



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL**

**PRONÓSTICO DE CRECIDAS EN TIEMPO REAL USANDO REDES
NEURONALES RECURRENTES**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

MILENA MUÑOZ BONACIC

**PROFESOR GUÍA:
XIMENA VARGAS MESA**

**MIEMBROS DE LA COMISIÓN:
JAMES MCPHEE TORRES
RICARDO GONZÁLEZ VALENZUELA**

**SANTIAGO DE CHILE
ENERO 2008**

RESUMEN DE LA MEMORIA
PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL
POR: MILENA MUÑOZ BONACIC
FECHA: 24/01/2008
PROF. GUÍA: XIMENA VARGAS

“PRONÓSTICO DE CRECIDAS EN TIEMPO REAL USANDO REDES NEURONALES RECURRENTES”

En el contexto socio económico actual de Chile la mayoría de los sistemas tradicionales de control de crecidas son reactivos y resulta imprescindible, en respuesta a un nivel recurrente sin prevenciones claras, ofrecer soluciones a los problemas asociados a fenómenos de crecidas e inundaciones, cambiando el modelo de acción defensiva por un modelo de acción proactiva.

En los últimos años se han realizado estudios de pronóstico de crecidas más adecuados a la realidad dentro del campo de las redes neuronales artificiales, mayormente prealimentadas. Continuando los estudios, en esta memoria se utilizan redes neuronales de tipo recurrente, redes dinámicas que liberan el supuesto de que los parámetros característicos de la cuenca sean estacionarios, capaces de modelar la alta NO linealidad del fenómeno y almacenar la historia relevante de la crecida.

Se aplican específicamente para el pronóstico de crecidas en la parte alta de la cuenca del río Mapocho, perteneciente a la hoya hidrográfica del río Maipo. Las variables relevantes son definidas a partir de registros de caudales, precipitaciones y temperaturas en tiempo horario. Posteriormente, se elige como red neuronal óptima aquella que entrega el menor error en el caudal pronosticado, cuya arquitectura, configuración y tipo de entrenamiento constituye un modelo de pronóstico.

Se concluye la eficiencia y menor complejidad de considerar una crecida con amplias componentes de frecuencia como set de datos de entrenamiento, en comparación a un set de crecidas “pegadas”, independiente del tipo de red neuronal utilizada y marcadamente en el caso de una red neuronal recurrente.

Los pronósticos de caudal obtenidos mediante redes neuronales recurrentes de Elman son de buena calidad tanto en valor como en el tiempo de ocurrencia. Se obtienen errores menores que para otros modelos de pronóstico como son las redes recurrentes de Hopfield, las redes prealimentadas y los modelos paramétricos.

En el caso de crecidas pluviales, en la cuenca del Estero Arrayán en La Montosa se obtienen pronósticos certeros hasta con 12 horas de anticipación y en la cuenca del río Mapocho en Los Almendros, se recomienda realizar pronósticos sólo hasta con 6 horas de anticipación.

En el caso de crecidas de deshielo, en ambas cuencas es posible obtener pronósticos certeros con más de 24 horas de anticipación.

Dedicada a mi Madre y a mis Nonos.

Milena.

Agradecimientos

Deseo expresar mi profundo cariño y agradecimiento a todas aquellas personas que durante los años de estudio de mi carrera fueron un apoyo y posibilitaron la realización de mi memoria.

Especialmente a mi Madre por todo su amor y apoyo incondicional en éste y en todos los proyectos de mi vida .A mis Abuelos, Nono y Nona, que si bien hoy no están presentes siempre me apoyaron con su cariño y sabios consejos.

Agradezco también a mi querido Robe, por todo su amor y compañía a lo largo de cada día, por apoyarme y aconsejarme en la realización de esta memoria.

Además, quiero agradecer a todos los profesores que me han formado en mis distintas etapas como estudiante, que además de formarme como profesional, me enseñaron a perseverar y a no desanimarme ante los problemas. En especial a la Profesora Sra. Ximena Vargas, por brindarme la oportunidad de realizar esta memoria.

Agradecer por supuesto a mis queridas amigas, Marcela, Ángela y Loreto por acompañarme siempre incondicionalmente. A mis amigos Pedro Sanzana, Camilo Gatica, Rodrigo Arbuch y Mauricio Charmín por toda su compañía, discusiones y aportes en que desarrollamos cada trabajo y por supuesto por todas aquellas risas compartidas que alivianaban los días.

A todos quienes me animaron y me han ayudado a llevar de una u otra forma a buen fin mi carrera.

Tabla de Contenidos

1	Introducción.....	1
1.1	Alcances y Objetivos	3
1.2	Contribuciones de la Memoria	3
1.3	Organización de la Memoria	4
2	Antecedentes Generales.....	5
2.1	Redes Neuronales	5
2.1.1	Redes Neuronales Recurrentes - Definiciones.....	6
2.1.2	Propiedades Generales de las Redes Neuronales.....	7
2.1.3	Modelos neuronales	8
2.2	Redes Neuronales Recurrentes	12
2.2.1	Propiedades y Características	12
2.2.2	Aplicación de las Redes neuronales Recurrentes al Procesamiento de secuencias	14
2.2.3	Modelos Neuronales Recurrentes Tradicionales	16
2.2.3.1	Redes de Primer Orden	16
2.2.3.2	Redes de Segundo Orden	19
2.2.3.3	Adición de una ventana temporal.....	20
2.2.3.4	Algoritmos de Entrenamiento	20
2.2.4	Red de Hopfield.....	29
2.2.4.1	Regla de Aprendizaje	33
2.3	Selección de Modelos y configuraciones de Redes.....	36
2.4	Identificación de Sistemas mediante Modelos Paramétricos	38
2.4.1	Algoritmos de ajuste de parámetros	40
2.5	Metodología General	42
2.5.1	Metodología de Pronóstico.....	42
2.5.2	Condición de aplicación	45
2.6	Análisis espectral de crecidas	50
2.6.1	Transformada de Fourier	50
2.7	Índices de Evaluación de desempeño	51
2.7.1	Índices de desempeño general	51
2.7.2	Índices para la Evaluación de la Calidad de un Modelo de Pronóstico.....	52
2.7.3	Índices de Comparación entre Modelos de Pronóstico.....	53
3	Caracterización de la Cuenca	54

3.1	Generalidades	54
3.1.1	Subcuenca Alta del Mapocho	54
3.1.2	Subcuenca Baja del Mapocho.....	54
3.2	Control de estaciones.....	55
3.2.1	Estación Fluviográfica Estero Arrayán en la Montosa	55
3.2.2	Estación Fluviográfica Mapocho en Los Almendros	56
3.2.3	Estación Pluviográfica Cerro Calán.....	58
3.2.4	Estación Meteorológica Tobalaba	59
3.2.5	Curvas Hipsométricas.....	59
4	Pronóstico de Crecidas	62
4.1	Introducción.....	62
4.2	Planteamiento General.....	63
4.3	Aplicación a Crecidas Pluviales	66
4.3.1	Cuenca del Estero Arrayán en La Montosa	66
4.3.1.1	Pronósticos mediante Redes Neuronales de pronóstico - Entrenamiento Unitario....	68
4.3.1.2	Pronóstico mediante Redes Neuronales de pronóstico -Entrenamiento Continuo.....	86
4.3.1.3	Comparación de modelos neuronales de pronóstico por crecidas.....	92
4.3.1.4	Pronóstico mediante Modelos Paramétricos	95
4.3.1.5	Pronóstico mediante Red Neuronal Recurrente de Hopfield-Identificación.....	98
4.3.1.6	Resumen y análisis de modelos de pronóstico a un paso de adelanto.....	102
4.3.1.7	Pronósticos a varios pasos de adelanto	106
4.3.2	Cuenca Mapocho en Los Almendros.....	112
4.3.2.1	Pronósticos mediante Redes Neuronales de pronóstico - Entrenamiento Unitario..	114
4.3.2.2	Pronóstico mediante Redes Neuronales -Entrenamiento Continuo	128
4.3.2.3	Comparación de Modelos Neuronales de Pronóstico por crecidas	132
4.3.2.4	Pronóstico mediante Modelos Paramétricos	133
4.3.2.5	Pronóstico mediante Red Neuronal Recurrente de Hopfield - Identificación.....	136
4.3.2.6	Resumen y Análisis de modelos de pronóstico a un paso de adelanto.....	139
4.3.2.7	Pronósticos a varios pasos de adelanto	144
4.4	Aplicación a Crecidas de deshielo.....	149
4.4.1	Cuenca del Estero Arrayán en La Montosa	149
4.4.1.1	Pronósticos mediante Redes Neuronales - Entrenamiento Unitario	150
4.4.1.2	Pronósticos mediante Redes Neuronales - Entrenamiento Continuo.....	164
4.4.1.3	Comparación de Modelos Neuronales de Pronóstico por crecidas	168
4.4.1.4	Pronóstico mediante Modelos Paramétricos	170
4.4.1.5	Resumen y análisis de modelos de pronósticos a un paso de adelanto	173
4.4.1.6	Pronósticos a varios pasos de adelanto	176

4.4.2	Cuenca del Mapocho en Los Almendros.....	180
4.4.2.1	Pronóstico mediante Redes Neuronales – Entrenamiento Unitario	181
4.4.2.2	Pronóstico mediante Redes Neuronales – Entrenamiento Continuo.....	193
4.4.2.3	Comparación de Modelos Neuronales de Pronóstico por crecidas	197
4.4.2.4	Pronóstico mediante Modelos Paramétricos	198
4.4.2.5	Resumen y análisis de modelos de pronóstico a un paso de adelanto.....	201
4.4.2.6	Pronósticos a varios pasos de adelanto	205
5	Conclusiones y Recomendaciones	208
5.1	Pronósticos Pluviales.....	213
5.1.1	Cuenca del Estero Arrayán en La Montosa	214
5.1.2	Cuenca Mapocho en Los Almendros.....	215
5.2	Pronósticos de Deshielo	216
5.2.1	Cuenca del Estero Arrayán en La Montosa	217
5.2.2	Cuenca del Mapocho en Los Almendros.....	218
5.3	Recomendaciones	219
	Referencias	221

Índice de Figuras

Figura 2.1 : Representación gráfica de la función logística, $1/(1+e^{-x})$, para x entre -10 y 10.	9
Figura 2.2: Modelo de red neuronal simple.	10
Figura 2.3: Modelo de red neuronal de tres capas.....	11
Figura 2.4: Bloque de retardo.....	12
Figura 2.5: Bloque integrador	12
Figura 2.6: Modelo red de Elman.....	17
Figura 2.7: Esquema y dinámica de una red recurrente simple de primer orden.	17
Figura 2.8: Esquema y dinámica de red parcialmente recurrente de primer orden.	18
Figura 2.9: Esquema y dinámica de red totalmente recurrente de primer orden.....	18
Figura 2.10: RRS desplegada en el instante t según retropropagación a través del tiempo.	26
Figura 2.11: Modelo red de Hopfield.....	30
Figura 2.12: Espacio de trayectorias de la red de Hopfield en un modelo bidimensional.	34
Figura 2.13: Errores de entrenamiento y validación durante el proceso de entrenamiento de la red neuronal.....	37
Figura 2.14: Esquema metodológico.....	45
Figura 2.15: Esquema Red Neuronal Recurrente de Elman.....	46
Figura 2.16: Análisis de pesos asociados a las variables de retardo. Crecidas Pluviales.....	48
Figura 2.17: Pesos Asociados a las variables de entrada a la red. Crecidas Pluviales.	49
Figura 2.18: Análisis de pesos asociados a las variables de retardo. Crecidas de Deshielo.....	49
Figura 2.19: Pesos asociados a las variables de entrada a la red. Crecidas de Deshielo.	50
Figura 3.1: Cuenca del Estero Arrayán en La Montosa.	56
Figura 3.2: Curva de Variación Estacional –Arrayán en La Montosa.	56
Figura 3.3: Curva de Variación Estacional - Mapocho en Los Almendros.....	57
Figura 3.4: Cuenca del Mapocho en Los Almendros.	58
Figura 3.5: Curva de Variación Estacional –Cerro Calán.	59
Figura 3.6: Curva de Hipsométrica –Mapocho en Los Almendros.....	61
Figura 3.7: Curva de Hipsométrica –Arrayán en La Montosa.	61
Figura 4.1: Transformada de Fourier de la serie de caudales –tiempo.....	67
Figura 4.2: Transformada de Fourier de la serie de caudales –tiempo.....	67
Figura 4.3: Errores de validación según crecida de entrenamiento. Modelos 2-7-7-1 y 3-7-7-1.	76
Figura 4.4: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	77
Figura 4.5: Errores Ponderados Modelos NEWELM y NEWFF-Validación.	81
Figura 4.6.b: Errores Ponderados Crecida 12 Modelos NEWELM y NEWFF.....	82
Figura 4.7: Errores de Validación Modelos NEWELM y NEWFF.	83

Figura 4.8: Errores de Validación Modelos NEWELM y NEWFF.	83
Figura 4.9: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.....	84
Figura 4.10: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.....	84
Figura 4.11: Índice TMEM entrenamiento. Modelos NEWELM y NEWFF.....	85
Figura 4.12: Índice EMEM validación crecida 5. Modelos NEWELM y NEWFF.	85
Figura 4.13: Índice EMRMOBS validación crecida 5. Modelos NEWELM y NEWFF.	86
Figura 4.14: Índice TEMEM validación crecida 5. Modelos NEWELM y NEWFF.....	86
Figura 4.15: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	88
Figura 4.16: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	88
Figura 4.17: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento	90
Figura 4.18: Pronósticos de crecidas Modelos NEWELM y NEWFF ³	93
Figura 4.19: Pronósticos de crecidas Modelos NEWELM cd. y NEWELM sd. ³	94
Figura 4.20: Pronósticos a un paso de adelanto. Modelo de Hopfield.....	102
Figura 4.21: Análisis gráfico de validación de crecidas, según modelo de pronóstico.....	104
Figura 4.22: Análisis de Modelos en el conjunto de validación.....	105
Figura 4.23: Análisis de Modelos en el conjunto de validación.....	106
Figura 4.24: Pronóstico de caudal máximo a t+x horas. Modelo NEWELM 3-7-7-1-Análisis continuo.	109
Figura 4.25: Pronóstico de caudal máximo a t+x horas. Modelo NEWELM 3-7-7-1-Análisis unitario.	110
Figura 4.26: Caudal máximo a t+16 horas. Modelos Recurrentes.	111
Figura 4.27: Transformada de Fourier de la serie de caudales –tiempo.....	113
Figura 4.28: Transformada de Fourier de la serie de caudales –tiempo.....	113
Figura 4.29: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	115
Figura 4.30: Errores de Validación para cada modelo NEWELM.....	119
Figura 4.31: Errores Ponderados para cada modelo NEWELM y NEWFF.....	125
Figura 4.32: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.....	127
Figura 4.33: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.....	127
Figura 4.34: Índice EMEM validación Crecida 13. Modelos NEWELM y NEWFF.	127
Figura 4.35: Índice EMRMOBS validación Crecida 13. Modelos NEWELM y NEWFF.	127
Figura 4.36: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento	129
Figura 4.37: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	131

