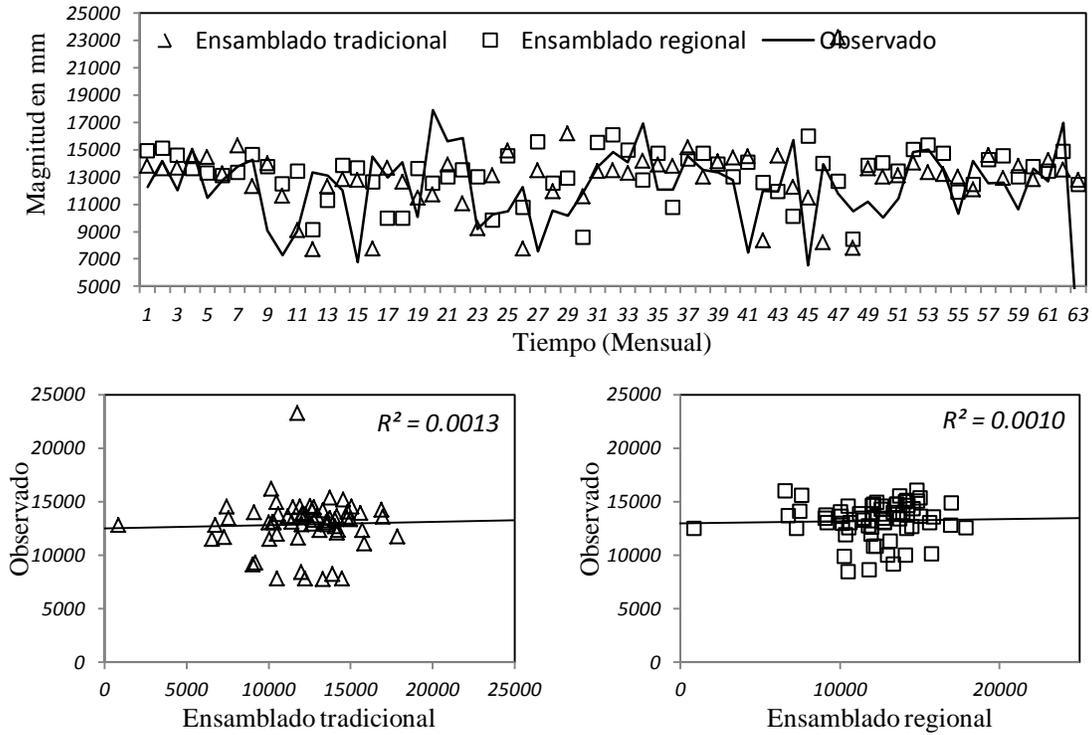


Figura 6.73. Pronostico semestral de lluvia acumulada en la estación USC00451233

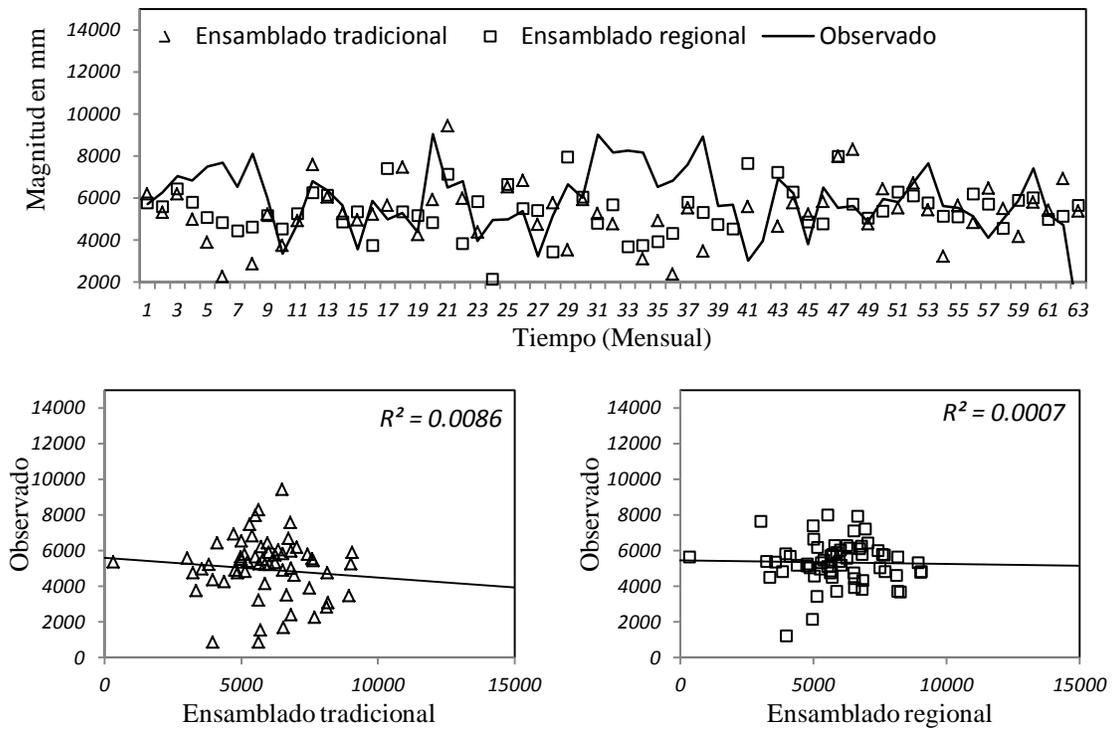


Figura 6.74. Pronostico semestral de lluvia acumulada en la estación USC00451276

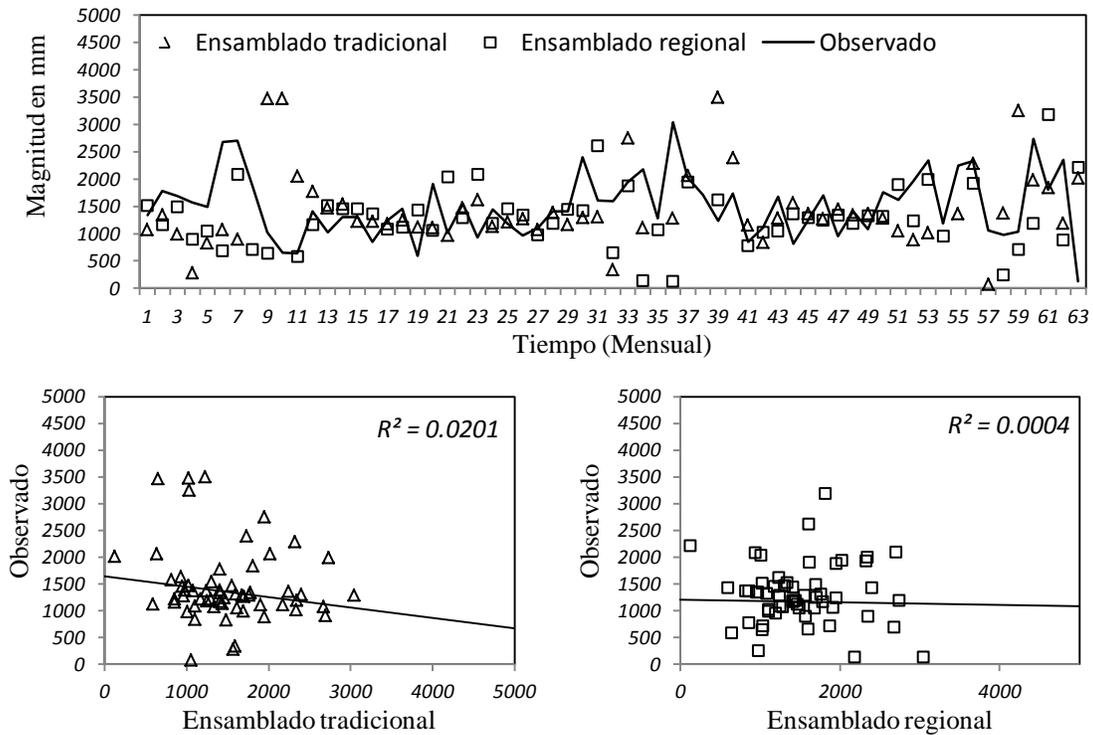


Figura 6.75. Pronostico semestral de lluvia acumulada en la estación USC00451350

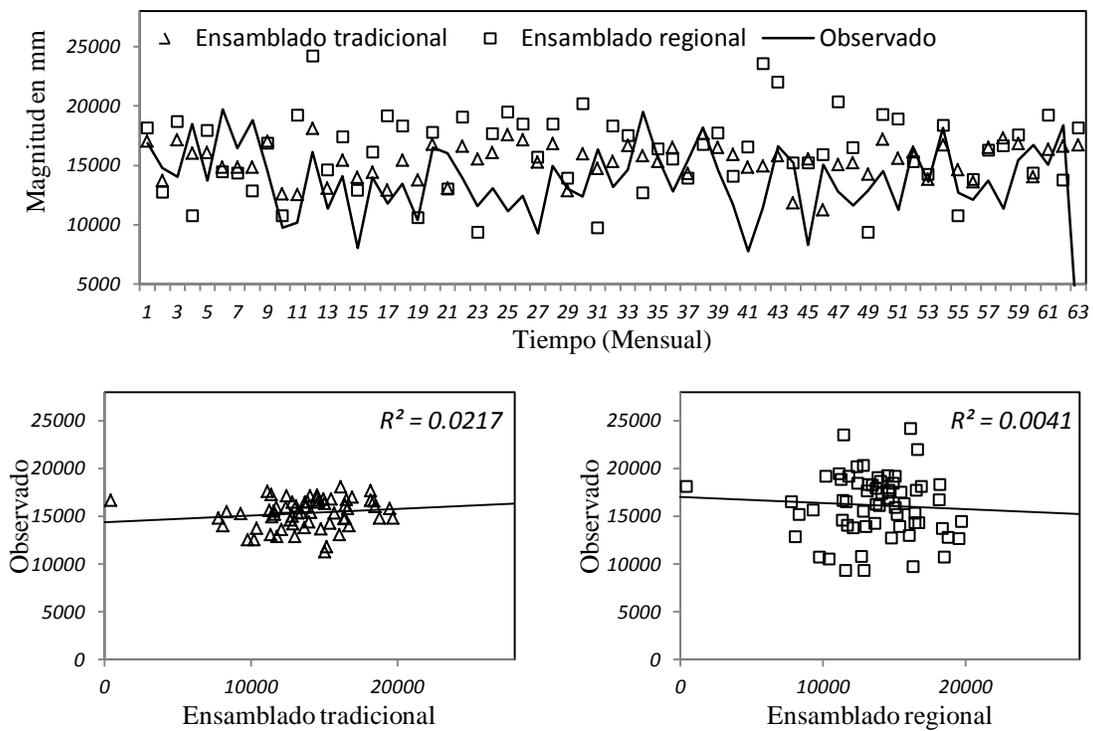


Figura 6.76. Pronostico semestral de lluvia acumulada en la estación USC00451496

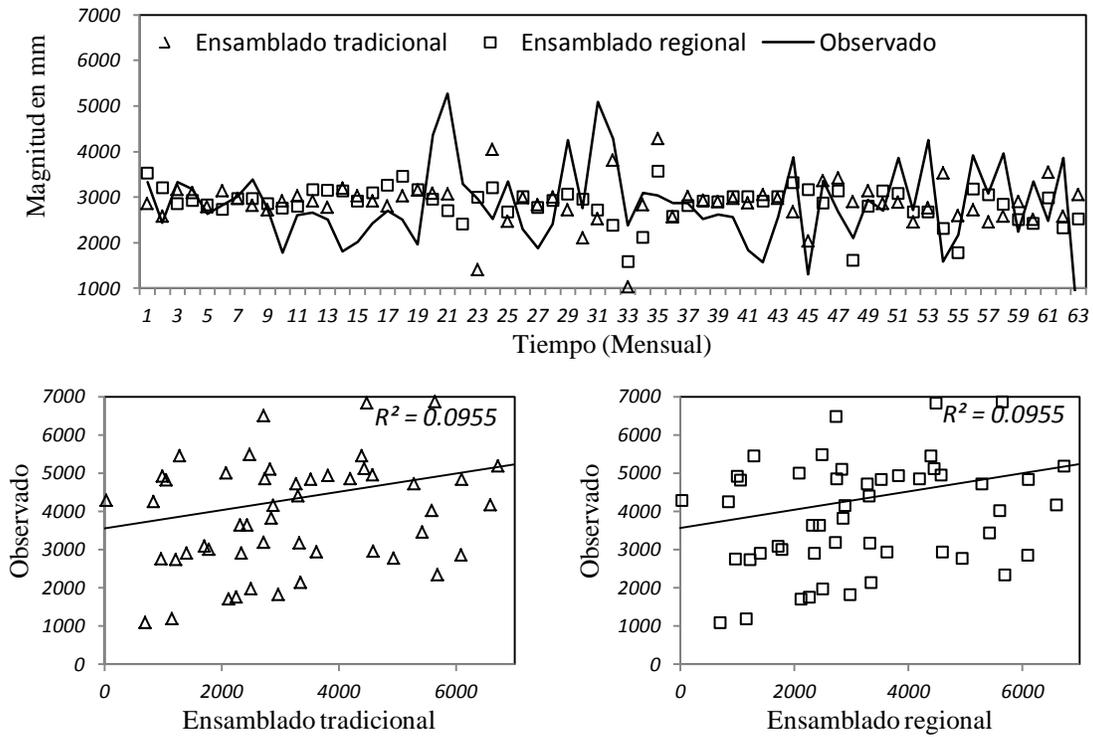


Figura 6.77. Pronostico semestral de lluvia acumulada en la estación USC00451504

6.34 Comparación de lluvias en las escalas mensual trimestral y semestral en Zacatecas.

En esta última parte de la tesis se realizó la comparación de los modelos regionales en la variable lluvia en la escala de tiempo mensual. Los modelos regionales comparados fueron RNAs, Ensamblado y Modular. Las series de lluvia fueron tomadas de las estaciones climatológicas del estado de Zacatecas (Figura 6.78).