Figura 4.38: Validación de crecidas según red neuronal y tipo de entrenamiento.....	132
Figura 4.39: Validación de crecidas según red neuronal y tipo de entrenamiento.....	133
Figura 4.40: Pronósticos a un paso de adelanto. Modelo de Hopfield.....	139
Figura 4.41: Análisis gráfico de validación de crecidas, según modelo de pronóstico.....	142
Figura 4.42: Análisis gráfico de validación de crecidas, según modelo de pronóstico.....	142
Figura 4.43: Análisis de Modelos en el conjunto de validación.....	144
Figura 4.44: Pronóstico de caudal máximo a t+x horas. Análisis Continuo NEWELM.....	146
Figura 4.45: Pronóstico de caudal máximo a t+x horas. Análisis Unitario NEWELM 3-5-5-1.	147
Figura 4.46: Caudal máximo a t+8 horas. Modelos Recurrentes.	148
Figura 4.47: Transformada de Fourier de la serie de caudales –tiempo.....	149
Figura 4.48: Transformada de Fourier de la serie de caudales –tiempo.....	150
Figura 4.49: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.....	152
Figura 4.50: ECM validación según número de neuronas en cada capa oculta y épocas de entrenamiento. Crecida 2 de entrenamiento.	152
Figura 4.51: Errores modelos NEWELM 1 con y sin gradiente.	156
Figura 4.52: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	157
Figura 4.53: Curvas de error ponderado modelos NEWELM y NEWFF.....	160
Figura 4.54: Curvas de error ponderado modelos NEWELM y NEWFF.....	161
Figura 4.55: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.....	162
Figura 4.56: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.....	163
Figura 4.57: Índice TMEM entrenamiento. Modelos NEWELM y NEWFF.....	163
Figura 4.58: Índice EMEM validación crecida 5. Modelos NEWELM y NEWFF.	163
Figura 4.59: Índice EMRMOBS validación crecida 5. Mod. NEWELM y NEWFF.....	164
Figura 4.60: Índice TMEM validación crecida 5. Modelos NEWELM y NEWFF.	164
Figura 4.61: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	165
Figura 4.62: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.	166
Figura 4.63: Análisis de Errores Ponderados según el modelo de pronóstico.	167
Figura 4.64: Validación de crecidas según modelo neuronal y tipo de entrenamiento.....	169
Figura 4.65: Análisis de Errores de validación según el modelo de pronóstico.....	175
Figura 4.66: Análisis de Modelos en el conjunto de validación.....	175
Figura 4.67: Caudal Pronosticado y observado (Validación continua). Modelo NEWELM.....	178
Figura 4.68: Transformada de Fourier de la serie de caudales –tiempo.....	180
Figura 4.69: Transformada de Fourier de la serie de caudales –tiempo.....	181

Figura 4.70: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.....	183
Figura 4.71: Errores Ponderados modelos NEWELM.....	185
Figura 4.72: Errores de validación modelos NEWELM.....	185
Figura 4.73: Errores de validación modelos NEWELM.....	186
Figura 4.74: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.....	187
Figura 4.75: Errores ponderados según crecida de entrenamiento. Modelo NEWFF y NEWELM. ..	190
Figura 4.76: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.....	191
Figura 4.77: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.....	192
Figura 4.78: Índice TMEM [Hr.] entrenamiento. Modelos NEWELM y NEWFF.....	192
Figura 4.79: Índice EMEM validación crecida 6. Modelos NEWELM y NEWFF.....	192
Figura 4.80: Índice EMRMOBS validación crecida 6. Modelos NEWELM y NEWFF.....	193
Figura 4.81: Índice TMEM [Hr.] validación crecida 6. Modelos NEWELM y NEWFF.....	193
Figura 4.82: Análisis de Errores de validación-Modelos NEWELM.....	194
Figura 4.83: Análisis de Errores de validación-Modelos NEWFF.....	196
Figura 4.84: Validación de crecidas según modelo neuronal y tipo de entrenamiento.....	198
Figura 4.85: Análisis de Errores de validación según el modelo de pronóstico.....	203
Figura 4.86: Análisis de Modelos en el conjunto de validación.....	204

Glosario de Términos

Notación

DGA	:	Dirección General de Aguas.
DMC	:	Dirección Meteorológica de Chile.
ECM	:	Error Cuadrático Medio.
LN	:	Línea de Nieve.
NEWELM	:	Red neuronal de Elman.
NEWFF	:	Red neuronal Prealimentada.
NEWHOP	:	Red Neuronal Recurrente de Hopfield
RNA	:	Red Neuronal Artificial.
RNR	:	Red Neuronal Recurrente.
RRS	:	Red Neuronal Recurrente Simple.
RTRL	:	Aprendizaje Recurrente en tiempo real.

Capítulo 1

1 Introducción

Dado el estado actual de desarrollo social y económico de Chile resulta ineludible ofrecer soluciones a los problemas asociados a fenómenos de crecidas e inundaciones, en respuesta a un nivel recurrente sin prevenciones ni soluciones claras. Por ello, en el campo de la hidráulica y la hidrología resulta imprescindible la utilización de modelos de previsión hidrológica, como base de la decisión en situaciones de crecidas, con dedicación al problema de explotación de embalses y para la obtención de soluciones aplicables en la práctica en tiempo real.

La mayoría de los sistemas tradicionales de control de crecidas son reactivos. Las intervenciones iniciales controlan principalmente las crecidas con elementos estructurales, que se consolidan posteriormente con algunas medidas no estructurales. Dentro de las primeras se encuentran las represas y reservorios, modificaciones a los canales de los ríos, diques y riberos, depresiones para desbordamiento, cauces de alivio y obras de drenaje, mientras que las segundas consisten en el control del uso de los terrenos aluviales mediante zonificación, los reglamentos para su uso, las ordenanzas sanitarias y de construcción, y la reglamentación del uso de la tierra de las cuencas hidrográficas.

Estas medidas tratan de resolver los problemas conforme van surgiendo y son de naturaleza unidisciplinaria. Desde hace ya algún tiempo se ha adquirido conciencia de que las estrategias de control de crecidas hasta ahora adoptadas están lejos de cumplir con las expectativas. Las medidas estructurales a menudo han afectado el equilibrio ecológico. La necesidad de cambiar la estrategia de acción defensiva por un modelo de acción proactiva está ampliamente reconocida, para lo cual se tendría que incentivar una cultura de prevención que gestione el riesgo de las crecidas y permita minimizar los daños asociados a ellas. Es en este contexto en el cual cabe la importancia de contar con algún sistema de pronóstico de crecidas, debido a los numerosos efectos negativos que tendrían la posibilidad de ser mitigados si se contara con un sistema de pronóstico adecuado a las necesidades del problema como son: una evacuación anticipada de la población circundante, el retiro de los bienes que pudiesen liberarse de ser afectados, la suspensión, en caso de que se amerite, del tránsito en calles,

caminos o puentes, y la correcta y oportuna construcción y operación de obras de control (compuertas, vertederos, etc.).

Cabe destacar que la utilidad de un modelo de pronóstico está altamente ligada a la antelación con la cual es factible obtener resultados, por lo tanto mientras mayor sea el desfase entre la causa (precipitación) y el efecto (crecida de río) mayor será la eficiencia del pronóstico.

Los estudios para pronosticar la ocurrencia de crecidas en Chile contemplan modelos que involucran funciones de transferencia (Márquez, 1991; Zavala, 1987 y López et al., 1995) los cuales presentan resultados aceptables basados en el supuesto de linealidad entre las variables precipitación y escorrentía superficial.

El posterior modelamiento del fenómeno P-Q mediante redes neuronales prealimentadas (Coulibaly, 2001; Döling y Varas, 2000; Vargas y Anguita, 1999; Smith y Velásquez, 1996; Toledo, 1996) libera el supuesto de linealidad entre precipitación y escorrentía superficial, siendo más acorde con las relaciones existentes entre las variables que modelan una crecida.

Consiguientemente se han realizado estudios que abarcan el uso de redes neuronales prealimentadas para el pronóstico de crecidas en cuencas pluvionivales (González, 2003; Troncoso, 1999, entre otros), mostrando predicciones más certeras y acordes con la realidad.

Nace por tanto, dentro del campo de las redes neuronales artificiales, la posibilidad de estudiar modelos de pronóstico que puedan resultar más eficientes que los ya estudiados a través de las redes neuronales recurrentes, no aplicadas con anterioridad al pronóstico de Crecidas en Chile, con existencia y aplicación en cuencas especialmente del continente Europeo (Imrie et al., 2000; Liong et al., 2000; Campolo et al., 1999; Dawson and Wilby 1999; Dawson and Wilby, 1998; Minns and Hall, 1997; Minns and Hall, 1996).

Las redes neuronales recurrentes son, en principio, un modelo computacional más potente que las clásicas redes neuronales hacia adelante. Esta mayor potencialidad radica en el hecho de que las redes recurrentes son capaces de procesar secuencias temporales gracias a la posibilidad de recordar parcialmente la historia relevante de la secuencia, por medio de una representación en forma de estado, son adaptables a distintas condiciones y con capacidad de mejorar su desempeño a medida que transcurre el tiempo (Kolen y Kremer, 2001; Haykin, 1999; Hertz et al. 1991)

Dentro del campo de aplicación de las redes neuronales recurrentes se encuentra el procesamiento de secuencias, entre las cuales se cuentan, la codificación del habla (Haykin y Li, 1995), ecualización de canales digitales (Kechriotis et al., 1994), predicción de melodías

inacabadas (Mozer, 1994), control de sistemas (Puskoriusy Feldkamp, 1994), predicción de series económicas (McCluskey, 1993), procesamiento del lenguaje humano (Elman, 1990).

1.1 Alcances y Objetivos

A modo general se pretende analizar la eficiencia de modelos de redes neuronales recurrentes aplicados al pronóstico de crecidas en tiempo real. Por tiempo real se entiende que el modelo sea capaz de procesar una muestra de señal antes de que ingrese al “sistema” o modelo la siguiente muestra y considerando la existencia de determinadas restricciones sobre el comportamiento temporal del modelo.

Aplicación realizada en la parte alta de la cuenca del río Mapocho, perteneciente a la hoya hidrográfica del río Maipo, dentro de la cual puede distinguirse una zona cordillerana y precordillerana de régimen nival y nivopluvial y otra zona baja mayoritariamente de régimen pluvial.

En consecuencia, la aplicación de este tipo de redes está referida a crecidas pluviales y de deshielo, cuyo pronóstico se enfoca en el primer y segundo caso a caudales instantáneos.

En términos específicos se pretende, encontrar los tipos de redes recurrentes, modelos, arquitecturas y configuraciones de ellas más apropiados para realizar pronósticos, demostrando sus ventajas y posibles limitaciones frente a otros métodos de pronóstico, ampliando el campo de herramientas que permitan realizar pronósticos de crecidas adecuados y con la intención de resultar útiles como alternativas preventivas y de mitigación.

1.2 Contribuciones de la Memoria

Como principales contribuciones de esta memoria se pueden mencionar lo siguiente:

- Entregar un marco teórico unificado de los recientes avances en el campo de redes neuronales para pronósticos de crecidas.
- Presentar las herramientas y programar los modelos más adecuados para el pronóstico de crecidas en tiempo real.
- Aplicar redes neuronales que reducen el número de variables de entrada necesarios para realizar pronósticos.

- Plantear el reconocimiento de patrones como alternativa de pronóstico de crecidas.
- Comparar todos los métodos de pronóstico de crecidas más comunes hasta ahora empleados con los nuevos modelos introducidos.
- Presentar nuevas soluciones a los problemas de pronóstico de crecidas en la parte alta de la cuenca del Río Mapocho, mejorando la eficiencia cuantitativa y cualitativa de dichos pronósticos.

1.3 Organización de la Memoria

Esta memoria centrada en el estudio de redes neuronales recurrentes para pronósticos de crecidas, está organizada como sigue:

En el capítulo 2 se presenta una breve introducción en torno a las redes neuronales generales, dando paso a una descripción de las redes recurrentes, partiendo de la base de las redes prealimentadas, posteriormente, una caracterización y descripción de las redes recurrentes más utilizadas en el ámbito de pronóstico, así como también las rutinas necesarias para su entrenamiento (algoritmos).

Incluye la formulación del problema y sus etapas, como también estrategias para evitar su sobre-entrenamiento y pérdida de generalidad.

Finalmente, se presenta una descripción de métodos de identificación de sistemas de tipo paramétricos de rápida obtención de resultados de pronósticos e índices de comparación de eficiencia de pronóstico.

El capítulo 3 contiene la caracterización de la parte alta de la cuenca del Río Mapocho, incluye análisis de registros en estaciones de tipo fluviográfico, pluviográfico y de temperatura.

En el capítulo 4 se incluye el pronóstico de crecidas pluviales y de deshielo en las subcuencas definidas por las estaciones fluviográficas: Arrayán en La montosa y Mapocho en Los Almendros por medio de la metodología de redes neuronales.

Se incluye análisis de tipo unitario, continuo, cuyos fundamentos serán explicados conforme se desarrollen y además, se realiza una comparación con otros métodos de pronóstico.

Finalmente, en el capítulo 5 se presentan las principales conclusiones obtenidas de este trabajo y se propone algunas líneas futuras de investigación.

Capítulo 2

2 Antecedentes Generales

2.1 Redes Neuronales

Una red neuronal puede verse como una máquina diseñada originalmente para modelar la forma en que el sistema nervioso de un ser vivo realiza una determinada tarea. Para lograr este objetivo, una red neuronal está formada por un conjunto de unidades de procesamiento interconectadas llamadas neuronas.

Cada neurona recibe como entrada un conjunto de señales discretas o continuas, las pondera e integra, y transmite el resultado a las neuronas conectadas a ella. Cada conexión entre dos neuronas tiene una determinada importancia asociada denominada peso sináptico o, simplemente, peso. En los pesos se suele guardar la mayor parte del conocimiento que la red neuronal tiene sobre la tarea en cuestión. El proceso mediante el cual se ajustan estos pesos para lograr un determinado objetivo se denomina aprendizaje o entrenamiento y el procedimiento concreto utilizado para ello se conoce como algoritmo de aprendizaje o algoritmo de entrenamiento. El ajuste de pesos es la principal forma de aprendizaje de las redes neuronales, aunque hay otras formas posibles.

El artículo de McCulloch y Pitts (1943) se considera como el punto de partida de la investigación en redes neuronales; este trabajo introdujo también la teoría de autómatas finitos como modelo computacional. McCulloch y Pitts propusieron un modelo simplificado de la actividad nerviosa real en el que cada neurona de una red neuronal podía activarse o desactivarse en función de lo que hicieran las neuronas conectadas a ella. Debido a que una neurona solo podía estar activada o desactivada, la capacidad computacional de la red completa se define en términos del conjunto de predicados lógicos que es capaz de computar. En este artículo ya aparecen redes neuronales recurrentes, a las que se denomina redes con ciclos.

Unos años después, Kleene (1956) reformuló algunos de estos resultados e introdujo una notación más compacta y general. A partir de ahí, el campo de las redes neuronales y el de la teoría de lenguajes comienzan a tomar caminos separados. De hecho, el segundo acaparó una

creciente atención desde entonces hasta nuestros días, mientras que el primero quedó prácticamente olvidado a raíz del trabajo de Minsky y Papert (1969). Salvo algunos trabajos aislados (Werbos 1974), las redes neuronales quedan relegadas a un segundo plano hasta la llegada de los ochenta.

Las redes neuronales destacan por su estructura de procesamiento paralelo y por su elevada capacidad de generalización (capacidad de producir salidas correctas para entradas no vistas durante el entrenamiento).

2.1.1 Redes Neuronales Recurrentes - Definiciones

La forma en que se interconectan las neuronas de una red neuronal define un grafo dirigido. Si este grafo es acíclico, la red se denomina red neuronal hacia delante y en el caso que posea ciclos se denomina red neuronal recurrente. En el primer caso están entre otros los perceptrones, las máquinas de vectores de soporte, etc.

Las redes neuronales recurrentes (Kolen y Kremer, 2001; Haykin, 1999; Hertz et al., 1991) son, en principio, un modelo computacional más potente que las clásicas redes neuronales hacia adelante. Esta mayor potencia proviene del hecho de que las redes recurrentes son capaces de procesar secuencias temporales gracias a la posibilidad de recordar parcialmente la historia relevante de la secuencia por medio de una representación en forma de estado. Esta *memoria* no existe de manera natural en las redes no recurrentes, utilizadas principalmente para el procesamiento de datos estáticos.

A simple vista el rasgo que diferencia las redes neuronales recurrentes de las que no lo son es simple: la existencia de, como mínimo, una conexión cíclica (*recurrente*) entre las neuronas que las configuran. Esta diferencia tiene, sin embargo, profundas implicaciones en la capacidad de computación del modelo y en los algoritmos de entrenamiento necesarios para conseguirla.

A la hora de clasificar las redes neuronales recurrentes se suele considerar la forma en la que el tiempo transcurre durante la actividad de la red. En este sentido se puede distinguir entre redes neuronales de tiempo discreto y redes neuronales de tiempo continuo. En las redes neuronales de tiempo discreto los cálculos ocurren a saltos, como si un reloj marcara el ritmo de procesamiento y se asume que en cada paso de tiempo la salida de cada neurona se obtiene instantáneamente a partir de sus entradas. En el caso de las redes neuronales de tiempo continuo (Pearlmutter 1995), las entradas y salidas de la red son función de una variable temporal continua y la dinámica de sus neuronas viene descrita por una ecuación diferencial.

En este trabajo se estudia la capacidad de las redes recurrentes para pronosticar crecidas. Ya que estas secuencias son de tiempo discreto, las redes también tienen que serlo forzosamente.

La predicción del siguiente elemento de una serie usando redes recurrentes es un tema clásico; véase, por ejemplo, los trabajos de Elman (1990) y Cleeremans et al. (1989). En este contexto es de esperar que la red desarrolle un modelo interno de la información relevante sobre una cierta cantidad de la historia pasada de la secuencia que le permita realizar predicciones acertadas sobre el futuro.

2.1.2 Propiedades Generales de las Redes Neuronales

- **Representación de relaciones de entrada /salida:** En aquellos casos en que se dispone de un conjunto de muestras de la relación entrada /salida a modelar se puede utilizar algún algoritmo de aprendizaje supervisado, que permita optimizar los pesos de la red de tal forma que la red neuronal modele de manera adecuada la relación entrada- salida por compleja que ésta sea.
- **Capacidad de generalización:** Una vez entrenada la red neuronal, se pueden presentar a la red distintos datos que los usados durante el proceso de aprendizaje. El correcto desempeño de la red al ser usada en los conjuntos de validación y prueba, dependerá de cuan buena sea su capacidad de generalización.
- **No linealidad:** Las redes neuronales modelan generalmente procesos no lineales aunque también pueden modelarse sistemas lineales (Muñoz, 1996).
- **Adaptabilidad:** Las redes neuronales son capaces de reajustar sus pesos para adaptarse a cambios en el entorno. Esto es especialmente útil cuando el entorno que suministra los datos de entrada es no estacionario, es decir, algunas de sus propiedades varían con el tiempo (Muñoz, 1996).
- **Tolerancia ante fallos:** Una red neuronal es tolerante ante fallos en el sentido de que los posibles fallos operacionales en partes de la red solo afectan débilmente al rendimiento de esta. Esta propiedad es debida a la naturaleza distribuida de la información almacenada o procesada en la red neuronal (Gómez, 1999).

- **Procesamiento Paralelo:** Las neuronas reales trabajan en paralelo, en el caso de las redes artificiales, resulta lógico que el uso de sólo un procesador no permite un proceso paralelo real, pero existe un paralelismo inherente; de tal manera que la estructura y modo de operación las hace especialmente adecuadas para el procesamiento paralelo real mediante multiprocesadores (Gómez, 1999).

2.1.3 Modelos neuronales

Dentro de una red neuronal, los elementos de procesamiento se encuentran agrupados por capas, siendo una capa una colección de neuronas. De acuerdo a la ubicación de la capa en la red neuronal artificial, ésta recibe diferentes nombres.

- **Capa de entrada:** Recibe las señales de la entrada de la red, algunos autores no consideran el vector de entrada como una capa pues allí no se lleva a cabo ningún proceso.
- **Capas ocultas:** Estas capas son aquellas que no tienen contacto con el medio exterior, sus elementos pueden tener diferentes conexiones y son éstas las que determinan las diferentes topologías de la red.
- **Capa de salida:** Recibe la información de la última capa oculta y transmite la respuesta al medio externo.

En el modelo más habitual de neurona se identifican cinco elementos básicos para la j -ésima neurona de una red de tiempo discreto:

- Un conjunto de n señales de entrada, $Z_i(t)$, $i=1, \dots, n$, que suministran a la neurona los datos del entorno; estos datos pueden ser externos a la red neuronal, pertenecientes a la salida de otras neuronas de la red o bien correspondientes a la salida anterior de la propia neurona.
- Un conjunto de sinapsis, caracterizada cada una por un peso propio $W_{j,i}$, $i=1, \dots, n$. El peso $W_{j,i}$ está asociado a la sinapsis que conecta la unidad i -ésima con la neurona j -ésima.

- Un sesgo W_j cuya presencia aumenta la capacidad de procesamiento de la neurona y que eleva o reduce la entrada a la neurona, según sea su valor positivo o negativo.
- Un sumador o integrador, que suma las señales de entrada, ponderadas con sus respectivos pesos, y el sesgo.
- Una función de activación g que suele limitar la amplitud de la salida de la neurona. La función de activación es la que define la salida de la neurona.

Las funciones de activación más utilizadas habitualmente son las siguientes:

1. **Función identidad:** Tiene la forma $g(x)=x$ y se utiliza cuando se desea acotar la salida de la neurona.

2. **Función escalón:** Adopta la forma

$$g_E(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2.1)$$

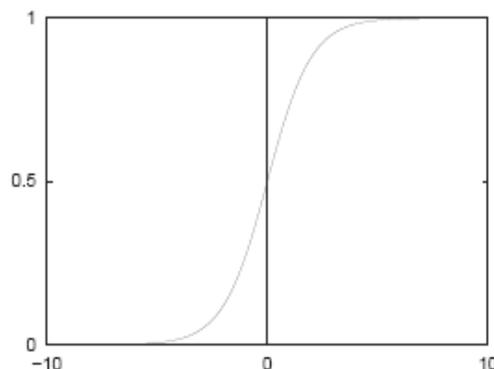
Y proporciona una salida con dos posibles valores.

3. **Función logística:** Las funciones sigmoideas son un conjunto de funciones crecientes, monótonas y acotadas que provocan una transformación no lineal de su argumento. Una de las más utilizadas es la función logística que se observa en la figura 2.1, definida por:

$$g_L = \frac{1}{1 + e^{-x}} \quad (2.2)$$

La función logística está acotada entre 0 y 1.

Figura 2.1 : Representación gráfica de la función logística, $1/(1+e^{-x})$, para x entre -10 y 10.



4. **Función tangente hiperbólica:** Otra función sigmoidea es la tangente hiperbólica.

$$g_T(x) = \tanh(x) \quad (2.3)$$

En este caso la función está acotada entre -1 y 1. La función logística y la tangente hiperbólica se relacionan mediante la ecuación:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.4)$$

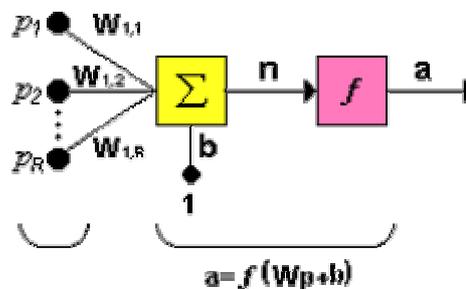
5. **Función de base radial:** Las más habituales son funciones gaussianas no monótonas del tipo.

$$g_B(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (2.5)$$

σ Define la anchura. La función alcanza su máximo valor cuando la entrada es cero.

En la figura 2.2 se presenta un modelo de red neuronal simple, donde es posible apreciar los elementos de cualquier modelo y el funcionamiento de ella. Posteriormente en la figura 2.3 se muestra un modelo multicapa, el cual posee mayores ventajas gracias a la capacidad de tener un mayor procesamiento de la información y generar un campo de decisión más amplio, como se ejemplifica a través de las neuronas tipo perceptrón.

Figura 2.2: Modelo de red neuronal simple.



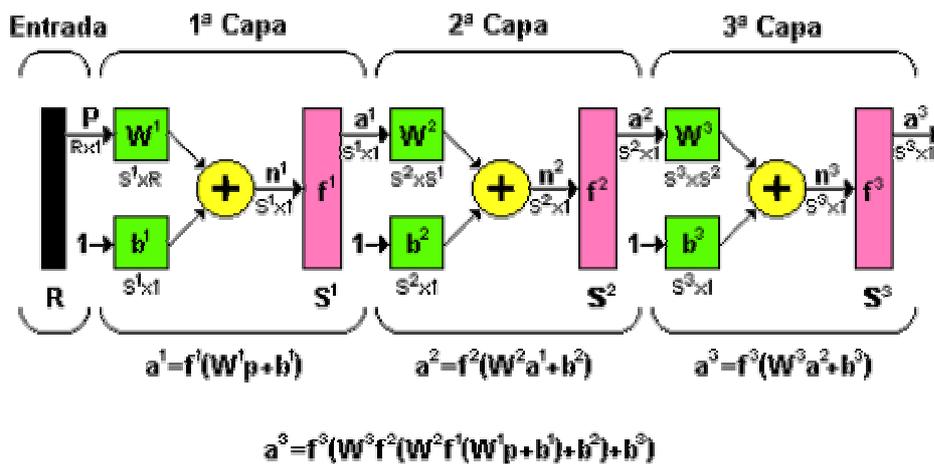
Utilizando la notación definida anteriormente se puede describir la operación de una neurona mediante la ecuación que determina su activación en el instante $t+1$.

$$Z_j[t+1] = g\left(\sum_{i=1}^n W_{ji}Z_i(t) + W_j\right) \quad (2.6)$$

Es habitual considerar el sesgo como un peso más de la red y no distinguirlo del resto de los pesos sinápticos. Por tanto mientras no se diga lo contrario, el término peso se refiere indistintamente tanto a $W_{j,i}$ como a W_j .

Si se tienen R entradas, S^1 neuronas en la primera capa, S^2 neuronas en la segunda capa, las cuales pueden ser diferentes, las salidas de las capas 1 y 2 son las entradas a las capas 2 y 3 respectivamente, así la capa 2 puede ser vista como una red de una capa con $R = S^1$ entradas, $S^1 = S^2$ neuronas y una matriz de pesos W^2 de dimensiones $S^1 \times S^2$.

Figura 2.3: Modelo de red neuronal de tres capas.