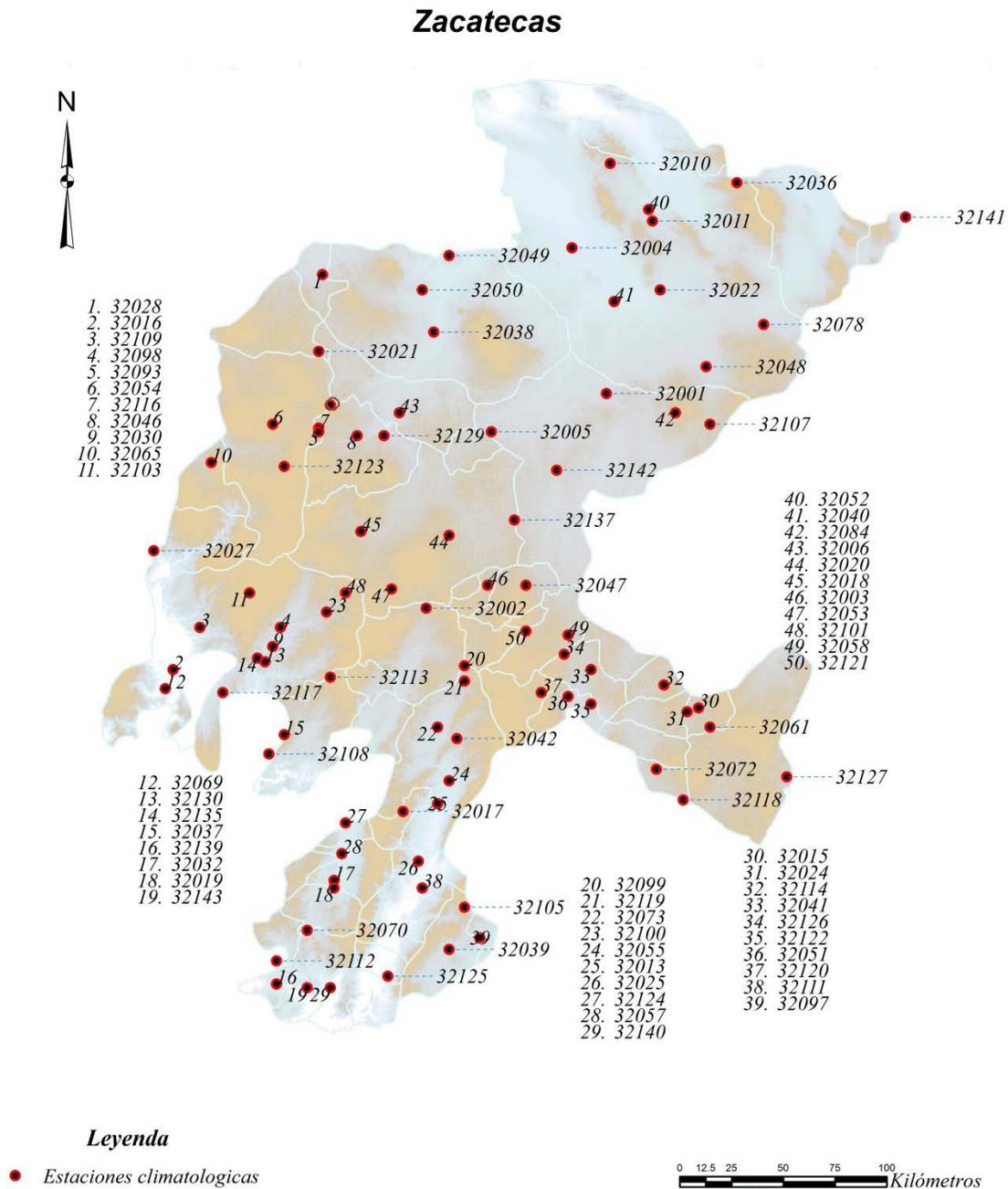


Figura 6.78. Estaciones climatológicas en Zacatecas.

En el estado de Zacatecas se cuenta con 87 estaciones climatológicas de acuerdo con el CLICOM. Para la evaluación de los modelos regionales se utilizaron series de lluvia acumulada mensual con longitud de registro de 71 años comprendidas en el periodo del año 1940 al año 2010.

6.35 Preparación de datos

Los datos utilizados en este estudio corresponden a la variable precipitación de lluvia en la escala mensual, sin embargo debido a la ausencia de datos en algunas estaciones se procedió a rellenar datos faltantes mediante la técnica interpolación de la distancia inversa (IDW).

Tabla 6.31. Estaciones climatológicas en Zacatecas.

Estaciones	Nombre	Latitud	Longitud	Elevación
32001	AGUA NUEVA, VILLA DE COS	23.78	102.17	2280
32002	BOCA DEL TEZORO, JEREZ	22.85	102.95	2045
32003	VICTOR ROSALES, CALERA	22.95	102.68	1153
32004	CAMACHO, MAZAPIL	24.42	102.32	1665
32005	CAÏTAS FELIPE PESCADOR	23.62	102.67	2090
32006	CAZADERO II, RIO GRANDE	23.70	103.07	1898
32010	COAPAS, MAZAPIL	24.78	102.15	2000
32011	CONCEPCION DE LA NORMA	24.53	101.97	2000
32013	EL CHIQUE, TABASCO	22.00	102.90	1575
32015	EL RUSIO, VILLA HIDALGO	22.42	101.77	2136
32016	EL PLATANITO, VALPARAISO	22.58	104.05	875
32017	EL TECOMATE, JALPA	21.97	103.05	1425
32018	EL SAUZ, FRESNILLO	23.18	103.23	2050
32019	EXCAME III, TEPECHITLA	21.63	103.35	1666
32020	FRESNILLO, FRESNILLO	23.17	102.85	2249
32021	GONZALEZ ORTEGA	23.97	103.42	2000
32022	GRUÏDORA, MAZAPIL	24.23	101.93	1919
32024	GPE. VICTORIA A.LA HONDA	22.40	101.82	2200
32025	HUANUSCO, HUANUSCO	21.75	102.98	1495
32027	JIMENEZ DE TEUL	23.10	104.13	2419
32028	JUAN ALDAMA, JUAN ALDAMA	24.30	103.40	2125
32030	LA FLORIDA, VALPARAISO	22.68	103.62	1950
32032	LA VILLITA, TEPECHITLAN	21.67	103.35	1686
32034	LOS PUERTOS(PTO.SAN FCO)	23.73	103.37	2000
32036	MAZAPIL, MAZAPIL	24.70	101.60	2250
32037	MONTE ESCOBEDO	22.30	103.57	2070
32038	NIEVES, GRAL.FCO.MURGUIA	24.05	102.92	1882
32039	NOCHISTLAN, NOCHISTLAN	21.37	102.85	1930
32040	NUEVO MERCURIO, MAZAPIL	24.18	102.13	2000
32041	OJOCALIENTE, OJOCALIENTE	22.58	102.23	2114
32042	PALOMAS, VILLANUEVA	22.28	102.82	1955
32046	SAIN ALTO, SAIN ALTO	23.60	103.25	2190
32047	SAN ANTONIO DEL CIPRES	22.95	102.52	2240
32048	SAN BENITO, MAZAPIL	23.90	101.73	2000
32049	SAN FRANCISCO	24.38	102.85	2210
32050	SAN GIL,GRAL.FCO.MURGUIA	24.23	102.97	2000
32051	SAN PEDRO PIEDRA GORDA	22.47	102.33	2055
32052	SAN RAFAEL, MAZAPIL	24.58	101.98	1706
32053	SANTA ROSA, FRESNILLO	22.93	103.10	2250
32054	SOMBRERETE, SOMBRERETE	23.65	103.62	2379
32055	TAYAHUA, VILLANUEVA	22.10	102.85	2000
32057	TLALTENANGO DE SANCHEZ R	21.78	103.32	1723

32058	TRANCOSO. GUADALUPE	22.73	102.33	2200
32061	VILLA HIDALGO,VILLA HGO.	22.33	101.72	2165
32069	SAN JUAN CAPISTRANO	22.50	104.08	1325
32070	TEUL DE GLEZ. ORTEGA	21.45	103.47	1787
32072	VILLA GARCIA (DGE)	22.15	101.95	2400
32073	VILLANUEVA, VILLANUEVA	22.33	102.90	1933
32078	SAN TIBURCIO, MAZAPIL	24.08	101.48	1890
32084	SAN ANDRES, VILLA DE COS	23.70	101.87	1900
32093	EL ARENAL, SOMBRERETE	23.62	103.42	2000
32097	TENAYUCA, APULCO	21.42	102.72	2000
32098	VALPARAISO, VALPARAISO	22.77	103.58	1950
32069	SAN JUAN CAPISTRANO	22.50	104.08	1325
32065	CHALCHIHUITES	23.48	103.88	2300
32099	MALPASO, VILLANUEVA	22.60	102.78	2077
32100	SAN ISIDRO DE LOS GLEZ.	22.83	103.38	2000
32101	SAN JOSE DE LLANETES	22.92	103.30	2000
32103	AMECA LA VIEJA	22.92	103.72	2280
32105	TLACHICHILA, NOCHISTLAN	21.55	102.78	2000
32107	SIERRA HERMOSA, V.DE COS	23.65	101.72	2155
32108	CORRALES, SOMBRERETE	22.22	103.63	2190
32109	SAN PEDRO DE LA SIERRA	22.77	103.93	1850
32111	JALPA, JALPA	21.63	102.97	1390
32112	MILPILLAS DE ALLENDE	21.32	103.60	1920
32113	PASTORIA, MONTE ESCOBEDO	22.55	103.37	1855
32114	VILLA GONZALEZ ORTEGA	22.52	101.92	2170
32116	ADJUNTAS DEL REFUGIO	23.63	103.42	2280
32117	LAS ANIMAS, VALPARAISO	22.48	103.83	2040
32118	LOS CAMPOS, VILLA GARCIA	22.02	101.83	2000
32119	F. ANGELES, VILLANUEVA	22.53	102.78	2090
32120	GENARO GODINA,CUAUHTEMOC	22.48	102.45	2130
32121	GUADALUPE, GUADALUPE	22.75	102.52	2265
32122	LUIS MOYA, LUIS MOYA	22.43	102.23	1940
32123	MESILLAS, SOMBRERETE	23.47	103.57	2300
32124	MOMAX, MOMAX	21.92	103.30	2000
32125	MOYAHUA DE ESTRADA	21.25	103.12	1108
32126	PALMILLAS, OJO CALIENTE	22.65	102.35	2205
32127	PINO SUAREZ, PINO	22.12	101.38	2160
32129	PRESA SUSTICACAN	23.60	103.13	2070
32130	EL ROMERILLO, VALPARAISO	22.62	103.65	2030
32135	POTRERO DE GALLEGOS	22.63	103.68	2075
32137	CHICHIMEQUILLAS	23.23	102.57	2060
32139	HUITZILA, TEUL GLEZ. O.	21.22	103.60	1675
32140	MEZQUITAL DEL ORO	21.20	103.37	1200
32141	SAN SALVADOR	24.55	100.87	2000
32142	TIERRA Y LIBERTAD	23.45	102.38	2030
32143	GARCIA DE LA CADENA	21.20	103.47	1740

6.36 Metodología

En el análisis de lluvia y escurrimientos se ensayaron los modelos tradicionales y los modelos regionales propuestos, estos últimos brindaron resultados satisfactorios en el pronóstico mensual, trimestral y semestral de dichas variables. En este análisis, se comparan los modelos regionales RNAs, Ensamblado y Modular para pronóstico de lluvia, las estradas que alimentan a los modelos corresponden a la lluvia acumulada mensual de varias estaciones que integran a

una misma región homogénea. El algoritmo de Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) fue utilizado para entrenar los modelos regionales.

En los modelos regionales propuestos se ha aprovechado la bondad del analisis no lineal mediante RNAs, esto mejora el pronóstico debido a una mayor flexibilidad y gama de información espacial. El entrenamiento es como sigue; la entrada de datos son las series acumuladas mensuales de lluvia. El entrenamiento corresponde al 90% de la serie mientras que el restante 10% es para validación y evaluación. Entonces una vez entrenado el modelo se pronostica en el periodo de evaluación la variable estudiada, que para este caso corresponde a la lluvia.

6.37 Delimitación de regiones homogéneas

El objetivo principal en la delimitación de regiones homogéneas es formar grupos que se aproximen satisfactoriamente a la condición de homogeneidad. Una alternativa es formar grupos mediante la técnica de momentos-L.

Una primera aproximación de la formación de regiones homogéneas se hace con el estadístico coeficiente de variación L (L-CV) y el coeficiente de asimetría L (Skewness-L) los grupos entonces son medidos a través de un proceso iterativo hasta que finalmente se obtiene regiones aceptablemente homogéneas.

El procedimiento es como sigue teniendo una región propuesta con N sitios donde cada sitio i tiene una longitud de registro ni y sus radios de momentos L t^i , t_3^i y t_4^i y se denotan como t^R , t_3^R y t_4^R siendo los estadísticos media regional L , coeficiente de asimetría L y a la curtosis L siendo proporcionales a los estadísticos de cada uno de los sitios de la región.

$$t^R = \frac{\sum_{i=1}^N ni t^i}{\sum_{i=1}^N ni} \tag{6.11}$$

Para calcular la desviación estándar de un sitio mediante Momentos-L (L-Cvs) es como sigue.

$$V = \left\{ \frac{\sum_{i=1}^N ni (t^i - t^R)^2}{\sum_{i=1}^N ni} \right\}^{\frac{1}{2}} \tag{6.12}$$

Después se debe encontrar con los radios de momentos L regionales la distribución kappa se simulan un numero largo N_{sim} de una región con N sitios para cada simulación de la región se calcula V de las simulaciones se determina la media y la desviación estándar de las N_{sim} valores de V denominados como μv y σv .