Las redes multicapa son de naturaleza estática, o sea su salida no evoluciona con el tiempo (para un patrón de entrada existe una salida asociada), pero pueden adquirir un comportamiento dinámico (para un patrón de entrada la salida posee un estado transitorio y converge a un valor en el estado estacionario) realimentando sus entradas con estados anteriores de sus salidas.

El caso de un perceptrón multicapa, la cual es una red con retroalimentación hacia adelante, compuesta por varias capas de neuronas entre la entrada y la salida de la misma, presenta la ventaja de permitir establecer regiones de decisión mucho más compleja que la de dos semiplanos, que establece dos regiones separadas por una frontera lineal en el espacio de entrada de los perceptrones en cambio un perceptrón con dos capas, puede formar cualquier

región convexa en el espacio. Las regiones convexas se forman mediante la intersección de regiones compuestas por cada neurona de la segunda capa (cada uno de estos elementos se comporta como un perceptrón simple).

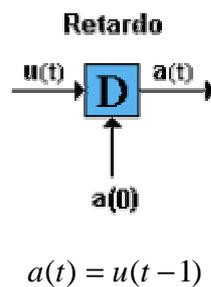
2.2 Redes Neuronales Recurrentes

2.2.1 Propiedades y Características

Las redes recurrentes conservan todas las propiedades de las redes neuronales hacia delante, siendo potencialmente más poderosas debido a su capacidad de recordar parcialmente la historia relevante de la secuencia por medio de una representación en forma de estado y a la introducción de dos conceptos, el bloque de retardo de la figura 2.4 y el bloque integrador de la figura 2.5.

- **Retardo**

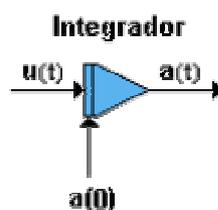
Figura 2.4: Bloque de retardo



La salida del bloque de retardo es el valor de entrada retrasado en un paso de tiempo, este bloque requiere que la salida sea inicializada con el valor $a(0)$ para el tiempo $t=0$; $a(0)$ se convierte en la salida de la red para el instante de tiempo inicial.

- **Integrador**

Figura 2.5: Bloque integrador



La salida del integrador es calculada de acuerdo a la expresión:

$$a(t) = \int_0^t u(t)dt + a(0) \quad (2.7)$$

Las redes neuronales recurrentes se dividen además de lo mencionado en el punto 2.1.1 en redes de estado continuo o de estado discreto, según la forma de las funciones de activación empleadas. Según la forma en que se utilicen, cabe también hablar de redes recurrentes de relajación y de redes recurrentes para el procesamiento temporal. Las primeras evolucionan durante una serie de iteraciones desde un estado inicial a un estado normalmente estacionario, momento en el que se consulta la salida de la red; la salida de las redes usadas para el procesamiento temporal, por otro lado, se consulta continuamente tras cada iteración.

Entre los grupos anteriores, este trabajo se centra en **las redes neuronales recurrentes de tiempo discreto con funciones de activación continuas aplicadas al procesamiento temporal de secuencias**. Se trabajará con sistemas dinámicos que poseen un estado que cambia conforme se procesa la secuencia de entrada y que proporciona una secuencia de salida a partir de las activaciones en cada instante de tiempo de las neuronas de salida. Más formalmente, una red neuronal recurrente de tiempo discreto de las que se han considerado en esta memoria puede verse (Forcada y Gori, 2001; Carrasco et al., 2000) como una séxtupla:

$$N = (X; U; Y; g_X; g_Y; x_0) \quad (2.8)$$

Donde:

- $X = [S_0, S_1]^{n \times 1}$ es el espacio de estados de la RNR. S_0 y S_1 son los valores que definen el rango de salida de las funciones de activación utilizadas en la función siguiente estado y n el número de unidades de estado.
- $U = R^{nu}$ es el espacio de los vectores de entrada y nu el número de componentes de la entrada.
- $Y = [T_0, T_1]^{ny}$ es el espacio de salida de la RNR T_0 y T_1 definen el rango de salida de las funciones de activación utilizadas en la función de salida y ny es el número de componentes del vector de salida.

- $g_x : X \times U \rightarrow X$ es la función de “siguiente estado”, que computa un nuevo estado $x(t)$ a partir del estado anterior $x(t-1)$ y la entrada actual $u(t)$.
- g_y es la función de salida, que habitualmente toma un de las dos formas siguientes:
 1. $g(y) : X \times U \rightarrow Y$. La salida $y(t)$ se obtiene a partir del estado anterior $x(t-1)$ y la entrada actual $u(t)$; este tipo de redes se conocen como máquinas neuronales de estados de Mealy.
 2. $g(y) : X \rightarrow Y$. La nueva salida $y(t)$ se obtiene a partir del estado recién alcanzado $x(t)$, estas redes se conocen como máquinas neuronales de estados de Moore.
- X_0 es el estado inicial de la RNR, es decir $x(0)$.

2.2.2 Aplicación de las Redes neuronales Recurrentes al Procesamiento de secuencias

A continuación se enumeran, siguiendo a Forcada y Gori (2001), algunas de las tareas relacionadas con el procesamiento de secuencias a las que se han aplicado las RNR. Los elementos de la siguiente lista no son excluyentes, es decir, alguna tarea podría englobarse en más de una categoría.

- **Predicción de series temporales.** Esta es una de las aplicaciones más habituales de las RNR. A partir de la historia pasada de una o más variables, la red neuronal debe proporcionar una predicción lo más correcta posible de su valor futuro. La mayor parte de los estudios de este tipo se centran en series económicas (McCluskey, 1993) o tomadas de fenómenos naturales (Aussem et al., 1995), pero hay otras aplicaciones como la continuación de melodías inacabadas (Mozer, 1994). Por otro lado, muchas de las tareas que se indican en los siguientes grupos pueden enfocarse como una tarea de predicción.
- **Procesamiento del lenguaje humano.** El análisis sintáctico de frases o el estudio de regularidades en el lenguaje son algunas de las tareas relacionadas con el lenguaje humano (escrito) a las que se han aplicado las RNR (Elman, 1990; 1991).

1. Esta condición puede relajarse para permitir que el intervalo de X sea abierto.
2. Si la función de salida no está acotada, el intervalo de Y puede ser abierto.

- **Ecuación de canales digitales.** Los efectos del canal sobre la señal transmitida en comunicaciones digitales pueden hacer que ésta sea irreconocible al llegar al receptor. Se hace necesario, por tanto, el uso de algún tipo de filtro inverso que deshaga estos efectos y proporcione una señal similar a la original. Esta tarea de traducción de señales se conoce normalmente como ecualización y varios trabajos se han acercado a ella con RNR (Ortiz Fuentes y Forcada, 1997; Cid-Sueiro et al., 1994; Kechriotis et al., 1994).
- **Codificación del habla.** Existe gran cantidad de técnicas para comprimir una señal de voz de manera que pueda ser transmitida por un canal con el menor número de bits por segundo posible (para una calidad de recepción determinada). Algunas de estas técnicas se basan en la llamada **codificación predictiva**, en ella lo que se envía no es la señal, sino la diferencia entre su valor real y el valor predicho por un determinado predictor. Si el predictor es bueno, esta diferencia será pequeña y se necesitarán pocos bits para codificarla. Las RNR también han sido usadas como predictores para la codificación del habla (Haykin y Li, 1995).
- **Reconocimiento del habla.** El reconocimiento del habla puede considerarse como una tarea de traducción de secuencias (por ejemplo, cuando se asigna una secuencia de fonemas a una secuencia de vectores acústicos obtenidos mediante el procesamiento de una señal de voz) o como una tarea de clasificación de secuencias (por ejemplo, al asignar una palabra a una serie de vectores acústicos). Existen varias referencias que utilizan RNR para el reconocimiento del habla (Robinson y Fallside, 1991).
- **Inferencia gramatical.** Dado un conjunto de cadenas pertenecientes a un cierto lenguaje, uno de los problemas más estudiados es la inferencia de un modelo (un autómata finito o una gramática independiente del contexto, por ejemplo) que describa de manera correcta ese lenguaje. Este es posiblemente el campo en el que las RNR han proporcionado mejores resultados (Carrasco et al., 2000; Castaño et al., 1995; Cleeremans et al., 1989).
- **Control de sistemas.** Las RNR pueden ser también entrenadas (Puskorius, 1994) para controlar un sistema real en que la salida siga un determinado patrón temporal.

2.2.3 Modelos Neuronales Recurrentes Tradicionales

En el contexto de las redes recurrentes existen redes dinámicas por naturaleza, como lo son la red de Elman, de Jordan y de Hopfield, y redes dinámicas que siendo de naturaleza estática como lo son las redes multicapa logran el comportamiento dinámico realimentando sus entradas con muestras anteriores de las salidas, el comportamiento dinámico de las redes recurrentes hace que sean una poderosa herramienta para simular e identificar sistemas dinámicos no lineales.

2.2.3.1 Redes de Primer Orden

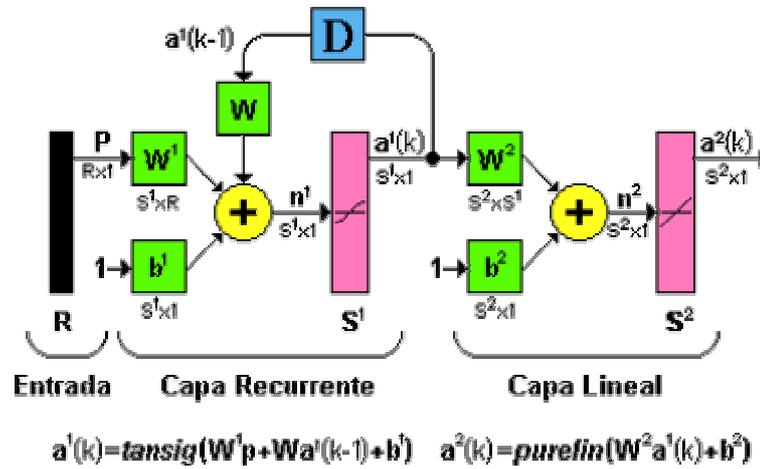
A. Red recurrente simple (Elman, 1990)

La red de Elman es una red neuronal recurrente simple de primer orden que está constituida por una capa de *sensores* o unidades de entrada, dos capas intermedias de unidades (una capa de *unidades ocultas* y otra capa de *unidades de contexto*) y una capa de *unidades de salida*. Los sensores de entrada y las unidades de salida recogen la información del entorno, es decir, la suministrada por el conjunto de patrones de entrenamiento, mientras que las demás no. Las unidades de salida reciben la señal suministrada por la salida de las unidades de la capa oculta ponderadas por los correspondientes pesos sinápticos y suelen utilizar como función de transferencia la función identidad o una función de tipo lineal. Sin embargo, la novedad de este modelo la introducen las unidades de contexto que se utilizan para memorizar las salidas de las unidades ocultas en la etapa anterior, de manera que cada unidad de contexto tiene como salida la salida de la unidad oculta correspondiente en la etapa anterior. Por lo tanto, esta red es sólo parcialmente recurrente y cada unidad de proceso oculta recibe como entrada las salidas de las unidades de contexto y de los sensores de entrada ponderadas por sus pesos sinápticos. De esta manera la salida de la red depende no sólo del patrón de entrada actual sino también de los patrones anteriores a través de las unidades de contexto. Las unidades de proceso de la capa oculta tienen como función de transferencia la función logística, la función tangente hiperbólica o la función identidad.

La figura 2.6 esquemáticamente pretende mostrar el modelo de la red de Elman, siendo P el vector de entradas, W la matriz de pesos que pondera las entradas y el vector b el correspondiente vector de sesgos. Este vector sumado al vector de salida anterior de la capa oculta ponderado por el correspondiente peso asociado entra nuevamente a la o las capas

ocultas, cuya salida entra a una capa que involucra una función de transferencia lineal y produce la salida deseada.

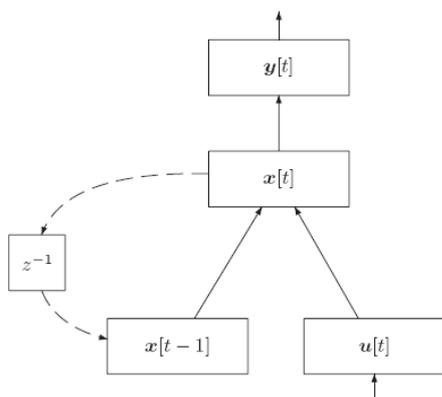
Figura 2.6: Modelo red de Elman.



Para establecer la dinámica de la red, apreciable en la figura 2.7, se consideran las funciones de activación g_x y g_y , n_x es el número de neuronas de estado, n_u el número de entradas a la red y n_y es el número de neuronas de salida.

Los superíndices indican el cálculo en el que está implicado el peso, por ejemplo $W_{i,j}^{y,u}$ indica que ese peso contribuye a determinar la salida y y a partir de la entrada u . Por otra parte W_i^x indica que este peso es un sesgo implicado en el cálculo del estado x . Los subíndices muestran las unidades concretas que se ven afectadas (conectadas) y van paralelos a los subíndices. Z^{-1} representa el vector de células de retardo temporal.

Figura 2.7: Esquema y dinámica de una red recurrente simple de primer orden.



$$y_i(t) = g_y(Y_i(t)) \quad i = 1, \dots, n_y.$$

$$Y_{ii}(t) = \sum_{j=1}^{n_x} W_{i,j}^{y,x} x_j(t) + W_i^y$$

$$x_i(t) = g_x(X_i(t)) \quad i = 1, \dots, n_x.$$

$$X_i(t) = \sum_{j=1}^{n_u} W_{i,j}^{x,u} u_j(t) + \sum_{j=1}^{n_x} W_{i,j}^{x,x} x_j(t-1) + W_i^x$$

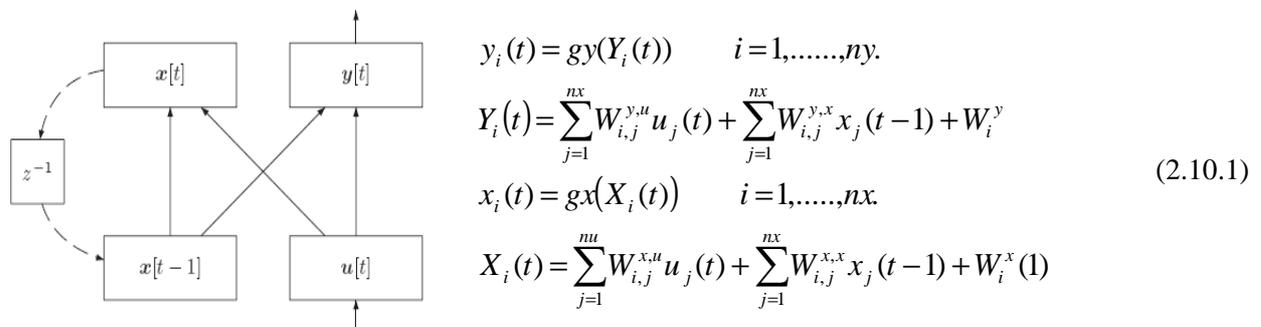
(2.9.1)

B. Red parcialmente Recurrente (Robinson y Fallside, 1991)

Red recurrente de propagación de errores o red parcialmente recurrente (Robinson y Fallside 1991) de primer orden, cuya dinámica viene dada por:

En este tipo de red, como se muestra en la figura 2.8, el estado $x(t)$ depende de las variables de entrada al sistema y del estado anterior a través de una función de retardo, a su vez no existe una relación directa entre el estado $x(t)$ y la salida $y(t)$, pero si existe una relación directa entre las variables de entrada, sus retardos y la salida $y(t)$, lo que la hace ser más “directamente” recurrente que una red de Elman, pero no totalmente debido a que la recurrencia no relaciona los estado $x(t)$ e $y(t)$.

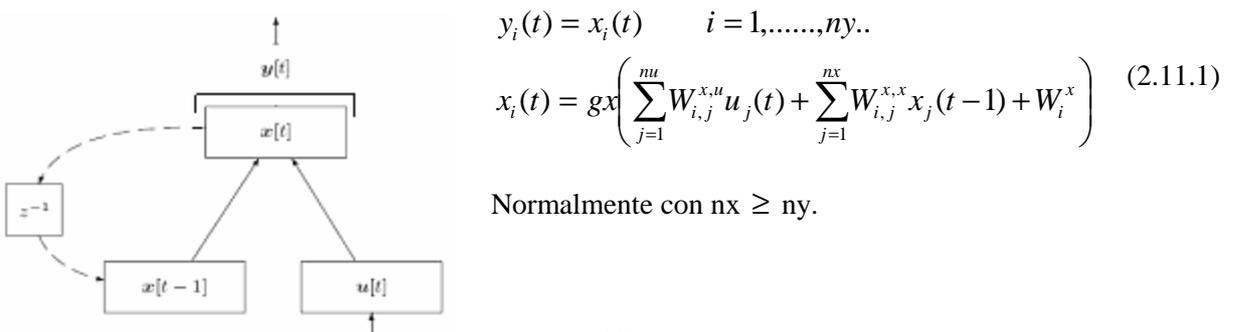
Figura 2.8: Esquema y dinámica de red parcialmente recurrente de primer orden.



C. Red totalmente recurrente (Williams y Zipser, 1989)

La red totalmente recurrente (Williams y Zipser 1989) de primer orden viene descrita por la relación de igualdad existente entre el estado $x(t)$ y la salida $y(t)$, luego, el estado $x(t)$ es el pronóstico o salida $y(t)$ de la red neuronal. El estado $x(t)$ queda descrito al igual que para una red parcialmente recurrente, es decir dependiente de las variables de entrada al modelo y del estado anterior ($x(t-1)$). Ver figura 2.9.

Figura 2.9: Esquema y dinámica de red totalmente recurrente de primer orden.



2.2.3.2 Redes de Segundo Orden

La red recurrente simple de segundo orden (RRS2), utilizada por Carrasco et. al (1996), viene dada por:

$$\begin{aligned}
 y_i(t) &= gy \left(\sum_{j=1}^{nx} W_{i,j}^{y,x} x_j(t) + W_i^y \right) \quad i = 1, \dots, ny. \\
 x_i(t) &= gx \left(\sum_{j=1}^{nx} \sum_{k=1}^{nu} W_{i,j,k}^{x,x,u} x_j(t-1) u_k(t) + W_i^x \right) \quad i = 1, \dots, nx.
 \end{aligned} \tag{2.9.2}$$

También puede obtenerse una versión de segundo orden (RPR2) de la red parcialmente recurrente. Esta topología ha sido utilizada, entre otros, por Omlin y Giles (1996) y responde a las ecuaciones:

$$\begin{aligned}
 y_i(t) &= gy \left(\sum_{j=1}^{nx} \sum_{k=1}^{nu} W_{i,j,k}^{y,x,u} x_j(t-1) u_k(t) + W_i^y \right) \quad i = 1, \dots, ny. \\
 x_i(t) &= gx \left(\sum_{j=1}^{nx} \sum_{k=1}^{nu} W_{i,j,k}^{x,x,u} x_j(t-1) u_k(t) + W_i^x \right) \quad i = 1, \dots, nx.
 \end{aligned} \tag{2.10.2}$$

La red totalmente recurrente de segundo orden (RTR2) propuesta por Giles et. al. (1992) se define a partir de:

$$\begin{aligned}
 y_i(t) &= x_i(t) \quad i = 1, \dots, ny. \\
 x_i(t) &= gx \left(\sum_{j=1}^{nx} \sum_{k=1}^{nu} W_{i,j,k}^{x,x,u} x_j(t-1) u_k(t) + W_i^x \right) \quad i = 1, \dots, nx.
 \end{aligned} \tag{2.11.2}$$

Donde, como RTR, se suele hacer $nx \geq ny$.

2.2.3.3 Adición de una ventana temporal

Todos los modelos anteriores de RNR pueden ser ampliados con la incorporación a sus entradas de una memoria explícita a corto plazo. De esta forma, la entrada a la red consistirá en el valor actual $u[t]$ concatenado con los $p-1$ valores anteriores $u[t-1]; \dots; u[t-p+1]$. A p se le conoce como orden de la memoria de entrada u *orden de entrada*. Una de las primeras utilizaciones de esta idea (con redes no recurrentes) fue el proyecto NetTalk (Sejnowski y Rosenberg, 1987).

Con esta ampliación, la red recurrente tiene a su disposición de forma explícita la historia mas reciente y puede, en teoría, utilizar el estado para almacenar información relativa a un pasado más remoto. Al incorporar una ventana de entradas de orden p la ecuación (1), por ejemplo, se convierte ahora en:

$$X_i(t) = \sum_{j=1}^{nu} \sum_{k=1}^p W_{i,j(k)}^{x,u} u_j(t-k-1) + \sum_{j=1}^{nx} W_{i,j}^{x,x} x_j(t-1) + W_i^x \quad (2.12)$$

Donde se ha utilizado $W_{i,j(k)}^{x,u}$ para referirse al peso que une $u_j(t-k+1)$ con la neurona i del estado. Las ecuaciones de las otras redes recurrentes se modificarían de manera similar.

Para efectos prácticos, es aconsejable no alterar las ecuaciones y considerar que la entrada de la red tiene tamaño $n'u = p \cdot nu$, donde nu es el número de componentes del vector de entrada. Además de controlar esa entrada realizando el desplazamiento de todos los componentes en cada instante de tiempo para descartar la información más antigua e incorporar la nueva.

2.2.3.4 Algoritmos de Entrenamiento

Los algoritmos de entrenamiento modifican los parámetros configurables de la red intentando minimizar el error cuadrático medio. Es un problema complejo de modificación sin restricciones para el que es necesario aplicar criterios heurísticos.

El entrenamiento de una red neuronal se realiza mediante un proceso de aprendizaje. Para realizar este proceso se debe inicialmente tener definida la topología de la red, esto es:

- Número de neuronas en la capa de entrada, el cual depende del número de componentes del vector de entrada.
- Cantidad de capas ocultas y número de neuronas de cada una de ellas, los cuales deben ser determinados tras experimentar con distintas configuraciones o usar límites

dados en la teoría para algunos tipos de tareas. Será explicado mayormente en el punto 2.3.

- Número de neuronas en la capa de la salida, el cual depende del número de componentes del vector de salida o patrones objetivo.
- Funciones de transferencia requeridas en cada capa, elegidas de acuerdo a las características del problema.

Con base en la topología escogida se asignan valores iniciales a cada uno de los parámetros que conforma la red.

En algunos casos, la red neuronal se somete a una fase de entrenamiento, tras la cual sus pesos se congelan. Durante esta fase a la red se le presentan los datos del llamado conjunto de entrenamiento. A continuación, esta red se evalúa sobre un nuevo conjunto de datos para determinar la corrección del aprendizaje. Este tipo de entrenamiento se conoce como **Entrenamiento fuera de línea**.

En otros casos, las fases de entrenamiento y evaluación no están tan claramente separadas y la salida de la red se usa simultáneamente como punto de referencia para cambiar los pesos y como producto utilizado en la resolución de la tarea en cuestión. Este tipo de entrenamiento se conoce como **Entrenamiento en línea**.

Los principales algoritmos de entrenamiento se basan en el cálculo del gradiente de la función de error, esto es, de la derivada de la función de error con respecto a los distintos parámetros ajustables de la red. Se intenta encontrar el mínimo de la función de error mediante la búsqueda de un punto donde el gradiente se anule.

La función de error (ECM) define una superficie multidimensional (hipersuperficie) conocida como **hipersuperficie de error**. Normalmente, ésta tiene un mínimo global (posiblemente múltiples mínimos globales debido a simetrías de la red) y muchos mínimos locales, que pueden no corresponder a una solución correcta del problema. Estos mínimos locales son consecuencia de la elevada dimensionalidad del espacio de búsqueda y son el mayor problema, al quedar atrapados en ellos, de casi todos los algoritmos de aprendizaje de redes neuronales, especialmente de aquellos que realizan una búsqueda local como los basados en el gradiente de la función de error.

A. Algoritmo de Descenso por Gradiente

Una de las variantes basadas en el gradiente más utilizadas es el **descenso por el gradiente**. En él los sucesivos ajustes realizados a los parámetros se hacen de forma individual para cada uno de ellos, digamos W_i , en sentido opuesto al vector de gradiente $dE(n)/dW_i(n)$:

$$W_i(n+1) = W_i(n) - \alpha \frac{dE(n)}{dW_i(n)} \quad (2.13)$$

Donde α es un parámetro conocido como tasa de aprendizaje, que ha de tomar un valor convenientemente pequeño.

Al pasar de la iteración n a la $n+1$, el algoritmo aplica la corrección:

$$\Delta W_i(n) = W_i(n+1) - W_i(n) = -\alpha \frac{dE(n)}{dW_i(n)} \quad (2.14)$$

Puede demostrarse (Haykin, 1999) que para valores positivos muy pequeños de la tasa de aprendizaje y funciones de error globales, la formulación del algoritmo de descenso por el gradiente permite que la función de error decrezca en cada iteración. La tasa de aprendizaje α tiene, por tanto, una enorme influencia en la convergencia del método de descenso por el gradiente. Si α es pequeña, el proceso de aprendizaje se desarrolla suavemente, pero la convergencia del sistema a una solución estable puede llevar un tiempo excesivo. Si α es grande, la velocidad de aprendizaje aumenta, pero existe el riesgo de que el proceso de aprendizaje diverja y el sistema se vuelva inestable.

Es habitual añadir un término de **momento** (Plaut et al., 1986; Rumelhart et al., 1986) que en ocasiones puede acelerar el aprendizaje y reducir el riesgo de que el algoritmo se vuelva inestable.

La nueva ecuación de actualización del parámetro ajustable W_i tiene la forma:

$$\Delta W_i(n) = W_i(n+1) - W_i(n) = -\alpha \frac{dE(n)}{dW_i(n)} + \gamma \Delta W_i(n-1) \quad (2.15)$$

Donde α es la tasa de aprendizaje y γ es la constante de momento.

El efecto del momento es el siguiente:

Si la derivada parcial del error tiene el mismo signo algebraico durante varias iteraciones seguidas (lo que indicaría que se está descendiendo por una “ladera”), el término $\Delta W_i[n]$ irá creciendo y el incremento del parámetro será mayor; si la derivada parcial cambia de signo

constantemente (indicación de que el algoritmo se encuentra en una zona complicada), el valor de $\Delta W_i[n]$ se va reduciendo y el parámetro se ajusta de forma más precisa.

Existen otros métodos de optimización más sofisticados (por ejemplo, métodos que consideran la información suministrada por las derivadas de segundo orden), que, en general, proporcionan mejores resultados que el descenso por el gradiente, a veces simplemente con una leve modificación.

Algunos de ellos son el método de Newton, el algoritmo de Levenberg-Marquardt o el método de los gradientes conjugados (Shepherd, 1997). Todos ellos han sido aplicados abundantemente sobre redes no recurrentes y escasamente en redes recurrentes (Chang y Mak, 1999; Chan y Szeto, 1999).

A continuación se describen dos algoritmos de entrenamiento que se basan en el cálculo del gradiente. El aprendizaje recurrente en tiempo real y retropropagación en el tiempo, de hecho usan el descenso por gradiente y pueden considerarse más bien como formas distintas de calcular el valor de la derivada correspondiente.

B. Aprendizaje Recurrente en tiempo real (RTRL)

Como ya se ha dicho RTRL se considera como una forma de calcular las derivadas parciales de la función de error, aunque algunos autores se refieren a él como a un algoritmo per-se al combinarlo con el ajuste de pesos realizado con el descenso por gradiente.

A continuación se derivan las ecuaciones de este tipo de entrenamiento para una red recurrente con la dinámica de la red de Elman.

Aplicando la regla de la cadena a la función de error y considerando un parámetro ajustable cualquiera, se tiene que:

$$E(t) = \frac{1}{2} * \sum_{i=1}^{ny} (d_i(t) - y_i(t))^2$$

$$\frac{\partial E(t)}{\partial \xi} = - \sum_{i=1}^{ny} (d_i(t) - y_i(t)) * \frac{\partial y_i(t)}{\partial \xi}$$
(2.16)

En función de lo anterior, la derivada $\frac{\partial y_i(t)}{\partial \xi}$ depende del parámetro concreto considerado.