Para calcular la medida de heterogeneidad es como sigue $H = \frac{V - \mu v}{\sigma v}$ y se define a la región como heterogénea si H es lo suficientemente grande en la literatura se sugiere $H \geq 2$ en caso contrario se tiene una región homogénea.

6.38 Resultados

En la delimitación de las regiones homogéneas en el estado de Zacatecas mediante la técnica de Momentos-L se obtuvieron 11 regiones potencialmente homogéneas como se muestra en la figura 6.80. La figura abajo muestra las primeras aproximaciones en la formación de regiones homogéneas mediante el coeficiente de asimetría-L y la media-L. Estas primeras aproximaciones de regiones homogéneas se refinaron mediante la técnica de Momentos-L.

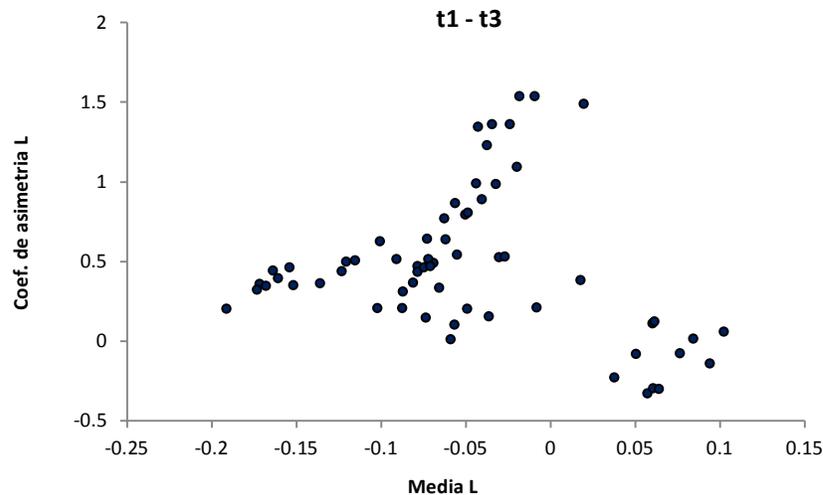


Figura 6.79. Agrupamiento de las estaciones climatológicas del acumulado anual de lluvia en el estado de Zacatecas.

Después de la formación de regiones homogéneas mediante la técnica de Momentos-L se compararon los modelos regionales RNAs, Ensamblado y Modular. Esto es en el periodo de evaluación mediante las medidas de eficiencia se compararon los resultados del pronóstico de lluvia mensual en cada uno de los grupos homogéneos formados. La arquitectura de las RNAs utilizadas fue la del perceptrón multicapa y el número de neuronas de cada capa se calculó mediante la expresión $\frac{4n+18}{n-1}$. Las funciones de transferencia usadas en las RNAs para cada uno de los modelos fueron la sigmoidea y la tangente hiperbólica. En el caso de los modelos Ensamblados y Modulares primero se entrenaron las RNAs que conforman a los miembros de ambos modelos. Particularmente en los modelos Ensamblados no se notó diferencias significativas en los algoritmos de clasificación mediante AdaBoost.M2, LPBoost, TotalBoost y RusBoost.

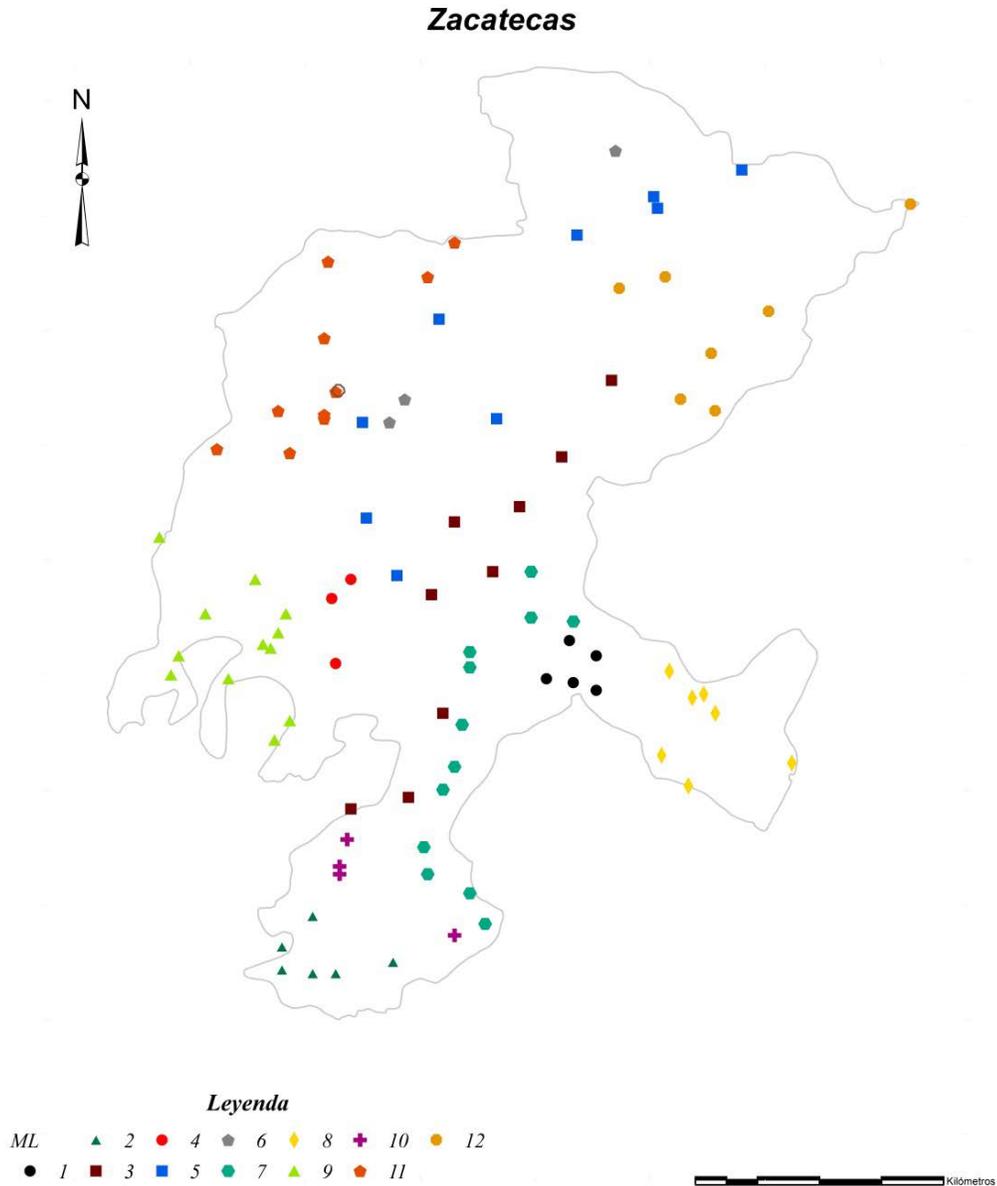


Figura 6.80. Formación de grupos homogéneos mediante Momentos-L en Zacatecas.

Los resultados de la comparación de los modelos regionales RNAs, Ensamblado y Modular en el pronóstico mensual de las series de lluvia en Zacatecas demostraron que el modelo Ensamblado siempre fue superior a los modelos RNAs y Modular regionales. Las medidas de eficiencia MSE, RMSE, MAE y d en general demuestran que el modelo Ensamblado es superior puesto que en el 63% de las estaciones analizadas fue superior a los modelos Modular y RNAs regional, esto es en el 37% restante. El modelo Modular obtuvo una mayor eficiencia en el 31% de las estaciones mientras que el modelo RNAs regional no obtuvo un porcentaje considerable en la comparación con los modelos Ensamblado y Modular regional puesto que solo fue eficiente en el 6% de las estaciones.

En la tabla 6.32 se muestran los resultados obtenidos de las medidas de eficiencia para los diferentes ajustes de los modelos RNAs, Ensamblado y Modular regionales en la región

homogénea uno de Zacatecas y en las figuras 6.80 a la 6.87 se muestran los ajustes obtenidos con los modelos analizados así como el coeficiente R^2 . Este último coeficiente ha favorecido al modelo Ensamblado regional pues en el 63% de las estaciones ha sido superior en el ajuste de las estaciones analizadas, mientras que en el 24% fue superior el modelo Modular regional, por último el modelo RNAs regional tan solo resultó ser eficiente en el 13% de las estaciones.

Tabla 6.32. Comparación de los modelos RNAs, Modular y Ensamblado regionales en la región 1.

	Estaciones	32001	32002	32017	32020	32073	32124	32137	32142
RNAs-Regional	MSE	7270.5	9570.6	8406.9	8167.6	13775	10124	9493.6	9833.3
	RMSE	85.267	97.829	91.689	90.374	117.36	100.62	97.435	99.163
	MAE	60.140	75.802	68.771	63.841	86.231	69.026	69.696	72.878
	<i>d</i>	0.3680	0.4661	0.8368	0.5718	0.5823	0.8207	0.3578	0.5077
Modular-Regional	MSE	4770.6	4398.9	4623.8	4779.7	8594.4	9725.8	6072.9	7275.2
	RMSE	69.07	66.324	67.999	69.135	92.706	98.619	77.928	85.295
	MAE	48.070	53.158	50.917	51.550	73.289	76.773	56.736	68.457
	<i>d</i>	0.5927	0.7889	0.9026	0.7573	0.7665	0.8088	0.6499	0.5937
Ensamblado-Regional	MSE	4769.8	4505.5	5744.8	4357.8	8805.3	6732.3	5062.7	5418.9
	RMSE	69.064	67.123	75.794	66.014	93.836	82.050	71.153	73.613
	MAE	48.385	51.888	56.868	47.911	74.929	58.217	55.882	56.265
	<i>d</i>	0.602	0.7839	0.8892	0.7807	0.7653	0.8886	0.7219	0.7421

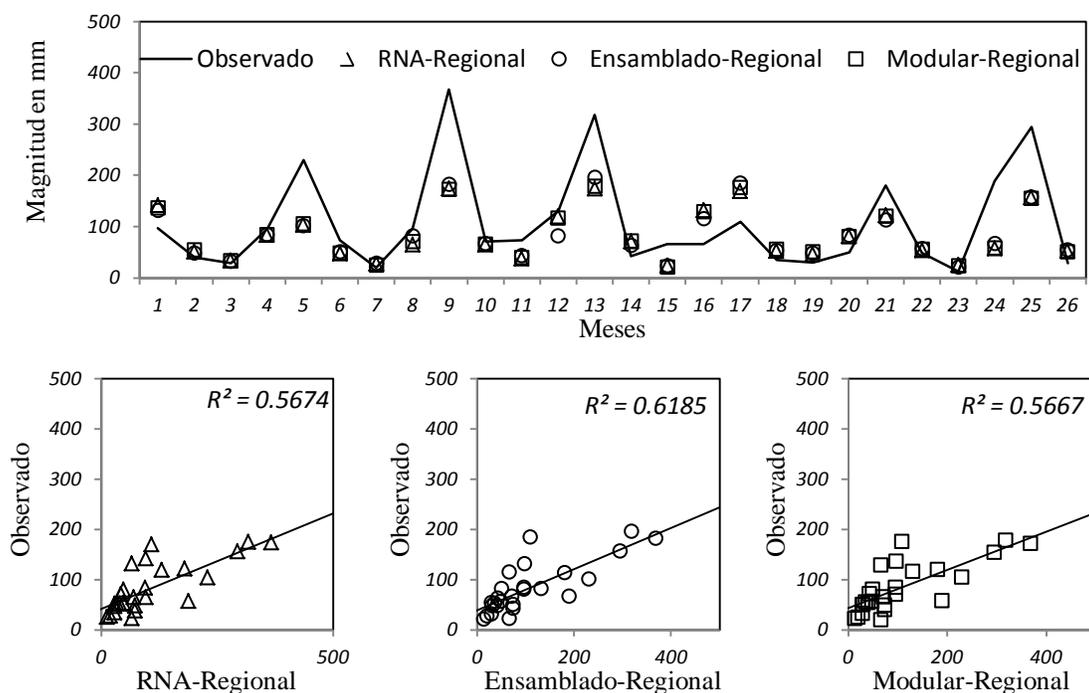


Figura 6.81. Pronóstico mensual de lluvia acumulada en la estación 32001.

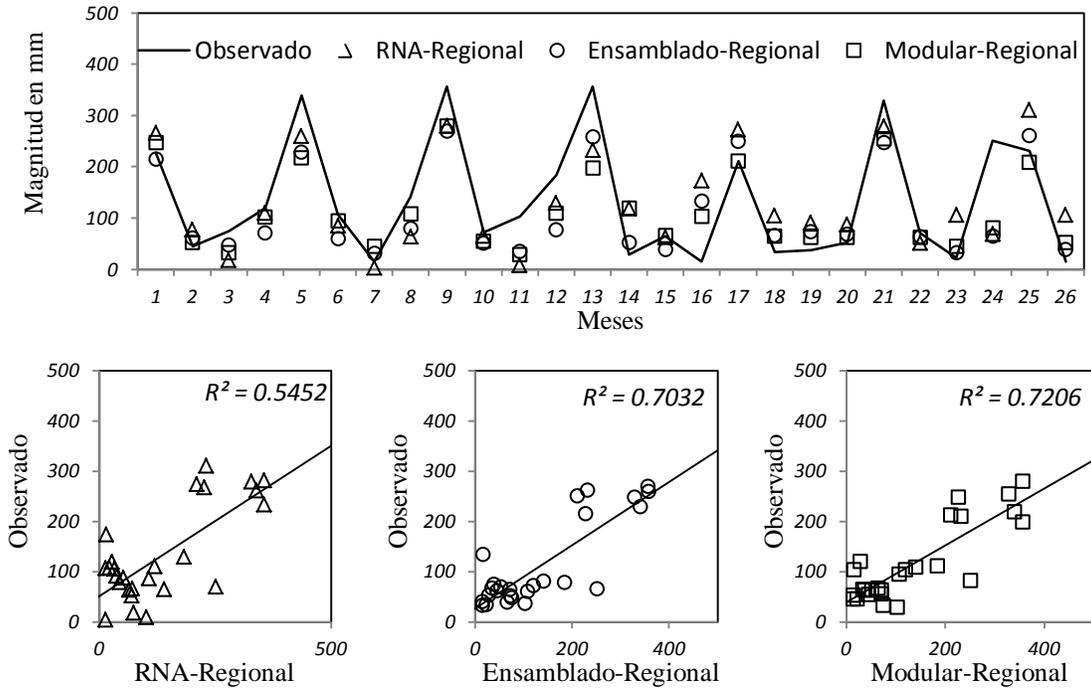


Figura 6.82. Pronostico mensual de lluvia acumulada en la estación 32002.

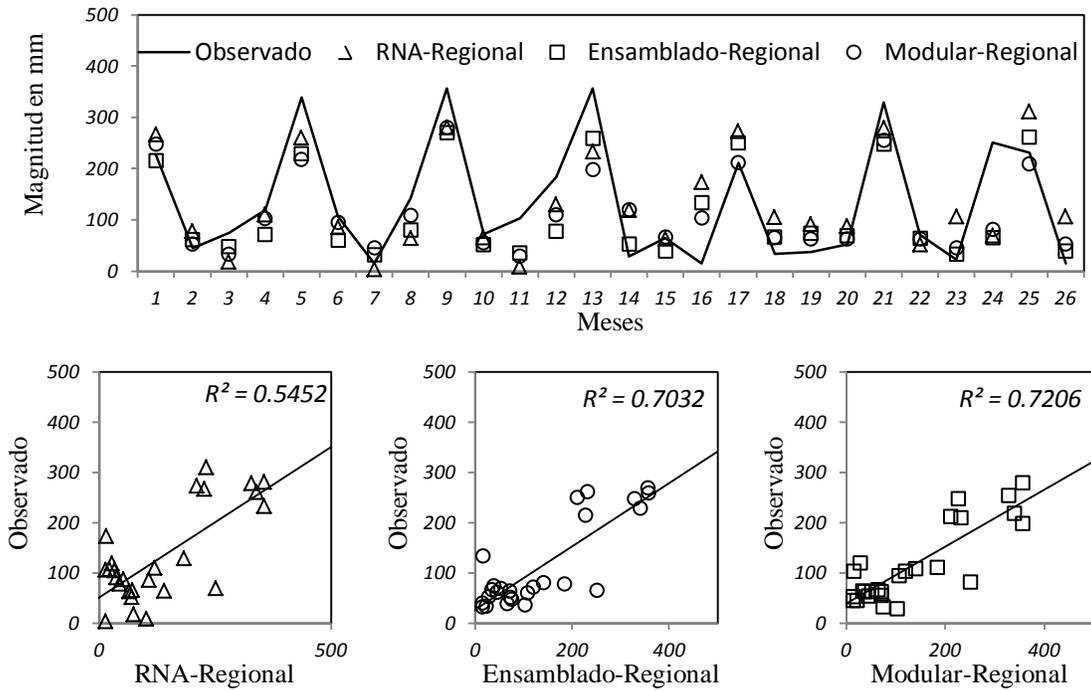


Figura 6.83. Pronostico mensual de lluvia acumulada en la estación 32017.

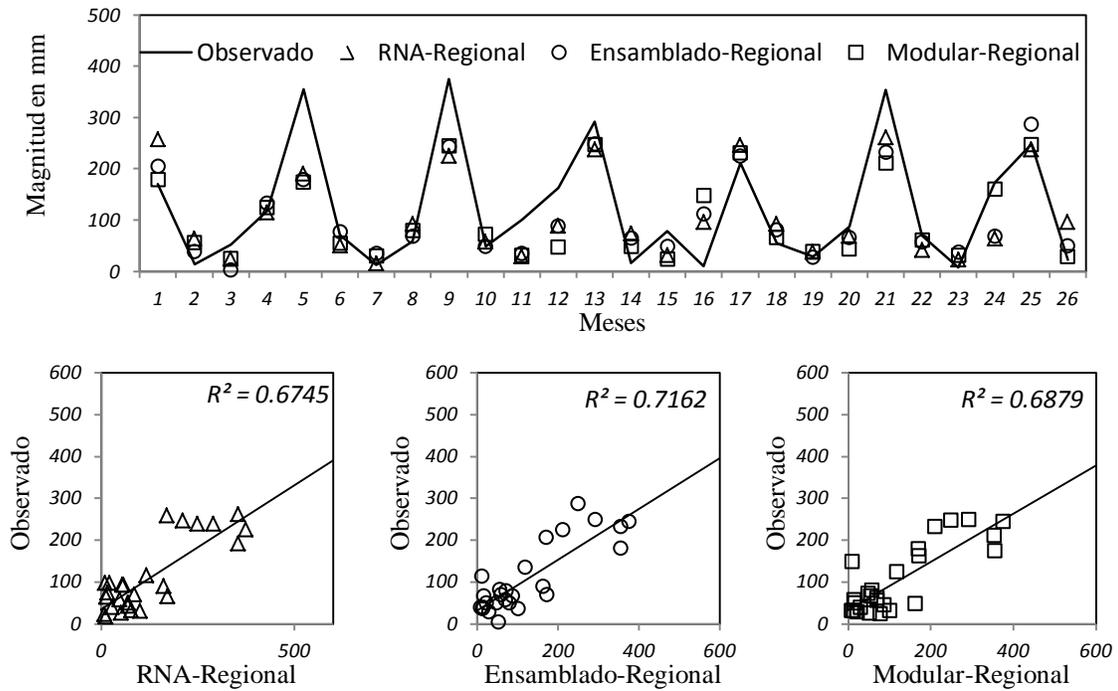


Figura 6.84. Pronostico mensual de lluvia acumulada en la estación 32020.

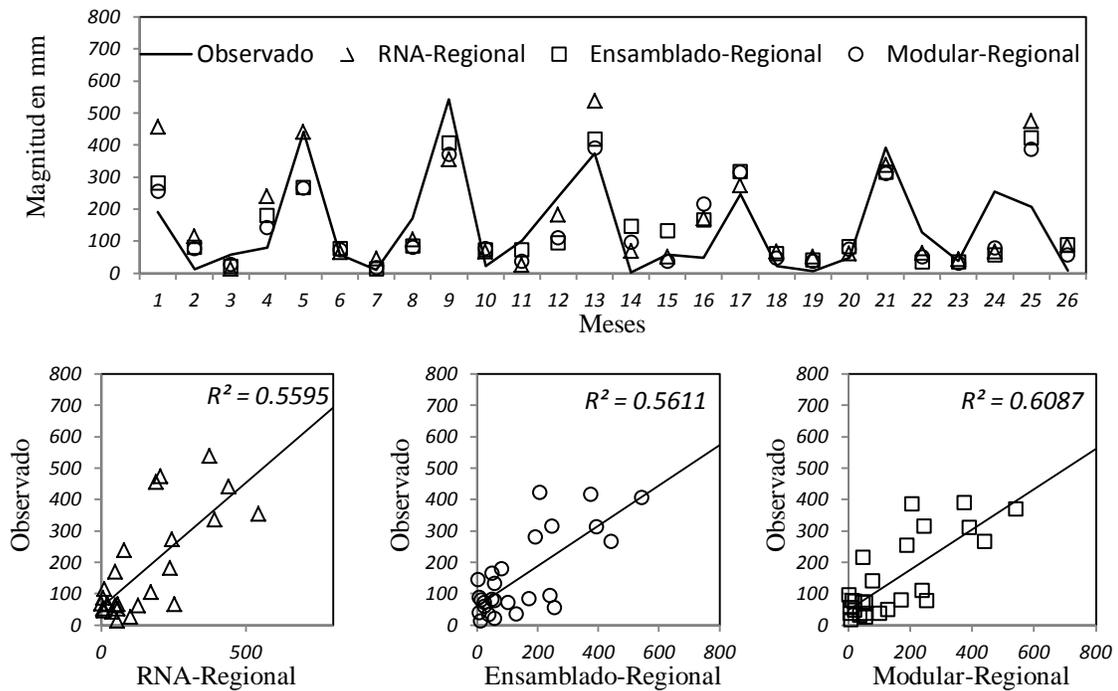


Figura 6.85. Pronostico mensual de lluvia acumulada en la estación 32073.

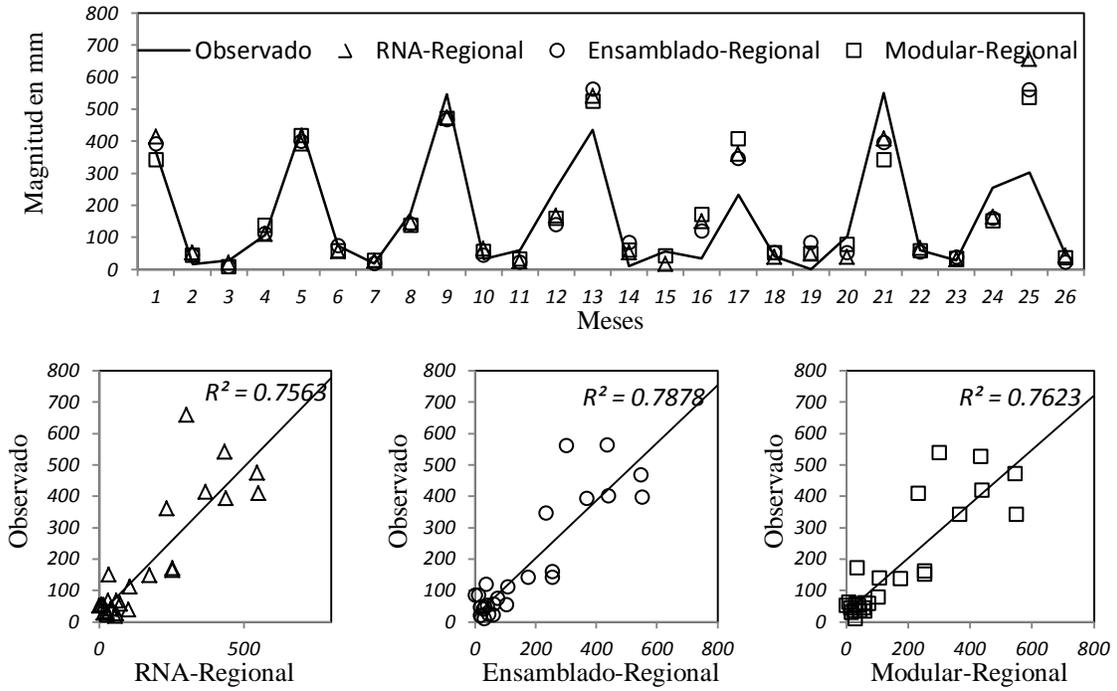


Figura 6.86. Pronostico mensual de lluvia acumulada en la estación 32124.

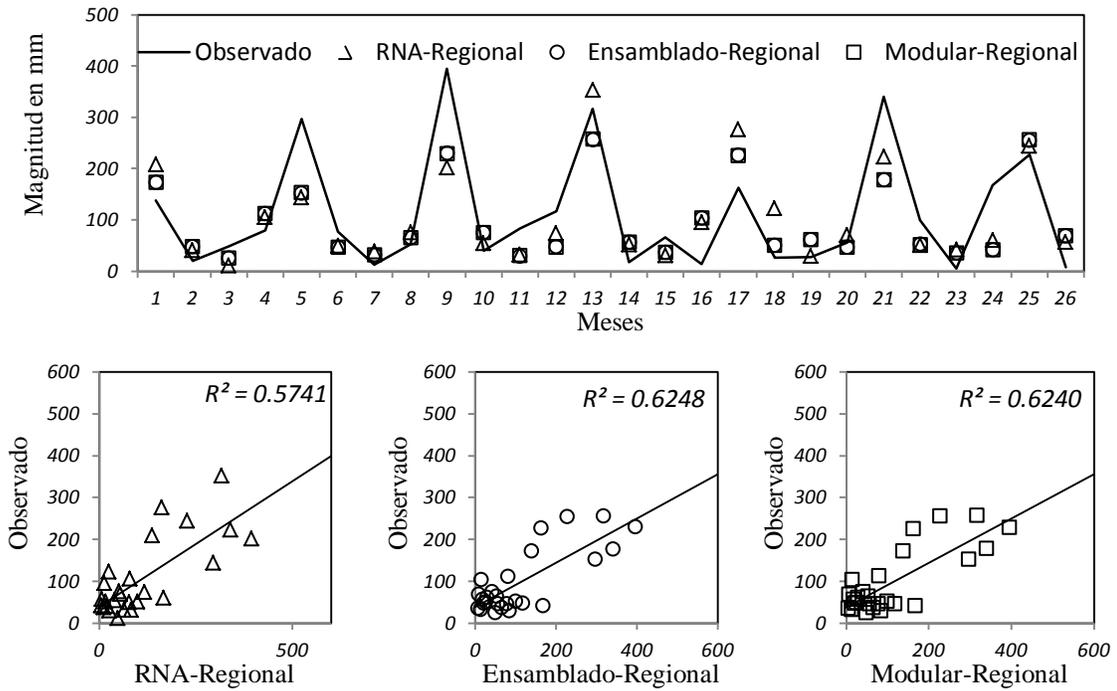


Figura 6.87. Pronostico mensual de lluvia acumulada en la estación 32137.