Luego, las expresiones de estas derivadas para todos los pesos y sesgos de la red.

$$\begin{aligned}
\frac{\partial y_i(t)}{\partial w_i^y} &= g'_Y(Y_i(t))\delta_{i,i} \\
\frac{\partial y_l(t)}{\partial w_{i,j}^{y,x}} &= g'_Y(Y_l(t))x_j(t)\delta_{l,i} \\
\frac{\partial y_i(t)}{\partial w_j^x} &= g'(Y_i(t))\sum_{i=1}^{nx} w_{i,j}^{y,x} \frac{\partial x_i(t)}{\partial w_j^x} \\
\frac{\partial y_i(t)}{\partial w_{j,k}^{x,u}} &= g'(Y_l(t))\sum_{i=1}^{nx} w_{i,j}^{y,x} \frac{\partial x_i(t)}{\partial w_{j,k}^{x,u}} \\
\frac{\partial y_i(t)}{\partial w_{j,k}^{x,x}} &= g'(Y_l(t))\sum_{i=1}^{nx} w_{l,i}^{y,x} \frac{\partial x_i(t)}{\partial w_{j,k}^{x,x}}
\end{aligned} \tag{2.17}$$

Para las derivadas de las ecuaciones anteriores debe tenerse en cuenta las siguientes expresiones:

$$\begin{aligned}
\frac{\partial w_i^y}{\partial w_j^y} &= \delta_{i,j} \\
\frac{\partial w_{i,j}^{y,x}}{\partial w_{k,l}^{y,x}} &= \delta_{i,k} \delta_{j,l} \\
\frac{\partial w_i^x}{\partial w_j^x} &= \delta_{i,j} \\
\frac{\partial w_{i,j}^{x,u}}{\partial w_{k,l}^{x,u}} &= \delta_{i,k} \delta_{j,l} \\
\frac{\partial w_{i,j}^{x,x}}{\partial w_{k,l}^{x,x}} &= \delta_{j,k} \delta_{j,l}
\end{aligned} \tag{2.18}$$

Donde la función $\delta_{i,j}$ es delta de Kronecker, función que toma valor 1 cuando $i = j$ y 0 en otros casos.

Las derivadas de la función del estado $x(t)$ de las ecuaciones son recurrentes en RTRL como resultado de la propia recurrencia de la red:

$$\begin{aligned}
\frac{\partial x_i(t)}{\partial W_j^x} &= g'_X(X_i(t)) \left[\delta_{i,j} + \sum_{k=1}^{nx} w_{i,k}^{x,x} \frac{\partial x_k(t-1)}{\partial W_j^x} \right] \\
\frac{\partial x_i(t)}{\partial W_{j,k}^{x,u}} &= g'_X(X_i(t)) \left[u_k(t)\delta_{i,j} + \sum_{m=1}^{nx} w_{i,m}^{x,x} \frac{\partial x_m(t-1)}{\partial W_{j,k}^{x,u}} \right] \\
\frac{\partial x_i(t)}{\partial W_{j,k}^{x,x}} &= g'_X(X_i(t)) \left[x_k(t-1)\delta_{i,j} + \sum_{m=1}^{nx} w_{i,m}^{x,x} \frac{\partial x_m(t-1)}{\partial W_{j,k}^{x,x}} \right]
\end{aligned} \tag{2.19}$$

C. Algoritmo de retropropagación en el tiempo. Backpropagation

Backpropagation es un tipo de red de aprendizaje supervisado, que emplea un ciclo propagación – adaptación de dos fases. Una vez que se ha aplicado un patrón a la entrada de la red como estímulo, éste se propaga desde la primera capa a través de las capas superiores de la red, hasta generar una salida. La señal de salida se compara con la salida deseada y se calcula una señal de error para cada una de las salidas.

Las salidas de error se propagan hacia atrás, partiendo de la capa de salida, hacia todas las neuronas de la capa oculta que contribuyen directamente a la salida. Sin embargo las neuronas de la capa oculta solo reciben una fracción de la señal total del error, basándose aproximadamente en la contribución relativa que haya aportado cada neurona a la salida original. Este proceso se repite, capa por capa, hasta que todas las neuronas de la red hayan recibido una señal de error que describa su contribución relativa al error total.

Basándose en la señal de error percibida, se actualizan los pesos de conexión de cada neurona, para hacer que la red converja hacia un estado que permita clasificar correctamente todos los patrones de entrenamiento.

La importancia de este proceso consiste en que, a medida que se entrena la red, las neuronas de las capas intermedias se organizan a sí mismas de tal modo que las distintas neuronas aprenden a reconocer distintas características del espacio total de entrada.

Después del entrenamiento, cuando se les presente un patrón arbitrario de entrada que contenga ruido o que esté incompleto, las neuronas de la capa oculta de la red responderán con una salida activa si la nueva entrada contiene un patrón que se asemeje a aquella característica que las neuronas individuales hayan aprendido a reconocer durante su entrenamiento. Y a la inversa, las unidades de las capas ocultas tienen una tendencia a inhibir su salida si el patrón de entrada no contiene la característica para reconocer, para la cual han sido entrenadas.

Backpropagation trabaja bajo aprendizaje supervisado y por tanto necesita un set de entrenamiento que le describa cada salida y su valor de salida esperado de la siguiente forma:

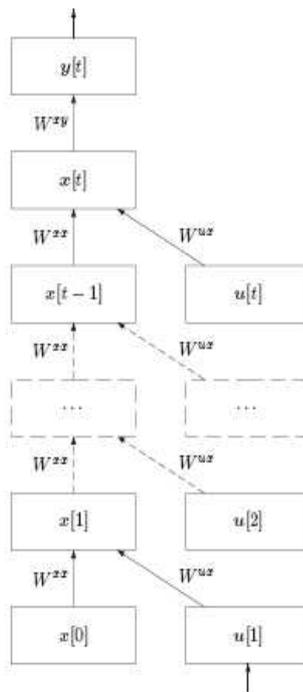
$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\}$$

Donde p_Q es una entrada a la red y t_Q es la correspondiente salida deseada para el patrón q -ésimo. El algoritmo debe ajustar los parámetros de la red para minimizar el error medio cuadrático.

Al calcular las derivadas parciales de la función de error se supone que el comportamiento temporal de la RNR puede ser desplegado en el espacio en forma de red hacia adelante, como se puede apreciar en la figura 2.10, de tal manera que la red va creciendo una y otra vez cada instante de tiempo. Así, suponiendo una RRS, las unidades de entrada y las unidades de estado del instante t se convierten en dos nuevas capas de una red extendida; y así sucesivamente hasta llegar al primer instante de tiempo. Las unidades de entrada y unidades ocultas del instante $t-1$, se convierten también en dos nuevas capas de la red extendida; y así sucesivamente hasta llegar al primer instante de tiempo. Como realmente sólo existe un conjunto de unidades de entrada y ocultas, los pesos equivalentes en las distintas capas virtuales han de tener idéntico valor.

En el caso de una secuencia de longitud relativamente extensa, las necesidades temporales y espaciales del algoritmo crecerían linealmente conforme la red fuera procesando las entradas. Por ello, en estos casos, la historia de la red se trunca y se considera irrelevante cualquier información anterior a t_0 instantes de tiempo. El valor de t_0 se conoce como umbral de truncamiento.

Figura 2.10: RRS desplegada en el instante t según retropropagación a través del tiempo.



En este apartado se deriva las ecuaciones de BPTT para una RRS con la dinámica de la red de Elman y sus ecuaciones definidas anteriormente. La red neuronal de la figura 2.10 es una red

no recurrente con lo que las derivadas de la función de error serán las mismas que las calculadas con la técnica de retropropagación para este tipo de redes (Rumelhart et al.1986). Si se utiliza el descenso por el gradiente, el algoritmo se limita a actualizar cada peso (no se muestran las ecuaciones de los sesgos) mediante la llamada *regla delta generalizada* como sigue:

$$\begin{aligned}\Delta W_{i,j}^{y,x}(t) &= \alpha \delta_i^y(t) x_j(t) \\ \Delta W_{i,j}^{x,x}(t) &= \alpha \sum_{\tau=1}^t \delta_i^x(\tau) x_j(\tau - 1) \\ \Delta W_{i,j}^{x,u}(t) &= \alpha \sum_{\tau=1}^t \delta_i^x(\tau) u_j(\tau)\end{aligned}\tag{2.20}$$

Donde la señal de error y la señal de error retropropagada se definen a partir de:

$$\begin{aligned}\delta_i^y(t) &= \frac{\partial E(t)}{\partial Y_i(t)} \\ \delta_i^x(t) &= g'_x(X_i(t)) \sum_{j=1}^t \delta_j^y(t) W_{j,i}^{y,x}(t)\end{aligned}\tag{2.21}$$

para $1 \leq \tau < t$

$$\delta_i^x(t) = g'_x(X_i(t)) \sum_{j=1}^t \delta_j^x(\tau + 1) W_{j,i}^{x,x}(t)$$

La señal de error retropropagada puede verse como un emisario del presente t que viaja hacia atrás en el tiempo para influir en el pasado de forma que este contribuya en mayor medida a los acontecimientos que se desearía haber observado en el presente.

D. Variaciones de Algoritmos

Considerando una expresión genérica de la forma:

$$W_i(n+1) = W_i(n) - \alpha_n * D_n\tag{2.22}$$

Donde:

W (k+1): Vector de pesos en la iteración k+1.

D_k: Dirección de búsqueda en la iteración k .

α_k : Tasa de aprendizaje.

Los distintos métodos de optimización basados en el gradiente se diferencian en la forma de obtener la dirección de búsqueda D_k a partir de la información de primer orden contenida en el gradiente ∇E_k , donde E es el error y, en algunos casos, de la estimación de información adicional de segundo orden (Muñoz, 1996). Son métodos de búsqueda local que reducen el problema de minimización n-dimensional a sucesivas minimizaciones unidimensionales, de tal manera que es posible tratar problemas de elevada dimensión de forma eficaz, sin que existan numerosas evaluaciones de la función de error y de su gradiente.

Las siguientes expresiones definen la dinámica de minimización de errores:

- Gradiente Conjugado: $D_k = -\nabla E_k + \gamma_k * D_{k-1}$ (2.23)

- Cuasi-Newton : $\alpha_k * D_k = -B_k * \nabla_k$ (2.24)

Con B estimación del inverso de la matriz Hessiana de E, dada por:

$$\nabla^2 E = \left| \frac{\partial^2 E}{\partial W_i * \partial W_j} \right| \quad (2.25)$$

- Levenberg-Marquardt : $\alpha_k * D_k = -[J^T * J + \mu * I]^{-1} * J * e$ (2.26)

Similar al método Cuasi-Newton, el cual aproxima el gradiente y la matriz Hessiana por:

$$\begin{aligned} \nabla E_k &= J^{-T} * e \\ H &= J^T * J \end{aligned} \quad (2.27)$$

Donde:

e = Vector de errores de la red neuronal.

H = Aproximación de la matriz Hessiana.

J = Matriz Jacobiana (derivadas de primer orden)

Si $\mu = 0$, se llega al método de Newton usando la aproximación del Hessiano y si μ es grande se aproxima al método del gradiente descendente con baja tasa de aprendizaje.

El algoritmo usa el parámetro μ variable y lo disminuye o aumenta, según el valor que toma la función de desempeño.

2.2.4 Red de Hopfield

Una de las mayores contribuciones al área de las redes neuronales fue realizada en los años 1980 por John Hopfield, quien estudió modelos autoasociativos que presentaban algunas similitudes con los perceptrones, pero incluía también grandes diferencias.

Las redes de Hopfield son redes de adaptación probabilística, recurrentes, funcionalmente entrarían en la categoría de las memorias autoasociativas, que aprenden a reconstruir los patrones de entrada que memorizan durante el “proceso de almacenamiento” (equivalente a un entrenamiento). Son arquitecturas de una capa con interconexión total, funciones de activación booleana de umbral (cada unidad puede tomar dos estados, 0,1, dependiendo de si la estimulación total recibida supera determinado umbral), adaptación probabilística de la activación de las unidades, conexiones recurrentes y simétricas, y regla de aprendizaje no supervisado. Mientras que las redes no recurrentes dan soluciones estables, los modelos recurrentes dan soluciones inestables (dinámicas), lo que no siempre es aconsejable. El principal aporte de Hopfield consistió precisamente en conseguir que tales modelos recurrentes fueran así mismo estables. Imaginó un sistema físico capaz de operar como una memoria asociativa, que almacenara información y fuera capaz de recuperarla aunque la misma se hubiera deteriorado.

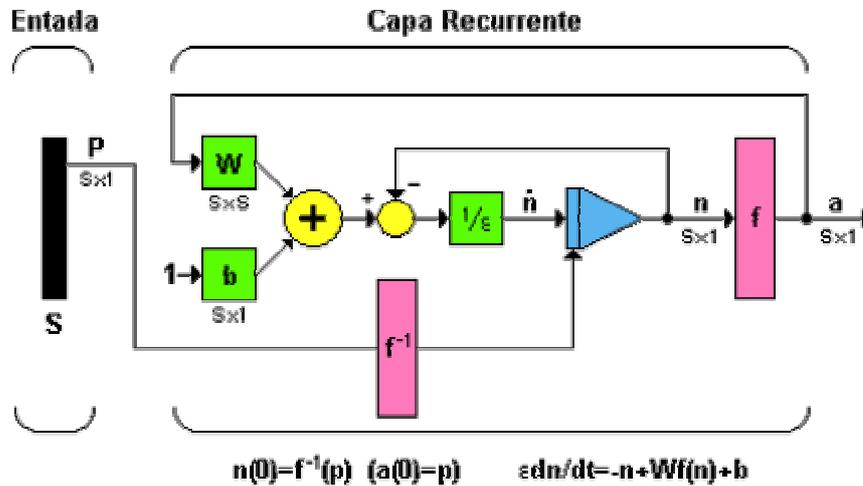
El concepto de memoria asociativa es bastante intuitivo, se trata simplemente de asociar dos patrones.