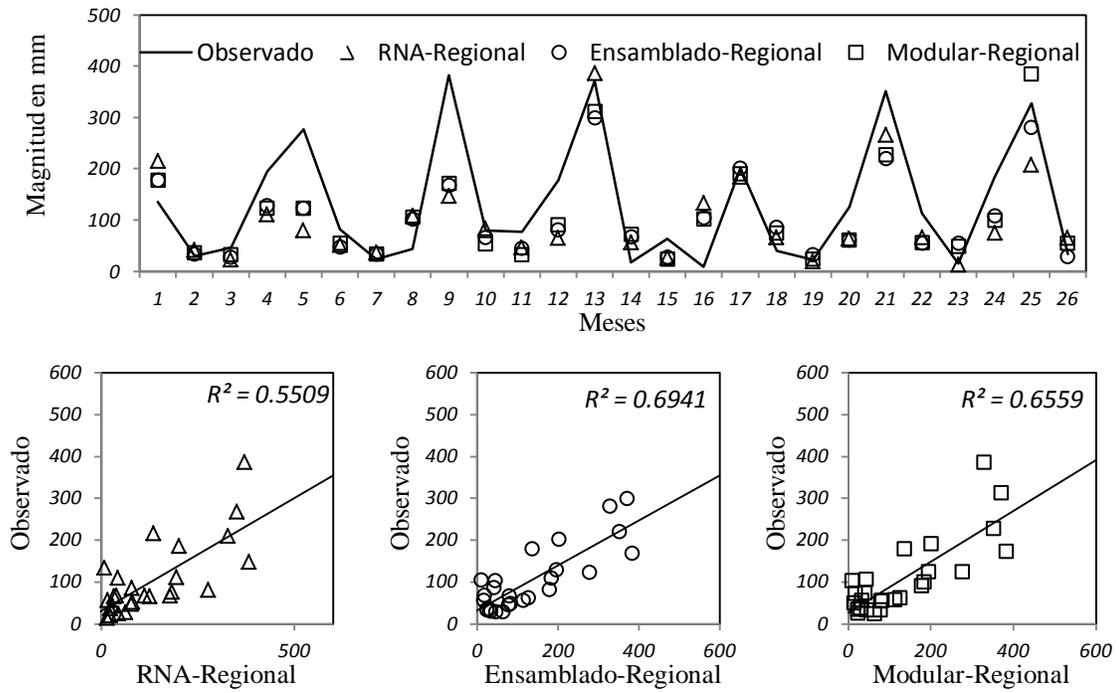


Figura 6.88. Pronostico mensual de lluvia acumulada en la estación 32142.

6.39 Referencias

- Abdel Karim M. Baareh, Alaa F Sheta y Khaled AI Khnaifes. (2006). Forecasting River Flow in the USA. A Comparison between Auto-Regression and Neural Network Non Parametric Models. *Journal of Computer Science* 2(10): 775-780.
- Ashakar, F.; T.B.M.J Ouarda, R. Roy and B. Bobée. (1993). Robust estimators in hydrologic frequency analysis, in *Engineering Hydrology*. Edited by C.Y, pp 347-352, Am. Soc. Civ. Eng.
- Cameron M. Zealand, Donald H. Burn, Slobodan P. Simonovic. (1999). Short term streamflow forecasting using artificial neural networks, *Journal of Hydrology* 214, 32–48.
- Cheleng A. Arslan. (2013). Artificial neural network models investigation for Euphrates river forecasting & back casting. *Journal of Asian Scientific Research*,3(11):1090-1104.
- French, M. N., Krajewski, W. F., and Cuykendall, R. R. (1992). Rainfall forecasting in space and time using neural network, *J. Hydrol.*, 137, 1–31.
- Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*. Macmillan, New York.
- Hikemet Kerem Cigizoglu. (2003). Estimation, forecasting and extrapolation of river flows by artificial neural networks *Hydrological Sciences–Journal–des Sciences Hydrologiques*, 48 (3).
- Hosking J.R.M. and Wallis J.R., 1997. *Regional Frequency Analysis. An Approach Based on L-moments*. Cambridge University Press, Cambridge, New York, Melbourne, 224 pp
- Jamileh Farajzadeh, Ahmad Fakheri Fard, Saeed Lotfi. (2014). Modeling of monthly rainfall and runoff of Urmia lake basin using “feed-forward neural network” and “time series analysis” model, *Water Resources and Industry Volumes 7–8, Pages 38–48*.
- John Abbot & Jennifer Marohasy. (2014). Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks, *Atmospheric Research Volume 138, Pages 166–178*.
- Jorge O. Pierini, Eduardo A. Gómez & Luciano Telesca. (2012). Prediction of water flows in Colorado River, Argentina, *Lat. Am. J. Aquat. Res.*, 40(4): 872-880.
- K. Mohammadi, H. R. Eslami and Sh. Dayyani Dardashti. (2005). Comparison of Regression, ARIMA and ANN Models for Reservoir Inflow Forecasting using Snowmelt Equivalent (a Case study of Karaj), *J. Agric. Sci. Technol. Vol. 7: 17-30*.
- Khalida Benmahdjoub, Zohra Ameer & Mina Boulifa. (2013). Forecasting of Rainfall Using Time Delay Neural Network in Tizi-Ouzou (Algeria). *Energy Procedia Volume 36, Pages 1138–1146*.
- Kin C. Luk, J.E. Ball, A. Sharma. (2001). An application of artificial neural networks for rainfall forecasting, *Mathematical and Computer Modelling Volume 33, Issues 6–7, Pages 683–693*.
- Lorenz E.N. (1963). Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20, 130-141.
- Lorenz E.N. (1969). Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences*, 26, 636-646.
- Lorenz E.N. (1982). Atmospheric predictability experiments with a large numerical model. *Tellus*, 34, 505-513.
- M. J. Diamantopoulou, P. E. Georgiou, and D. M. Papamichail. (2006). A time delay artificial neural network approach for flow routing in a river system, *Hydrol. Earth Syst. Sci. Discuss.*, 3, 2735–2756.
- Mehmet C. Demirel & Martijn J. Booij. (2009). Identification of an appropriate low flow forecast model for the Meuse River. *Hydroinformatics in Hydrology, Hydrogeology and Water Resources (Proc. of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad, India. IAHS Publ. 331*.
- Mehmet C. Demirel, Anabela Venancio, Ercan Kahya. (2009). Flow forecast by SWAT model and ANN in Pracana basin, Portugal. *Advances in Engineering Software* 40, 467–473.
- N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi. (2009). An artificial neural network model for rainfall forecasting in Bangkok, Thailand, *Hydrol. Earth Syst. Sci.*, 13, 1413–1425.
- P.T. Nastosa, K.P. Moustrisb, I.K. Larissic y A.G. Paliatsosd. (2013). Rain intensity forecast using Artificial Neural Networks in Athens, Greece, *Atmospheric Research Volume 119, Pages 153–160*.
- Ray, C. and Klindworth, K. (2000). "Neural Networks for Agrichemical Vulnerability Assessment of Rural Private Wells." *J. Hydrol. Eng.*, 5(2), 162–171.
- Wang W. (2006). *Stochasticity, nonlinearity and forecasting of streamflow processes*. IOS Press.

CAPÍTULO**SIETE**

CONCLUSIONES Y RECOMENDACIONES

En esta tesis se ha propuesto un enfoque regional para la predicción de lluvia y de escurrimientos en hidrología. Esto es mediante modelos no lineales, como lo son Redes Neuronales Artificiales RNAs, modelos Ensamblados basados en RNAs y mediante modelos Modulares también basados en RNAs.

En la evaluación de las series de lluvias mensuales se han aplicado los modelos de pronóstico regionales propuestos y tradicionales. Esto es RNAs regional, Ensamblado regional y Modular regional. Los modelos considerados tradicionales son RNAs, Ensamblado y Modular. Las zonas de estudio fueron Australia, Canadá, la cuenca del Amazonas, la Costa Oeste de Estados Unidos y Zacatecas. En el pronóstico de lluvia en Australia se analizaron 118 estaciones climatológicas. Los grupos homogéneos formados fueron 15. De acuerdo a los resultados obtenidos mediante medidas de eficiencia MSE, RMSE, MAE y el índice de concordancia d , el enfoque regional propuesto ha demostrado ser más eficiente que los modelos tradicionales. Los modelos regionales fueron desarrollados mediante la combinación de dos pasos: 1) El primero consiste en la agrupación de estaciones climatológicas mediante el algoritmo de Fisher para después lograr un refinamiento mediante Momentos- L y 2) El segundo paso consiste en utilizar RNAs, Modelos Ensamblados basados en RNAs y Modulares basado en RNAs para hacer un pronóstico regional. El promedio del porcentaje de las medidas de eficiencia en el modelo RNAs regional fue de un 60% mientras que el modelo RNAs tradicional logró una eficiencia favorable en el 40%. En el caso de los modelos Ensamblados, el porcentaje del promedio de las medidas de eficiencia favoreció al modelo regional en el 63.5% de las estaciones mientras que para el modelo tradicional fue de 36.5%. Finalmente para el modelo Modular regional el promedio de las medidas de eficiencia fue del 58.5% y para el modelo Modular tradicional de 41.5%. Uno de los principales aportes de los modelos regionales es mejorar la precisión beneficiando la predicción de la lluvia, esto es con la selección de los criterios adecuados en el modelo. Los resultados de la simulación demuestran que los modelos regionales propuestos son más eficientes que los modelos tradicionales.

Para el pronóstico de escurrimientos mensuales se aplicaron los modelos tradicionales y regionales propuestos; estos son RNAs, Ensamblados y Modulares. Los modelos regionales fueron utilizados mediante grupos homogéneos delimitados mediante c-Medias y ML. Los modelos regionales similares a los de pronóstico de lluvias fueron desarrollados mediante la combinación de dos pasos: 1) Clasificación regional mediante el análisis c-Medias y Momentos- L y 2). Los modelos Regionales fueron comparados con los modelos tradicionales para el pronóstico de escurrimientos mensuales usando datos de 103 estaciones hidrométricas en el suroeste y sureste de Canadá. Para aplicar los modelos regionales se formaron 11 regiones homogéneas. Los resultados de la comparación en promedio de las medidas de eficiencia para los modelos RNAs regionales fue de 42% para RNAs-ML, de 38% para RNAs-c-Medias y finalmente de 21% para el modelo tradicional de RNAs. En la comparación de los modelos Ensamblados el promedio de las medidas de eficiencia fue del 46% en el modelo Ensamblado regional mediante ML del 28% en el modelo Ensamblado regional mediante c-Medias y del 25% para el modelo Ensamblado tradicional, por último el modelo Modular regional mediante ML fue mejor en un 46%, después el modelo Modular tradicional con 32%. Los modelos regionales proporcionaron resultados más precisos que los modelos tradicionales. Se cree que el éxito de los modelos regionales al ser más precisos es debido a que los Momentos-L son una poderosa herramienta para la determinación de las regiones homogéneas debido a que más información se utiliza en el análisis "en el lugar" en lugar de datos de un solo sitio, ya que existe la posibilidad de una mayor precisión en la estimación de los cuantiles (Hoskin, 1997). Momentos-L mejora la capacidad de pronóstico del modelo de RNAs para una mayor precisión en la estimación.

Para la evaluación de escurrimientos en periodos de tiempo mensual, trimestral y semestral se comparó un modelo ensamblado tradicional y otro mediante el enfoque regional propuesto, ambos basados en RNAs. La zona de estudio correspondió a la cuenca del Amazonas donde se tomaron 18 estaciones hidrométricas. Los resultados obtenidos de la comparación del modelo ensamblado propuesto (Ensamblado regional) y el modelo tradicional de redes neuronales (Ensamblado) indican que el modelo Ensamblado regional ha sido superior en un 90% de las estaciones analizadas, esto es el promedio de las medidas de eficiencia para series mensuales. En el caso de las series trimestrales se observó que el modelo Ensamblado regional fue superior al modelo tradicional y finalmente en las series semestrales también el modelo regional fue favorecido con un 83%.

En el análisis de la lluvia para los periodos de tiempo mensual, trimestral y semestral se utilizó un modelo Ensamblado mediante el enfoque regional propuesto que nace de la combinación con Momentos-L (Ensamblado regional) y el modelo tradicional Ensamblado (Ensamblado). La región de estudio correspondió a la Costa Oeste de Estados Unidos en la cual se formaron 12 regiones homogéneas mediante la técnica de Momentos-L. Se utilizaron las series de lluvia tomados de 56 estaciones climatológicas de la red de estaciones de la National Oceanic and Atmospheric Administration (NOOA) ubicadas en la Costa Oeste de Estados Unidos. Los resultados demostraron que para las series mensuales en más de un 60% el modelo Ensamblado regional fue más eficiente. En el caso de las series trimestrales el modelo más eficiente fue el modelo Ensamblado regional, pues obtuvo mejores resultados en más del 80% de las estaciones analizadas finalmente en las series semestrales el modelo Ensamblado regional fue superior en un 85%. Los resultados que se obtuvieron en la comparación de las

estaciones dentro de la región Costa Oeste de Estados Unidos sugieren que en esta zona y para la lluvia el modelo Ensamblado regional provee una mejor predicción en la mayoría de las estaciones y para los diferentes periodos analizados.