Dentro de este concepto se definen diferentes tipos de memorias asociativas:

- **Memoria heteroasociativa:** Establece una correspondencia F entre dos vectores X , Y de tal manera que $F(x_i)=y_i$, y si un x arbitrario está más próximo a x_i que a cualquier otro x_j , entonces $F(x)=y_i$. En esta definición, el estar más próximo quiere decir con respecto a la distancia de Hamming.
- **Memoria asociativa interpoladora:** Establece una correspondencia F entre X e Y de tal manera que $F(x)=y_i$, pero si el vector de entrada difiere de uno de los dos ejemplares en el vector d , de tal modo que $x= x_i+d$, entonces la salida será y_i+d .
- **Memoria autoasociativa:** Supóngase que $Y=X$, y se aplica la primera definición (memoria heteroasociativa).

La red de Hopfield en notación compacta se muestra en la figura 2.11 en donde el vector p no se considera como la entrada a la red sino como la condición inicial de la red.

Figura 2.11: Modelo red de Hopfield.



Como se observa la red de Hopfield está compuesta de neuronas dinámicas altamente interconectadas gobernadas por ecuaciones diferenciales no lineales, esta red funciona como una memoria asociativa no lineal que puede procesar patrones presentados de forma incompleta o con ruido, siendo útil como una poderosa herramienta de optimización.

La ecuación que describe el comportamiento de la red de Hopfield según el esquema mostrado es:

$$\begin{aligned} \epsilon * \frac{dn(t)}{dt} &= -n(t) + Wf(n) + b \\ n(0) &= f^{-1}(p) \\ a(0) &= p \end{aligned} \quad (2.28)$$

Para lograr la estabilidad de la red se aplica la teoría de estabilidad de Lyapunov, a través del teorema de La Salle y para su utilización el primer paso es escoger una función de Lyapunov, para lo cual Hopfield sugirió la siguiente función:

$$V(a) = -\frac{1}{2} a^T W a + \sum_{i=1}^S \left[\int_0^{a_i} f^{-1}(u) du \right] - b^T a \quad (2.29)$$

Donde a es la salida de la red, W es la matriz de pesos y b es el vector de ganancias.

La elección de esta particular función, fue clave en el desarrollo de Hopfield, pues el primer y el tercer termino de esta ecuación conforman una función cuadrática, las cuales pueden aproximar gran cantidad de funciones en un pequeño intervalo, especialmente cerca de puntos donde se encuentre un mínimo local.

Para usar el teorema de La Salle se necesita evaluar la derivada de la ecuación (2.29). Por claridad se evaluará cada uno de los tres términos de forma independiente, tomando la derivada del primer término de la ecuación (2.29) se obtiene:

$$\frac{d}{dt} \left[-\frac{1}{2} a^T W a \right] = -\frac{1}{2} * \nabla [a^T W a]^T \frac{da}{dt} = -[W a]^T \frac{da}{dt} = -a^T W \frac{da}{dt} \quad (2.30)$$

Derivando el segundo termino de la ecuación (2.29), el cual consiste de una sumatoria de integrales y considerando una de estas integrales se obtiene:

$$\frac{d}{dt} \left[\int_0^{a_i} f^{-1}(u) du \right] = \frac{d}{da_i} \left[\int_0^{a_i} f^{-1}(u) du \right] \frac{da_i}{dt} = f^{-1}(a_i) \frac{da_i}{dt} = n_i \frac{da_i}{dt} \quad (2.31)$$

Tomando en consideración todas las integrales, en forma matricial la derivada del segundo término es:

$$\frac{d}{dt} \left[\sum_{i=1}^s \left[\int_0^{a_i} f^{-1}(u) du \right] \right] = n^T \frac{da}{dt} \quad (2.32)$$

Derivando el tercer término de la ecuación (2.29) y apoyándose en las propiedades de las funciones cuadráticas se obtiene la ecuación (2.33).

$$\frac{d}{dt} [-b^T a] = -\nabla [b^T a]^T \frac{da}{dt} = -b^T \frac{da}{dt} \quad (2.33)$$

La derivada total de la ecuación (2.29) se obtiene al unir los resultados de las ecuaciones (2.30), (2.32) y (2.33).

$$\frac{d}{dt} V(a) = -a^T W \frac{da}{dt} + n^T \frac{da}{dt} - b^T \frac{da}{dt} = [-a^T W + n^T - b^T] \frac{da}{dt} \quad (2.34)$$

Comparando con la ecuación (2.28) del modelo de Hopfield, se tiene que:

$$[-a^T W + n^T - b^T] \frac{da}{dt} = -\varepsilon \left[\frac{dn(t)}{dt} \right]^T \quad (2.35)$$

Esto permite describir la ecuación (2.34) como sigue:

$$\frac{d}{dt}V(a) = -\varepsilon \left[\frac{dn(t)}{dt} \right]^T \frac{da}{dt} = -\varepsilon \sum_{i=1}^s \left(\frac{dn_i}{dt} \right) \left(\frac{da_i}{dt} \right) \quad (2.36)$$

Ya que $n_i = f^{-1}(a_i)$, es posible expandir la derivada de n_i de la siguiente forma:

$$\frac{dn_i}{dt} = \frac{d}{dt} [f^{-1}(a_i)] = \frac{d}{da_i} [f^{-1}(a_i)] \frac{da_i}{dt} \quad (2.37)$$

Con esto la ecuación (2.36) puede ser rescrita como:

$$\frac{d}{dt}V(a) = -\varepsilon \sum_{i=1}^s \left(\frac{dn_i}{dt} \right) \left(\frac{da_i}{dt} \right) = -\varepsilon \sum_{i=1}^s \left(\frac{d}{da_i} [f^{-1}(a_i)] \right) \left(\frac{da_i}{dt} \right)^2 \quad (2.38)$$

Si se supone que $f^{-1}(a_i)$ es una función incremental, entonces:

$$\frac{d}{da_i} [f^{-1}(a_i)] > 0 \quad (2.39)$$

Este resultado implica en la ecuación (2.39) que:

$$\frac{d}{dt}V(a) \leq 0 \quad (2.40)$$

De esta manera, si $f^{-1}(a_i)$ es una función incremental, todos los valores propios de la función $dV(a)/dt$ son no positivos lo cual implica que la red sea estable, entonces $V(a)$ es una función de Lyapunov válida.

Los atractores de Hopfield son puntos estacionarios de la función de Lyapunov que satisfacen la ecuación (2.41).

$$\frac{da}{dt} = 0 \quad (2.41)$$

Estos puntos estacionarios son puntos donde se encuentra un mínimo de la función $V(a)$ descrita en la ecuación (2.29), en estos puntos el gradiente de la función $V(a)$ es igual a cero.

$$\nabla V(a) = \left[\frac{\partial V}{\partial a_1} \quad \frac{\partial V}{\partial a_2} \quad \dots \quad \frac{\partial V}{\partial a_s} \right]^T = 0 \quad (2.42)$$

La función de Lyapunov descrita por la ecuación (2.29) puede simplificarse si se considera que la ganancia γ es grande, como sucede en los amplificadores con los que se implementa la

red, una función de transferencia típica para estos amplificadores no lineales se muestra a continuación:

$$a = f(n) = \frac{2}{\pi} \tan^{-1} \left(\frac{\gamma n}{2} \right) \quad (2.43)$$

Para evaluar el segundo término de la función de Lyapunov se requiere el cálculo de $f^{-1}(u)$.

$$f^{-1}(u) = \frac{2}{\gamma\pi} \tan \left(\frac{\pi u}{2} \right) \quad (2.44)$$

Si la ganancia γ es muy grande y la salida de la red se mantiene en el rango $-1 < a < 1$, el segundo término de la función de Lyapunov tiende a cero y puede definirse la función de alta ganancia de Lyapunov como:

$$V(a) = -\frac{1}{2} a^T W a - b^T a \quad (2.45)$$

2.2.4.1 Regla de Aprendizaje

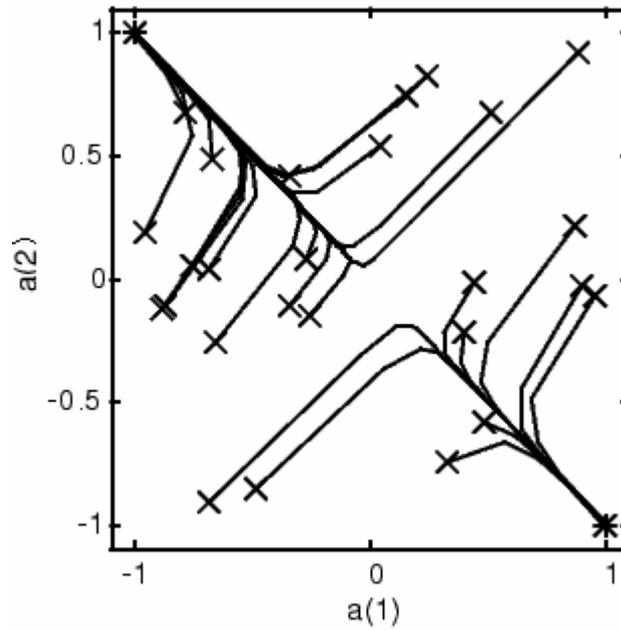
La red de Hopfield no tiene una ley de aprendizaje asociada, esto significa que la red no es entrenada ni realiza un proceso de aprendizaje, sin embargo es posible determinar la matriz de pesos por medio de un procedimiento basado en la función de alta ganancia de Lyapunov descrita por la ecuación (2.45).

El procedimiento consiste en escoger la matriz de pesos W y el vector de ganancias b tal que V toma la forma de la función que se quiere minimizar, convirtiendo el problema que se quiere resolver, en un problema de minimización cuadrática, puesto que la red de Hopfield minimizará a V .

Una red de Hopfield puede diseñarse como una memoria asociativa, en este caso es llamada memoria de contenido direccionable, porque la memoria recupera la información almacenada con base en parte de su contenido.

Cuando se le presenta un patrón de entrada a la red de Hopfield, el estado inicial de la salida es el mismo patrón de entrada y luego la red converge al patrón prototipo almacenado que se encuentra más cercano (o que más se parece) al patrón de entrada. Por lo tanto, el patrón de entrada sigue una trayectoria en que pasa por diferentes estados dependiendo del número de iteraciones necesarias para alcanzar la convergencia, tiempo en que alcanza algún patrón almacenado. Ver figura 2.12

Figura 2.12: Espacio de trayectorias de la red de Hopfield en un modelo bidimensional.



Para que la red memorice un patrón prototipo, este debe ser un mínimo de la función de Lyapunov.

Se asume que los patrones prototipo son $\{p_1, \dots, p_q\}$ y que cada uno de estos vectores se compone de S elementos, al asumir que $Q \ll S$, el espacio de estado es amplio y los patrones prototipo se encuentran bien distribuidos y por lo tanto no están cercanos uno de otro.

Para garantizar que los patrones prototipo a almacenar son mínimos de la función de Lyapunov, se propone la siguiente función para evaluar el error en la aproximación.

$$J(a) = -\frac{1}{2} \sum_{q=1}^Q ([p_q]^T a)^2 \quad (2.46)$$

Si los elementos de a son restringidos a valores de ± 1 , la función es minimizada en los patrones prototipo como se muestra a continuación:

Suponiendo que los patrones prototipo son ortogonales, y evaluando el error en uno de ellos, se tendrá que:

$$J(a) = -\frac{1}{2} \sum_{q=1}^Q ([p_q]^T p_j)^2 = -\frac{1}{2} ([p_q]^T p_j)^2 = -\frac{S}{2} \quad (2.47)$$

La segunda igualdad de la ecuación (2.47) se debe a la ortogonalidad de los patrones prototipo y la última igualdad a que todos los elementos de p_j son ± 1 . Evaluando el error del patrón

aleatorio de entrada, el cual presumiblemente no está cercano a ningún patrón prototipo, cada elemento de la sumatoria en la ecuación (2.46) es el producto punto entre un patrón prototipo y la entrada. El producto punto se incrementa cuando la entrada se mueve cerca del patrón prototipo, sin embargo, si la entrada no se encuentra cerca de algún patrón prototipo, todos los términos de la sumatoria serán pequeños y por lo tanto $J(a)$ será la mayor (menos negativa) y cuando a sea igual a alguno de los patrones prototipo $J(a)$ será mas pequeña (más negativa).

La ecuación (2.46) es una función cuadrática que indica con precisión el desempeño del contenido de la memoria direccionable, el próximo paso es escoger la matriz de pesos W y ganancias b , tal que la función de Lyapunov de Hopfield V sea equivalente al desempeño de la función cuadrática J .

Para calcular la matriz de pesos (con patrones objetivo iguales a los patrones de entrada) se utiliza:

$$W = \sum_{q=1}^Q p_q(p_q)^T \quad Y \quad b = 0 \quad (2.48)$$

Entonces la función de Lyapunov será:

$$V(a) = -\frac{1}{2} \left[\sum_{q=1}^Q p_q(p_q)^T \right] a = -\frac{1}{2} \sum_{q=1}^Q a^T p_q(p_q)^T a \quad (2.49)$$

Y puede ser reescrita como:

$$V(a) = -\frac{1}{2} \sum_{iq=1}^Q [(p_q)^T a]^2 = J(a) \quad (2.50)$$

Se observa que la función de Lyapunov es igual al desempeño del error del contenido de la memoria direccionable, la salida de la red de Hopfield tenderá a converger a los patrones prototipo almacenados. En el caso que todos los patrones prototipo sean ortogonales, cada uno será un punto de equilibrio de la red, la cual puede tener muchos otros puntos de equilibrio indeseables, una regla práctica para evitarlos consiste en que el número de patrones almacenados no debe superar en gran cantidad el número de neuronas de la red, dicha cantidad es cuantificable en base a pruebas.

2.3 Selección de Modelos y configuraciones de Redes

Es un problema fundamental saber de antemano cuan grande debe ser una red para aplicación específica, por ello existen dos herramientas al alcance, las cuales permiten mejorar la generalización de la red o dicho de otro modo, evitar que se produzcan errores pequeños en el entrenamiento y errores lo suficientemente mayores en la validación, éstas son:

- Regularización.
- Validación Temprana.