De acuerdo a los resultados de la comparación de lluvia y escurrimientos en diferentes regiones y para los periodos de tiempo mensual, trimestral y semestral los mejores ajustes los brindaron los modelos regionales RNAs, Ensamblado y Modular. Es por ello que en esta última comparación se decidió evaluar únicamente a los modelos regionales y así poder determinar cuál de los tres modelos de predicción mediante el enfoque regional propuesto era el más eficiente. La zona de estudio evaluada fue el estado de Zacatecas de donde se tomaron 87 estaciones climatológicas de la base de datos CLICOM y se formaron 12 grupos homogéneos mediante la técnica de Momentos-L. Los resultados indicaron que en promedio de las medidas de eficiencia el modelo Ensamblado regional fue superior en un 63%, mientras que el modelo RNAs regional en un 31% y finalmente el modelo RNAs regional en el 6% del total de las estaciones analizadas. Fue notable que el modelo Ensamblado ha sido el más eficiente en el pronóstico mensual de lluvia para el estado de Zacatecas. En cuanto a los algoritmos de asignación AdaBoost.M2, LPBoost, TotalBoost y RusBoost no se encontraron diferencias significativas.

En referencia a los objetivos originales de este estudio, se determinó que el enfoque Regional propuesto mediante Momentos-L en combinación con modelos de RNAs, Ensamblados y Modulares basados en RNAs pueden ser utilizados con una alta precisión para la predicción de lluvia y escurrimientos. El uso de las series de lluvia y de escurrimientos de las regiones homogéneas como entradas a los modelos regionales propuesto ayudó a proporcionar pronósticos más exactos de acuerdo con las medidas de eficiencia. Los resultados indican que el enfoque regional propuesto desarrolla nuevos métodos prometedores en la predicción de lluvias y escurrimientos a corto plazo en hidrología. A pesar de la mejora en la predicción mediante los modelos regionales planteados en esta tesis se espera que la principal contribución sea la de inspirar futuras investigaciones y el uso de los modelos regionales en las dependencias encargadas en la planeación de los recursos hídricos del país.

En estudios futuros se espera una mejora en la formación de regiones homogéneas para obtener un mejor rendimiento en la predicción de la lluvia a diferentes escalas incluyendo escalas horarias.

Apéndice A

A1 Componentes básicos de la red neuronal

Una red neuronal es un paradigma computacional matemático general que modela las operaciones de los sistemas neuronales biológicos. En 1943, McCulloch un neurobiólogo y Pitts un estadístico, publicaron un artículo seminal titulado “Un cálculo lógico de las ideas inminentes en la actividad nerviosa” en el boletín de Biofísica Matemática. Este artículo inspiró el desarrollo de la computadora digital moderna o el cerebro electrónico como lo llamó John Von Neumann. Aproximadamente al mismo tiempo, Frank Rosenblatt también fue motivado por este trabajo para investigar el proceso de cálculo del ojo, lo que finalmente llevó a la primera generación de redes neuronales, conocido como el perceptrón. En esta sección se ofrece una breve descripción de modelos de RNAs.

A2 Modelo neuronal de McCullochand Pitts

Entre los numerosos modelos de redes neuronales que se han propuesto en los últimos años, todos comparten un bloque de construcción común conocido como neurona y una estructura de interconexión en la red. El modelo de neurona más utilizado se basa en el trabajo de McCulloch y Pitts y se ilustra en la Figura A.1.

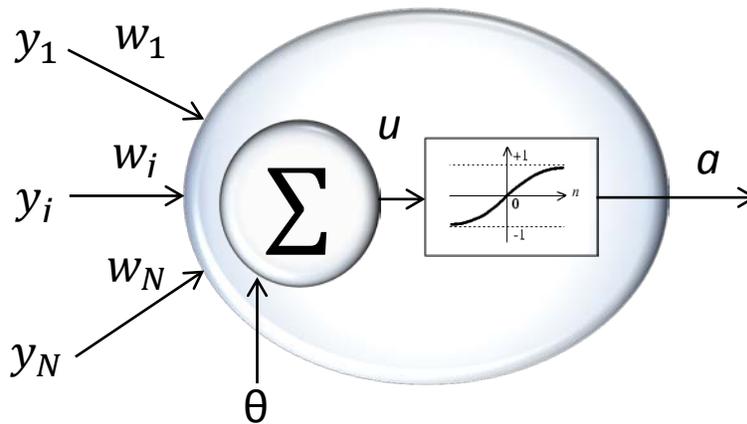


Figura A1. Modelo neuronal de McCulloch y Pitts

En la figura A1 cada neurona consiste en dos partes: la función de la red y la función de activación. La función de la red determina las entradas de la red ($y_j; 1 \leq j \leq N$) son combinadas dentro de la red. Una combinación lineal es adoptada.

$$u = \sum_{j=1}^N w_j y_j + \theta \tag{A1}$$

($w_j; 1 \leq j \leq N$) Son parámetros conocidos como pesos sinápticos. El parámetro θ es conocido como bias y es usado como umbral. En la literatura otros tipos de combinación de entradas en las redes han sido propuestos también. El resumen se muestra en la siguiente tabla.

La salida de la neurona denotada por a esta relacionada a la entrada de la red u_j mediante una transformación lineal o no lineal, llamada función de activación.

$$a = f(u)$$

Tabla A1. Resumen de funciones de red.

Función de red	Fórmula	Comentario
Lineal	$u = \sum_{j=1}^N w_j y_j + \theta$	Más comúnmente usada
De alto orden (fórmula de segundo orden)	$u = \sum_{j=1}^N \sum_{k=1}^N w_{jk} y_j y_k + \theta$	u_i Es una combinación lineal de pesos de alto orden polinomial. El número de entradas es igual N^d donde d es el orden del polinomio.
Delta ($\Sigma - \Pi$)	$u = \prod_{j=1}^N w_j y_j$	Raramente usada

En varios modelos de redes neuronales han sido propuestas diferentes funciones de activación. Las funciones de activación más comúnmente usadas son mostradas en la siguiente tabla.

Tabla A2. Funciones de activación neuronal.

Función de activación	Fórmula $a = f(u)$	Derivada $\frac{df(u)}{du}$	Comentario
Sigmoidal	$f(u) \frac{1}{1 + e^{-u}}$	$f(u) [1 - (f(u))^2]$	Más comúnmente usada
Tangente hiperbólica	$f(u) \tanh(u)$	$f(u) [1 - f(u)^2]$	Es también comúnmente usada
Tangente inversa	$f(u) = \frac{2}{\pi} \tan^{-1}(u)$	$\frac{2}{\pi} \cdot \frac{1}{1 + u^2}$	Usada con menos frecuencia
Acotada	$f(u) = \begin{cases} 1 & u > 0 \\ -1 & u < 0 \end{cases}$	<i>Las derivadas no existen para $u = 0$</i>	
Base radial Gaussiana	$f(u) = \exp[-\ u - m\ ^2 / \sigma^2]$	$-2(u - m) f(u) / \sigma^2$	Usada para redes neuronales de base radial
Lineal	$f(u) = au + b$	a	

A3 Arquitectura de la red neuronal

Como ya se ha comentado una red neuronal artificial (RNAs) es un esquema matemático inspirado en la estructura del sistema nervioso basándose en la neurona como elemento fundamental (Hilera, 1995). La arquitectura de una red neuronal es formada conectando múltiples procesadores elementales (Solomatine, 2002), y se clasifica según el número de capas como redes monocapa y redes multicapa, ambas arquitecturas forman un sistema con múltiples parámetros, teniendo la capacidad de reproducir complejas relaciones no lineales.

A4 Redes monocapa

En esta arquitectura las neuronas que componen esta única capa deben cumplir las tres funciones: de entrada de datos, de procesamiento de datos y de salida de información

pudiendo tener conexiones autoconcurrentes o no. Algunas de las implementaciones más conocidas son: Shunting Grossberg (SG) (Carpenter et al., 1987); Optimal Linear; Boltzman Machine (Ackley et al., 1985) Cauchy Machine (Ackley et al., 1985).

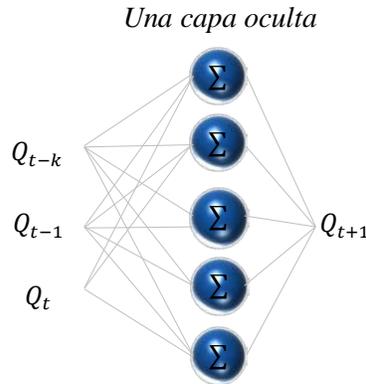


Figura. A.2. Red monocapa

A5 Redes multicapa

Este tipo de redes disponen de conjuntos de neuronas agrupadas en varios niveles donde las conexiones pueden estar solas en una capa, en varias o en todas, por lo que cabe señalar los dos tipos básicos de redes multicapa: Feedforward: donde todas las señales neuronales se propagan hacia delante a través de las capas de la red.

Feedforward/feedback: en este tipo de redes circula información tanto hacia delante como hacia atrás durante el funcionamiento de la red. Algunas de las implementaciones más conocidas son: Learning Vector Quantizae (LVQ) (Kohonen, 2001); Topology Preserving Map (TPM) o S.O.M. (Kohonen, 2001); Temporal Associative Memory (Wan et al., 1990).

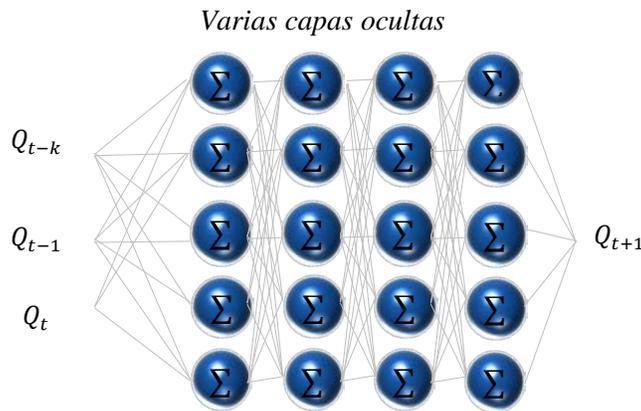


Figura. A.3. Red multicapa

Apéndice B

B1 Quasi-Newton Backpropagation BFGS

Este algoritmo se llama así porque se asemeja mucho al método de Newton, la idea principal es la aproximación de la matriz Hessiana mediante un proceso iterativo. Tomando en cuenta que el método de Newton calcula la búsqueda de dirección como.

$$r_k = -H_f(w_k)^{-1} \nabla f(w_k) \quad B1$$

Donde $H_f(w_k)$ debe ser definida positiva. Con el fin de evitar problemas con que no esté definida positiva o que no se pueda invertir, los métodos quasi-Newton aproximan la matriz Hessiana $H_f(w_k)$ mediante la matriz B_k utilizando una fórmula de actualización $B_{k+1} = B_k + U_k$.

La actualización debe ser tal que en cada paso la nueva información de curvatura se construye en la aproximación Hessiana. Usando la expansión de segundo orden de Taylor de la función f se tiene.

$$T(w_k + r) \approx f(w_k) + \nabla f(w_k)^T r + \frac{1}{2} r^T H_f(w_k) r \quad B2$$

De esta relación se obtiene

$$\nabla f(w_k + r) \approx T(w_k + r) = \nabla f(w_k) + H_f(w_k) r \quad B3$$

Si se toma $r = r_k$ y se denota a $y_k = \nabla f(w_{k+1}) - \nabla f(w_k)$ se obtiene

$$y_k \approx H_f(w_k) r_k \quad B4$$

Esta ecuación es llamada condición de Quasi-Newton esto es $y_k = B_k r_k$ esto es valido para todos los B_k y para cada búsqueda de dirección $r_k = w_{k+1} - w_k$

La actualización de la matriz $B_{k+1} = B_k + \alpha_k u_k u_k^T$ esto se requiere definir como

$$B_{k+1} = B_k + \frac{1}{(y_k - B_k r_k)^T r_k} (y_k - B_k r_k)(y_k - B_k r_k)^T \quad B5$$

Esto se denomina rango simétrico, en general después de actualiza la aproximación de la aproximación de la matriz Hessiana se debe obtener su inversa, para poder calcular la nueva dirección de búsqueda, afortunadamente esto se puede hacer utilizando la fórmula de Sherman-Morrison, formula que puede actualizar directamente la matriz inversa, se denota de la siguiente forma

$$M_k = B_k^{-1} \quad B6$$

$$M_{k+1} = M_k + \frac{1}{(r_k - M_k y_k)^T y_k} (r_k - M_k y_k)(r_k - M_k y_k)^T \quad B7$$

Más tarde, Broyden-Fletcher-Goldfarb-Shanno BFGS, (Battiti et al,1990) método que fue descubierto por Broyden, Fletcher, Goldfarb, y Shanno cada uno de forma independiente, alrededor de 1970, las fórmulas actualizadas son.

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T r_k} - \frac{B_k r_k r_k^T B_k}{r_k^T B_k r_k} \quad B8$$

$$M_{k+1} = M_k + \frac{(r_k - M_k y_k)(r_k - M_k y_k)^T}{y_k^T r_k} - \frac{M_k y_k y_k^T M_k}{y_k^T r_k} + \frac{y_k^T M_k y_k r_k r_k^T}{(y_k^T r_k)^2} \quad B9$$

$$r_k = -M_k \nabla f(w_k) \quad B10$$

$$w_{k+1} = w_k + \lambda r_k \quad B11$$

B2 Scaled Conjugate Gradiante Backpropagation

Este método también puede ser visto como una modificación del método del paso descendente, donde el fin es evitar el zigzag de la búsqueda, en cada iteración la dirección de la búsqueda es modificada por una combinación de las direcciones anteriores.

$$r_k = -\nabla f_k + B_k r_{k-1} \quad B12$$

Las correcciones que se hacen aseguran que $r_1, r_2, r_3, \dots, r_n$ son las llamadas direcciones conjugadas, lo cual indica que existe una matriz A tal que $r_i^T A_j = 0, \forall i \neq j$.

Por ejemplo las direcciones (vectores unitarios) son conjugadas, tomando la matriz A como la matriz unitaria, la idea radica en que A es la inversa de la matriz Hessiana, con ello se puede deducir que si se busca en la línea exacta se puede alcanzar el óptimo en la mayoría de los n pasos de las funciones cuadráticas.

Una vez que se tiene la dirección r_k la siguiente iteración se calcula como sigue

$$w_{k+1} = w_k + \lambda r_k \quad B13$$

Donde λ es la longitud del paso óptimo o su aproximación, el parámetro B_k puede ser calculado usando diferentes fórmulas, Hestenes y Stiefel (1952) sugieren.

$$B_k = \frac{\nabla f_k^T (\nabla f_k - \nabla f_{k-1})}{r_k^T (\nabla f_k - \nabla f_{k-1})} \quad B14$$

Después, Fletcher y Reeves (1964) examinaron

$$B_k = \frac{\|\nabla f_k\|^2}{\|\nabla f_{k-1}\|^2} \quad B15$$

Y finalmente la fórmula de Polak and Ribière (1969) es

$$B_k = \frac{\nabla f_k^T (\nabla f_k - \nabla f_{k-1})}{\|\nabla f_{k-1}\|^2} \quad B16$$

Estas fórmulas están basadas en el caso cuadrático donde $f(x) = \frac{1}{2}x^T Ax + b^T x + c$ para A positiva. Para esta función es obtener direcciones conjugadas $-A$ para $r_i^T A_j, \forall i \neq j$, en $r_k^T A r_{k-1} = 0$ dado $-\nabla f_k^T A r_{k-1} + B_k r_{k-1}^T A r_{k-1} = 0$ de tal manera que

$$B_k = \frac{\nabla f_k^T A r_{k-1}}{r_{k-1}^T A r_{k-1}} \quad B17$$

Ahora con $\nabla f(x) = Ax + b$ obteniendo $\nabla f(x_k) = A(x_{k-1} + \lambda r_{k-1}) + b = \nabla f_{k-1} + \lambda A r_{k-1}$ así que $\nabla f_k = \nabla f_{k-1} + \lambda A r_{k-1}$ por lo tanto

$$B_k = \frac{\nabla f_k^T A r_{k-1}}{r_{k-1}^T A r_{k-1}} = \frac{\nabla f_k^T (\nabla f_k - \nabla f_{k-1})}{r_{k-1}^T (\nabla f_k - \nabla f_{k-1})} \quad B18$$

Esta es exactamente la fórmula de Hestenes y Stiefel. De hecho, para el caso cuadrático las tres fórmulas son iguales, y el óptimo se encuentra en algunos pasos.

B3 One Step Secant Backpropagation

El término de un solo paso secante proviene del hecho de que las derivadas son aproximadas por secantes evaluadas en dos puntos de la función (en este caso la función es el gradiente). La ventaja de este método se menciona en Battiti (1992, 1994).

La factibilidad de este método es que la actualización en la dirección de búsqueda dirección (Ec. 3.29) es calculada basada solo en los vectores determinados por el gradiente, la nueva dirección de búsqueda r_{i+1} se obtiene como sigue:

$$r_k = -\nabla f_k + A_k s_k + B_k q_k \quad B19$$

$$s_k = \theta_{k+1} - \theta_k = p_k \quad B20$$

$$A_k = - \left[1 + \frac{q_k^T q_k}{s_k^T q_k} \right] \frac{s_k^T g_k}{s_k^T q_k} + \frac{q_k^T g_k}{s_k^T q_k}; \quad B_k = \frac{s_k^T g_k}{s_k^T q_k} \quad B21$$

Los vectores q_k y p_k son determinados como sigue

$$q_k = \nabla f_{k+1} - \nabla f_k = \nabla^2 J(\theta) p_k \quad B22$$

$$p_k = \alpha_k r_k \quad B23$$

$$w_{k+1} = w_k + \lambda r_k \quad B24$$

B4 Resilient Backpropagation

El algoritmo RBP fue introducido por M. Riedmiller en 1993 es un sistema de búsqueda local de aprendizaje adaptativo. El principio básico de RBP es eliminar la influencia del tamaño de la derivada parcial en cada etapa del peso, como consecuencia de ello solo se considera el signo de la derivada para indicar la dirección de la actualización del peso, este algoritmo introduce la actualización para cada peso w_{ij} de forma individual, esto es determinando el tamaño de la actualización Δ_{ij} del peso (Riedmiller & Braun, 1993).

Esto se lleva a cabo introduciendo una segunda regla de aprendizaje, que determina la evolución del valor de actualización Δ_{ij} esta estimación se basa en el comportamiento observado de la derivada parcial durante dos pasos sucesivos de la actualización del peso

$$\Delta_{ij}(t) = \left\{ \begin{array}{l} \eta^+ \Delta_{ij}(t+1), \text{ si } \frac{\partial E}{\partial w_{ij}}(t) \cdot \frac{\partial E}{\partial w_{ij}}(t-1) > 0 \\ \eta^- \Delta_{ij}(t+1), \text{ si } \frac{\partial E}{\partial w_{ij}}(t) \cdot \frac{\partial E}{\partial w_{ij}}(t-1) < 0 \\ \Delta_{ij}(t+1), \text{ en otro caso} \end{array} \right\} \quad B25$$

Donde

$$0 < \eta^- < 1 < \eta^+$$

La regla de adaptación es como sigue

Cada vez la derivada parcial del correspondiente peso w_{ij} cambia de signo lo que indica que la ultima actualización involucraba saltos demasiados grandes esto provocaba saltos de más de un mínimo local, lo que se solución mediante el factor η^- . Si la derivada conserva su signo, la actualización del valor aumenta ligeramente con el fin de acelerar la convergencia en regiones poco profundas.

Una vez que se actualiza cada peso, la actualización siguiente sigue una regla sencilla, si la derivada es positiva (incremento de error), el peso es decrementado en su actualización, si la derivada es positiva se incrementa el peso en la actualización.

$$\Delta_{ij}(t) = \begin{cases} -\Delta_{ij}(t), & \text{si } \frac{\partial E}{\partial w_{ij}}(t) > 0 \\ +\Delta_{ij}(t), & \text{si } \frac{\partial E}{\partial w_{ij}}(t) < 0 \\ 0, & \text{de lo contrario} \end{cases} \quad B26$$

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}(t) \quad B27$$

Sin embargo, hay una excepción: Si la derivada parcial cambia de signo, es decir, el paso anterior era demasiado grande y el mínimo se ha perdido, la anterior actualización del peso es revertido:

$$\Delta w_{ij}(t) = -\Delta w_{ij}(t - 1) \quad B28$$

$$\text{si } \frac{\partial E}{\partial w_{ij}}(t) \cdot \frac{\partial E}{\partial w_{ij}}(t - 1) < 0 \quad B29$$

Debido a que el peso se actualiza en cada paso, la derivada cambiaría su signo, una vez más en la siguiente el paso.

Con el fin de evitar un castigo en actualización, no debería haber ninguna adaptación de la actualización del valor en la etapa subsiguiente. En la práctica se termina de ajustar cuando.

$$\frac{\partial E}{\partial w_{ij}}(t - 1) = 0 \quad B30$$

LISTA DE FIGURAS

i	Diagrama de la tesis.	iv
2.1	Arquitectura de MFNN típica con el algoritmo de aprendizaje de retropropagación. .	15
2.2	Superficie de una función error (E) en dos parámetros.	17
3.1	Estructura de la red ensamblada.	31
3.2	Esquema del algoritmo AdaBoost.M2.	32
3.3	Esquema del algoritmo LPBoost.	33
3.4	Esquema del algoritmo TotalBoost.	34
3.5	Esquema del algoritmo RUSBoost.	36
4.1	Arquitectura del modelo Modular con el algoritmo de aprendizaje de retropropagación.	42
6.1	Diagrama de flujo del modelo Regional.	63
6.2	Ubicación de las estaciones climatológicas de estudio en Australia.	65
6.3	Agrupación jerárquica de las regiones homogéneas.	70
6.4	Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007027.	72
6.5	Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007064.	73
6.6	Lluvia mensual estimada mediante los modelos RNAs y RNAs- Regional en el periodo de prueba— Estación ASN00007197.	74
6.7	Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007027.	76
6.8	Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007064.	77
6.9	Lluvia mensual estimada mediante los modelos Ensamblado y Ensamblado-Regional en el periodo de prueba— Estación ASN00007197.	78
6.10	Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007027.	80
6.11	Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007064.	81
6.12	Lluvia mensual estimada mediante los modelos Modular y Modular-Regional en el periodo de prueba— Estación ASN00007197.	82
6.13	Localización de las estaciones hidrométricas en el sureste y suroeste de Canadá. . . .	83
6.14	Localización de los sitios que conforman cada grupo delimitado mediante c-Medias	86
6.15	L-Radios de los Momentos-L de los 10 sitios en la región 9.	87
6.16	Ajuste de la distribución Kappa a ocho sitios de la región 9.	88
6.17	Localización de los sitios en cada grupo de estaciones en Canadá mediante c-Medias.	88
6.18	Estimación mensual de escurrimientos mediante el modelo RNAs y RNAs-Regional en el periodo de prueba—Estación Spray river at banff (2BC001).	92
6.19	Estimación mensual de escurrimientos mediante el modelo RNAs y RNAs-Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).	93

6.20	Estimación mensual de escurrimientos mediante el modelo Ensamblado y Ensamblado Regional en el periodo de prueba—Estación Spray river at banff (2BC001).	97
6.21	Estimación mensual de escurrimientos mediante el modelo Ensamblado y Ensamblado Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).	97
6.22	Estimación mensual de escurrimientos mediante el modelo Modular y Modular Regional en el periodo de prueba—Estación Spray river at banff (2BC001).	101
6.23	Estimación mensual de escurrimientos mediante el modelo Modular y Modular Regional en el periodo de prueba—Estación Bow river near seebe (02BE004).	102
6.24	Distribución de las estaciones climatológicas en la cuenca del Amazonas.	103
6.25	Regiones homogéneas en la cuenca del Amazonas.	106
6.26	Dendograma de los grupos formados en el análisis clúster.	107
6.27	Ajuste de la distribución Kappa al grupo homogéneo 1.	108
6.28	Ajuste de la distribución Kappa al grupo homogéneo 2.	109
6.29	Ajuste de la distribución Kappa al grupo homogéneo 3.	109
6.30	Ajuste de la distribución Kappa al grupo homogéneo 4.	110
6.31	Pronostico mensual de escurrimiento en la estación 14710000.	111
6.32	Pronostico mensual de escurrimiento en la estación 15275100.	111
6.33	Pronostico mensual de escurrimiento en la estación 15860000.	112
6.34	Pronostico mensual de escurrimiento en la estación 10099800.	112
6.35	Pronostico mensual de escurrimiento en la estación 10100000.	113
6.36	Pronostico trimestral de escurrimiento en la estación 14710000.	114
6.37	Pronostico trimestral de escurrimiento en la estación 15275100.	114
6.38	Pronostico trimestral de escurrimiento en la estación 15860000.	115
6.39	Pronostico trimestral de escurrimiento en la estación 10099800.	115
6.40	Pronostico trimestral de escurrimiento en la estación 10100000.	116
6.41	Pronostico semestral de escurrimiento en la estación 14710000.	117
6.42	Pronostico semestral de escurrimiento en la estación 15275100.	117
6.43	Pronostico semestral de escurrimiento en la estación 15860000.	118
6.44	Pronostico semestral de escurrimiento en la estación 10099800.	118
6.45	Pronostico semestral de escurrimiento en la estación 10100000.	119
6.46	Estaciones climatológicas en la Costa Oeste de Estados Unidos.	120
6.47	Formación de grupos homogéneos la Costa Oeste de EEUU.	124
6.48	Pronostico mensual de lluvia acumulada en la estación USC00042239.	125
6.49	Pronostico mensual de lluvia acumulada en la estación USC00042805.	125
6.50	Pronostico mensual de lluvia acumulada en la estación USC00043452.	126
6.51	Pronostico mensual de lluvia acumulada en la estación USC00046719.	126
6.52	Pronostico mensual de lluvia acumulada en la estación USC00450008.	127
6.53	Pronostico mensual de lluvia acumulada en la estación USC00451233.	128
6.54	Pronostico mensual de lluvia acumulada en la estación USC00451276.	128
6.55	Pronostico mensual de lluvia acumulada en la estación USC00451350.	129
6.56	Pronostico mensual de lluvia acumulada en la estación USC00451496.	129
6.57	Pronostico mensual de lluvia acumulada en la estación USC00451504.	130
6.58	Pronostico trimestral de lluvia acumulada en la estación USC00042239.	131

6.59	Pronostico trimestral de lluvia acumulada en la estación USC00042805.....	131
6.60	Pronostico trimestral de lluvia acumulada en la estación USC00043452.....	132
6.61	Pronostico trimestral de lluvia acumulada en la estación USC00046719.....	132
6.62	Pronostico trimestral de lluvia acumulada en la estación USC00450008.....	133
6.63	Pronostico trimestral de lluvia acumulada en la estación USC00451233.....	134
6.64	Pronostico trimestral de lluvia acumulada en la estación USC00451276.....	134
6.65	Pronostico trimestral de lluvia acumulada en la estación USC00451350.....	135
6.66	Pronostico trimestral de lluvia acumulada en la estación USC00451496.....	135
6.67	Pronostico trimestral de lluvia acumulada en la estación USC00451504.....	136
6.68	Pronostico semestral de lluvia acumulada en la estación USC00042239.....	137
6.69	Pronostico semestral de lluvia acumulada en la estación USC00042805.....	137
6.70	Pronostico semestral de lluvia acumulada en la estación USC00043452.....	138
6.71	Pronostico semestral de lluvia acumulada en la estación USC00046719.....	138
6.72	Pronostico semestral de lluvia acumulada en la estación USC00450008.....	139
6.73	Pronostico semestral de lluvia acumulada en la estación USC00451233.....	140
6.74	Pronostico semestral de lluvia acumulada en la estación USC00451276.....	140
6.75	Pronostico semestral de lluvia acumulada en la estación USC00451350.....	141
6.76	Pronostico semestral de lluvia acumulada en la estación USC00451496.....	141
6.77	Pronostico semestral de lluvia acumulada en la estación USC00451504.....	142
6.78	Estaciones climatológicas en Zacatecas.....	143
6.79	Agrupamiento de las estaciones climatológicas del acumulado anual de lluvia en el estado de Zacatecas.....	147
6.80	Formación de grupos homogéneos mediante ML en Zacatecas.....	148
6.81	Pronostico mensual de lluvia acumulada en la estación 32001.....	149
6.82	Pronostico mensual de lluvia acumulada en la estación 32002.....	150
6.83	Pronostico mensual de lluvia acumulada en la estación 32017.....	150
6.84	Pronostico mensual de lluvia acumulada en la estación 32020.....	151
6.85	Pronostico mensual de lluvia acumulada en la estación 32073.....	151
6.86	Pronostico mensual de lluvia acumulada en la estación 32124.....	152
6.87	Pronostico mensual de lluvia acumulada en la estación 32137.....	152
6.88	Pronostico mensual de lluvia acumulada en la estación 32142.....	153
A.1	Modelo neuronal de McCulloch y Pitts.....	158
A.2	Red monocapa.....	160
A.3	Red multicapa.....	160

LISTA DE TABLAS

5.1	Valores críticos de D_i para la prueba de discordancia (Hosking & Wallis, 1997).	52
6.1	Medidas de eficiencia de los modelos.	64
6.2	Ubicación de las estaciones climatológicas estudiadas en Australia.	65
6.3	Relaciones de los radios Momentos-L, estadístico de discordancia D_i y medidas de homogeneidad H Y H^* de las diferentes regiones homogéneas formadas.	68
6.4	RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs-Regional para las regiones estudiadas (periodo de prueba).	71
6.5	RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs-Regional para las regiones estudiadas (periodo de prueba).	75
6.6	RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs-Regional para las regiones estudiadas (periodo de prueba).	79
6.7	Información de estaciones hidrométricas en el sureste y suroeste de	83
6.8	Transformación de las características de los sitios.	86
6.9	Radio de los Momentos-L y estadístico de discordancia D_i para escurrimientos en la región 9.	87
6.10	Resumen de las regiones finales mediante Momentos-L.	89
6.11	RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs-Regional para las regiones estudiadas (periodo de prueba).	89
6.12	RMSE, MAE e índice de concordancia d de los modelos RNAs y RNAs-Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).	92
6.13	RMSE, MAE e índice de concordancia d de los modelos Ensamblado y Ensamblado Regional para las regiones estudiadas (periodo de prueba). . .	94
6.14	RMSE, MAE e índice de concordancia d de los modelos Ensamblado y Ensamblado Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).	96
6.15	RMSE, MAE e índice de concordancia d de los modelos Modular y Modular Regional para las regiones estudiadas (periodo de prueba).	98
6.16	RMSE, MAE e índice de concordancia d de los modelos Modular y Modular Regional para las estaciones 05BC001 y 05BE004 (periodo de validación y prueba).	101
6.17	Estaciones hidrométricas en la Cuenca del Amazonas.	104
6.18	Resultados por clase.	107
6.19	Distancias entre los objetos centrales:	107
6.20	Grupos de estaciones utilizadas.	108
6.21	Comparación del modelo Ensamblado regional y tradicional vía redes neuronales en la región 2.	110
6.22	Comparación del modelo Ensamblado regional y tradicional vía redes neuronales en la región 2.	113

6.23	Comparación del modelo Ensamblado regional y tradicional vía redes neuronales en la región 2.	116
6.24	Estaciones climatológicas en la Costa Oeste de EEUU.	121
6.25	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A.	124
6.26	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B.	127
6.27	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A.	130
6.28	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B.	133
6.29	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región A.	136
6.30	Comparación de los modelos Ensamblado regional y tradicional vía redes neuronales en la región B.	139
6.31	Estaciones climatológicas en Zacatecas.	144
6.32	Comparación de los modelos RNAs, Modular y Ensamblado regionales en la región 1.	149
A1	Resumen de funciones de red.	159
A2	Funciones de activación neuronal.	159

LISTA DE ACRÓNIMOS

BP	retropropagación (Backpropagation BP)
CCA	Canonical Correlation Analysis
ENN	Ensemble Neural Network
FFMLPs	Feed forward-Multilayer Perceptrons
FFN	Feed Forward Nerwork
IA	Inteligencia Artificial
LM	Levenberg Marquardt
LPBoost	linear programming boosting
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MFNN	Multilayer Feedforward Neural Network
ML	Momentos-L
MLP	Multilayer Perceptrons
MNN	Modular Neural Network
MSE	Mean Square Error
OSSB	One Step Secante Backpropagation
QNBFGS	Quasi-Newton Backpropagation BFGS
RB	Resilient Backpropagation
RBF	red de función de base radial
RMSE	Root Mean Sqare Error
RNAM	Redes Neuronales Artificiales Modulares
RNAs	Redes Neuronales Artificiales
ROI	Región Of Influence
RUSBoost	Random Undersampling Boosting
SCGB	Scaled Conjugate Gradiente Backpropagation
SVM	Support Vector Machine
TotalBoost	Total Boosting

CODIGOS

Redes neuronales

%El siguiente código describe la rutina usada en esta tesis para pronóstico de lluvia y %escurrimiento mediante redes neuronales artificiales (RNAs) en MATLAB R12. La variable en se refiere a la %entrada de los datos. Las variables salidas1 y salidas2 se refieren a la salida de los datos. %Finalmente la variable 'valor' se refiere al valor pronosticado.

```
clear en;clear salidas1;clear salidas2;clear valor;clear ee;
e=e1;n=length(e);ls=length(e);minimo=min(e);maximo=max(e);
```

```
tam=10;pron=1;
    long=ls-tam;nu=1;clear valor;
        en(1:long)=e(1:long);
            ene(long+1:n)=e(long+1:n);
                en=en;valor(long)=e(long);
                    ene(long)=e(long);
```

```
screensz = get(0,'ScreenSize');f=figure('Position',[1 screensz(4)/2 screensz(3)/2 screensz(4)/2.5]);whitebg([1 1 1]);set(gcf,'Color',[1,1,1]);plot(e,'k-o','LineWidth',2,'MarkerFaceColor','k','MarkerSize',10);
axis([long, n, minimo, maximo]);title('Pronostico de acumulados mensuales','FontWeight','bold','FontSize',25)
    xlabel('Tiempo (años)', 'FontWeight','bold','FontSize',25,'LineWidth',4);
    ylabel('Acumulado mensual', 'FontWeight','bold','FontSize',25);
set(gca,'yTick',(minimo:(maximo-minimo)/10:máximo),'YLim',[minimo,máximo],'Box','on','YGrid','on','FontSize',20);minimo)/10:maximo),'YLim',[minimo,maximo],'Box','on','YGrid','on','FontSize',20);
Grid on;
set(gca,'Xcolor',[0.5 0.5 0.5]);set(gca,'Ycolor',[0.8 0.8 0.8]);Caxes = copyobj(gca,gcf);
set(Caxes, 'color', 'none', 'xcolor', 'k', 'xgrid', 'off', 'ycolor', 'k', 'ygrid', 'off');
```

```
hold on;
```

```
        for i=long+1:pron:n
            for m=long+1:(i-1)
                ene(m)=e(m);
            end
        end
        for k=1 :1: pron
```

```
T = num2cell(en');
    capa=[floor(length('e')/12:);
```

```
        net = narnet(1:1,capa);
            net = init(net)
                net.biasConnect = [0;0];
                    net.layers{1}.transferFcn = 'tansig';net.layers{2}.transferFcn = 'logsig';
                        net.trainFcn = 'trainlm'; [Xs,Xi,Ai,Ts] = preparets(net,{},{},T);
                            %net.divideParam.trainRatio = 1;
                                %net.divideParam.valRatio = 0.1;
```

Gráficas

RNAs

RNAs

```

%net.divideParam.testRatio = 0.1;
    net.divideFcn = "";
    net.trainParam.epochs=1000; %Maximum number of epochs to train
    net.trainParam.goal=1; %Performance goal
    net.trainParam.min_grad=0;%Minimum performance gradient
    net.trainParam.mu_max=10e100;net.trainParam.max_fail=1000;

```

```

        if i>long+1
            net = configure(net,T);
            net.IW = pe1;net.LW = pe2;
        end

```

```
net = train(net,Xs,Ts,Xi,Ai); pe1=net.IW;pe2=net.LW;
```

```
Y = net(Xs,Xi);perf = perform(net,Ts,Y);
```

```
en(i)=i; T1 = num2cell(en');
```

```

    salidas1=sim(net,T1);salidas1=cell2mat(salidas1);T1=cell2mat(T1);salidas2=salidas1';
    en(i)=e(i); %pronostico de un dato; en(i)=salidas2(i);%pronostico del pronostico

```

Pronóstico

```

valor(i)=salidas2(i);hold on;plot(valor','b -o','LineWidth',2.5);legend('Serie real','Pronostico Real');
i=i+1;
end
end
ee=(sum((valor(long+1:ls) -e(long+1:ls)).^2)).^0.5;
text(ls-tam/1.005,máximo(maximo-minimo)/20,['EEA=',num2str(ee)],'FontSize',15,.....
'EdgeColor','red')
valor'

```

Medida de discordancia D

%El siguiente código describe la rutina usada para obtener la medida de discordancia D en %MATLAB R12. La variable r es la matriz que corresponde a un grupo posiblemente homogéneo. %Finalmente la variable D se refiere al valor calculado.

```

clear b0;clear b1;clear b2;clear b3;clear PK;clear P;
clear xord;clear xi;clear x;clear f;clear D;clear B;clear ue;clear u;clear A;
global xord P g w reg1 reg3 reg4 x t t3 t4 x;

```

```
%tam=8;r(1:length(s1),1:tam)=[s1,s2,s3,s4,s5,s6,s7,s8];
```

```
[m,tam]=size(r);s1=r(:,1);
```

```

for k=1:tam;
    i(1:length(s1),k)=(1:length(s1));
end;

```

```
xi= mean(r)';desvest=std(r);media=mean(r);
```

```
b0=mean(r);
```

```

b1=(1/(length(s1)*(length(s1)-1)))*sum(r(1:length(s1)-1,1:tam).*(length(s1)-i(1:length(s1)-1,1:tam)));
b2=(1/(length(s1)*(length(s1)-1)*(length(s1)-2)))*sum(r(1:length(s1)-2,1:tam).*(length(s1)-
i(1:length(s1)-2,1:tam)).*(length(s1)-1-i(1:length(s1)-2,1:tam)));
b3=(1/(length(s1)*(length(s1)-1)*(length(s1)-2)*(length(s1)-3)))*sum(r(1:length(s1)-
3,1:tam).*(length(s1)-i(1:length(s1)-3,1:tam)).*(length(s1)-1-i(1:length(s1)-3,1:tam)).*(length(s1)-2-
i(1:length(s1)-3,1:tam)));

```

```

l1=b0;l2=2*b1-b0;l3=6*b2-6*b1+b0;l4=20*b3-30*b2+12*b1-b0;t1r=l2./l1;t3r=l3./l2;t4r=l4./l2;

```

```

u=[t1r;t3r;t4r];% ue=[mean(t1r);mean(t3r);mean(t4r)]
ue=[mean(u)'];

```

```

A=0;
for k=1:tam;
    A=(u(:,k)-ue)*(u(:,k)-ue)'+A;
end;
B=inv(A);
for k=1:tam;
    D(k)=(tam/3)*(u(:,k)-ue)'*B*(u(:,k)-ue);
end;
D=D'

```

Momentos-L

%El siguiente código describe la rutina usada para el cálculo de los ratios Momentos-L en MATLAB
 %R12. La variable r es la matriz que corresponde a un grupo posiblemente %homogéneo.
 %Finalmente las variables t1r, t2r y t3r se refiere a los ratios.

```

clear b0;clear b1;clear b2;clear b3;clear PK;clear P;clear s1;clear Ln
clear xord;clear xi;clear x;clear f;clear rord;
global xord P g w reg1 reg3 reg4 x t t3 t4 x t3r t4r;

```

```

% tam=3;r(1:length(s1),1:tam)=[s1,s2,s3];
[m,tam]=size(r);s1=r(1:m,1);

```

```

for k=1:tam;
    i(1:length(s1),k)=(1:length(s1));
end;

```

```

xi= mean(r)';desvest=std(r);media=mean(r);rord=sort(r,'descend');

```

```

b0=mean(r);
b1=(1/(m*(m-1)))*sum(rord(1:m-1,1:tam).*(m-i(1:m-1,1:tam)));
b2=(1/(length(s1)*(length(s1)-1)*(length(s1)-2)))*sum(rord(1:length(s1)-2,1:tam).*(length(s1)-
i(1:length(s1)-2,1:tam)).*(length(s1)-1-i(1:length(s1)-2,1:tam)));
b3=(1/(length(s1)*(length(s1)-1)*(length(s1)-2)*(length(s1)-3)))*sum(rord(1:length(s1)-
3,1:tam).*(length(s1)-i(1:length(s1)-3,1:tam)).*(length(s1)-1-i(1:length(s1)-3,1:tam)).*(length(s1)-2-
i(1:length(s1)-3,1:tam)));

```

```

l1=b0;l2=2*b1-b0;l3=6*b2-6*b1+b0;l4=20*b3-30*b2+12*b1-b0;t1r=l2./l1;t3r=l3./l2;t4r=l4./l2;

```