La regularización permite a través de la modificación de la función de desempeño (error cuadrático medio), obtener pesos de menor magnitud, lo cual fuerza a que la respuesta de la red sea más suave y a la vez reduce la probabilidad de que se produzca sobreentrenamiento.

La nueva función de desempeño, involucra un término que corresponde al promedio de la suma de los cuadrados de los pesos y sesgos de la red, según sigue:

$$msereg = \gamma * mse + (1 - \gamma) * msw$$
$$msw = \frac{1}{n} * \sum_{j=1}^n W_j^2 \quad (2.51)$$

Donde γ es la razón de desempeño y debe ser un valor tal que no produzca sobre ajuste (demasiado grande) y tampoco muy pequeño, ya que la red no ajustará de manera adecuada los datos del entrenamiento.

En el método de detección temprana la idea base es que el número de parámetros efectivos (aquellos cuyo valor difiere significativamente de cero) crece en el curso del aprendizaje. Al detener el entrenamiento cuando se eleva el error de validación se limita la complejidad de la red (Estévez, 1999).

Los datos se dividen en tres conjuntos: Entrenamiento, validación y prueba.

El segundo es por lo tanto para establecer el tiempo al cual el entrenamiento de la red debe detenerse, tal tiempo se denomina número de épocas de entrenamiento.

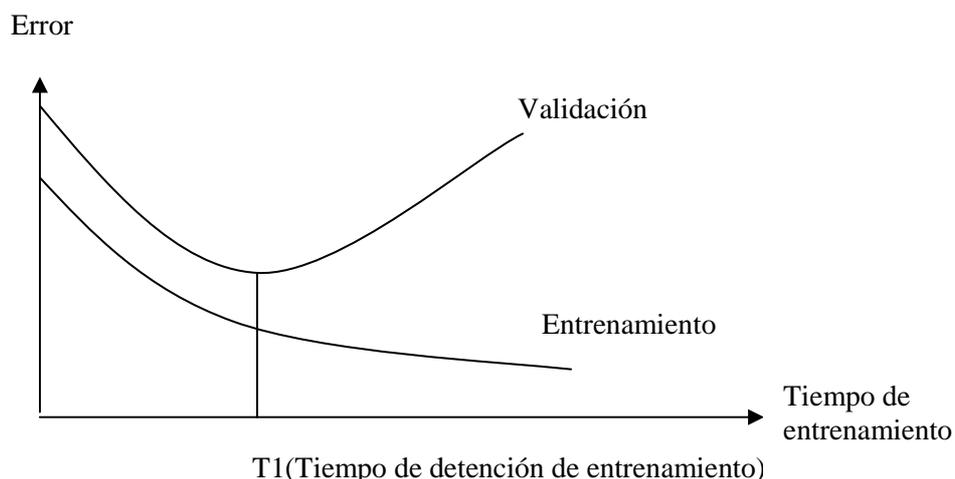
El tiempo al cual el entrenamiento debe detenerse, se define examinando los errores de validación en base al número de épocas de entrenamiento y número de neuronas en las capas ocultas, por medio de lo cual es posible determinar el óptimo de épocas con el objetivo minimizar el error de validación.

La selección del número de neuronas en la (s) capa (s) oculta (s) consiste en un proceso de prueba - análisis de tal manera que se deben escoger diversas arquitecturas neuronales de acuerdo a un análisis previo de la complejidad del problema y necesidad de procesamiento de la información. Aquella arquitectura que presente el menor error ponderado de entrenamiento, validación y prueba, es denominada arquitectura de red más satisfactoria u “óptima” en términos de minimizar los errores de pronóstico. Como condición inicial para probar arquitecturas se parte de la base de problemas de similares características y su arquitectura de red, considerando que fenómenos complejos o con un rápido tiempo de respuesta requieren de más de una capa oculta.

El tercer conjunto de datos prueba se utiliza para verificar la capacidad de generalización de la red neuronal ajustada.

En la figura 2.13 se observa el comportamiento típico de los errores de entrenamiento y validación, en el cual se observa que dichos errores decrecen durante la fase inicial del entrenamiento, pero cuando la red comienza a sobre-ajustar los datos, el error de validación comienza a aumentar. Luego, cuando el error de validación se incrementa durante un determinado número de iteraciones, el entrenamiento se detiene.

Figura 2.13: Errores de entrenamiento y validación durante el proceso de entrenamiento de la red neuronal.



2.4 Identificación de Sistemas mediante Modelos Paramétricos

Estos métodos requieren la elección de una posible estructura del modelo, de un criterio de ajuste de parámetros, y por último de la estimación de los parámetros que mejor ajustan el modelo a los datos observados.

A continuación se presentan los métodos de identificación existentes:

1. Métodos de identificación off-line (a posteriori), son utilizados en aquellas aplicaciones en que no se requiera un ajuste continuado del modelo. En estos casos, se realiza la identificación previa del problema, considerándose que la validez de los parámetros obtenidos no se verá alterada con el paso del tiempo.

2. Métodos de identificación on-line (identificación recursiva), son aquellos en los que los parámetros se van actualizando continuamente a partir de los nuevos datos de entrada-salida obtenidos durante la evolución del proceso.

Dependiendo del criterio de ajuste de los parámetros, existen diversos métodos matemáticos para ajustar los parámetros de una estructura a un conjunto de datos de entrada-salida. Algunos de los más utilizados en el campo de la identificación son el método de mínimos cuadrados y el método de las variables instrumentales.

La expresión más general de un modelo paramétrico es de la forma:

$$s(t) = \eta(t) + w(t) \quad (2.52)$$

Donde $w(t)$ es el término que modela la salida debida a las perturbaciones, $\eta(t)$ la salida debida a la entrada, y $s(t)$ la salida medible del sistema. Cada uno de estos términos puede desarrollarse de la siguiente forma:

$$\begin{aligned} \eta(t) &= G(q^{-1}, \theta)u(t) \\ w(t) &= H(q^{-1}, \theta)e(t) \\ s(t) &= A(q^{-1}, \theta)y(t) \end{aligned} \quad (2.53)$$

Donde q^{-1} es el operador retardo, θ representa un vector de parámetros, $u(t)$ y $e(t)$ son la entrada al sistema y el ruido de entrada al mismo respectivamente e $y(t)$ es la salida de interés del sistema (que puede no coincidir con la salida medible).

Tanto $G(q^{-1}, \theta)$ como $H(q^{-1}, \theta)$ son polinomios del tipo:

$$\begin{aligned} G(q^{-1}, \theta) &= \frac{B(q^{-1})}{F(q^{-1})} = \frac{b_1 * q^{-nk} + b_2 * q^{-nk-1} + \dots + b_{nb} * q^{-nk-nb+1}}{1 + f_1 * q^{-1} + \dots + f_{nf} * q^{-nf}} \\ H(q^{-1}, \theta) &= \frac{C(q^{-1})}{F(q^{-1})} = \frac{c_1 * q^{-nk} + c_2 * q^{-nk-1} + \dots + c_{nb} * q^{-nk-nb+1}}{1 + d_1 * q^{-1} + \dots + d_{nf} * q^{-nf}} \end{aligned} \quad (2.54)$$

Y $A(q^{-1}, \theta)$ un polinomio del tipo:

$$A(q^{-1}, \theta) = 1 + a_1 * q^{-1} + \dots + a_{na} * q^{-na} \quad (2.55)$$

El vector de parámetros θ contiene los coeficientes a_i , b_i , c_i , d_i y f_i de las funciones de transferencia anteriores. La estructura genérica de estos modelos es por tanto:

$$A(q^{-1}) * y(t) = G(q^{-1}, \theta) * u(t) + H(q^{-1}, \theta) * e(t) = \frac{B(q^{-1})}{F(q^{-1})} * u(t) + \frac{C(q^{-1})}{H(q^{-1})} * e(t) \quad (2.56)$$

Para elegir la estructura de este tipo de modelos hay que determinar el orden de cada uno de los polinomios anteriores, es decir n_a , n_b , n_c , n_d , n_f y el retardo entre la entrada y la salida n_k . Una vez elegidos estos valores, sólo queda determinar el vector de coeficientes q (a_i , b_i , c_i , d_i y f_i) que hacen que el modelo se ajuste a los datos de entrada - salida del sistema real.

La anulación de alguno de los polinomios, resultando estructuras simplificadas, facilita el proceso de ajuste de parámetros. Cada una de las estructuras que se forman tiene sus propias características y debe ser elegida fundamentalmente en función del punto en el que se prevé que se añade el ruido en el sistema. En cualquier caso, es necesario ensayar con varias estructuras y con varios órdenes dentro de una misma estructura hasta encontrar un modelo satisfactorio.

Tabla 2.1: Diferentes Estructuras de modelos paramétricos.

Tipo de modelo	Condición	Estructura resultante
Modelo ARX	$F(q^{-1})=D(q^{-1})=C(q^{-1})=1$	$A(q^{-1})y(t) = B(q^{-1}) \cdot u(t) + e(t)$
Modelo Output Error (OE)	$C(q^{-1})=D(q^{-1})=A(q^{-1})=1$	$y(t) = \frac{B(q^{-1})}{F(q^{-1})} \cdot u(t) + e(t)$
Modelo ARMAX	$F(q^{-1})=D(q^{-1})=1$	$A(q^{-1}) \cdot y(t) = B(q^{-1}) \cdot u(t) + C(q^{-1}) \cdot e(t)$
Modelo Box Jenkins (BJ)	$A(q^{-1})=1$	$y(t) = \frac{B(q^{-1})}{F(q^{-1})} \cdot u(t) + \frac{C(q^{-1})}{D(q^{-1})} \cdot e(t)$

2.4.1 Algoritmos de ajuste de parámetros

Para determinar el valor de los parámetros que identifican el sistema considerado se ajusta la respuesta del modelo a los datos de entrada - salida experimentales.

Todo modelo matemático es capaz de predecir el valor de la salida del sistema en función de las entradas y salidas en instantes anteriores. Se llama error de predicción $\varepsilon(t, \theta)$ a la diferencia entre la salida estimada por el modelo y la salida real del sistema en un determinado instante de tiempo:

$$\varepsilon(t, \theta) = y(t) - y_e(t, \theta) \quad (2.57)$$

Donde $y_e(t)$ es la salida estimada por el modelo en el instante t .

Por otra parte, se dice que una estructura posee regresión lineal cuando la salida estimada puede expresarse como:

$$y_e(t, \theta) = \varphi^T(t) * \theta \quad (2.58)$$

Donde $\varphi^T(t)$ es un vector columna formado por las salidas y entradas anteriores (conocido como vector de regresión), y θ es el vector de parámetros del modelo.

Luego, la expresión del error de pronóstico aplicando los criterios fijados anteriormente es:

$$\varepsilon(t, \theta) = y(t) - \varphi^T(t) * \theta \quad (2.59)$$

Se define la siguiente función de error:

$$V_N(\theta) = \frac{1}{N} * \sum_{t=1}^N \frac{1}{2} [y(t) - \varphi^T(t) * \theta]^2 \quad (2.60)$$

Conocida como **criterio de mínimos cuadrados** para una regresión lineal.

Existe un valor de θ que minimiza la función anterior y que constituye la estimación del modelo por mínimos cuadrados:

$$\theta_{LSE} = sol \left\{ \frac{1}{N} * \sum_{t=1}^N \varphi^T(t) * [y(t) - \varphi^T(t) * \theta] = 0 \right\} \quad (2.61)$$

Para este vector de parámetros, la función de error V_N toma su valor mínimo, siendo éste la función de pérdidas del modelo estimado.

Los métodos de identificación descritos, pueden implementarse como recursivos.

Un algoritmo de identificación recursiva típico es el siguiente:

$$\theta(t) = \theta(t-1) + K(t) * (y(t) - y_e(t)) \quad (2.62)$$

Donde $\theta(t)$ es el vector de parámetros estimado en el instante t , $y(t)$ la salida real del sistema en dicho instante de tiempo, $y_e(t)$ la salida estimada con los parámetros actuales, y $\theta(t-1)$ el vector de parámetros del modelo en el instante de tiempo anterior. De esta forma, $K(t)$ determina el modo en que el error de predicción $(y(t)-y_e(t))$ afecta en la actualización on-line de los parámetros del modelo.

$$K(t) = Q(t) * \psi(t) \quad (2.63)$$

Donde $\psi(t)$ representa el gradiente de la salida estimada respecto al vector de parámetros $y_e(t|\theta)$. Por otro lado, la matriz $Q(t)$ puede ser escogida según diferentes criterios, dando lugar a diferentes métodos de adaptación.

Suponiendo que el modelo puede escribirse como una regresión lineal, uno de los métodos más sencillos de escoger la matriz $Q(t)$ está basado en el filtro de Kalman, dando lugar al siguiente algoritmo:

$$\begin{aligned} \theta(t) &= \theta(t-1) + K(t) * (y(t) - y_e(t)) \\ y_e(t) &= \psi^T(t) * \theta(t-1) \\ K(t) &= Q(t) * \psi(t) \end{aligned} \quad (2.64)$$

$$\begin{aligned} \Rightarrow \theta(t) &= [1 - K(t) * \psi^T(t)] * \theta(t-1) + K(t) * y(t) \\ \Leftrightarrow \theta(t) &= H(t) * \theta(t-1) + Z(t) \end{aligned}$$

Por tanto, el filtro de Kalman permite utilizar la información incompleta del modelo para mejorar de forma recursiva la estimación del estado del sistema proporcionada por la medición.

En cada paso el filtro proyecta la estimación del estado actual y de la covarianza actual hacia delante en el tiempo para obtener una estimación a priori para el siguiente paso. Después utiliza los resultados de la medición real para mejorar esta estimación y obtener una estimación a posteriori. Este proceso puede verse también como un ciclo de predicción corrección.

2.5 Metodología General

La metodología a seguir, que se esquematiza en la figura 2.14, se describe a través de los siguientes pasos. Cabe destacar que se utiliza MATLAB 7.0 como programa de modelación de redes neuronales, usando el TOOLBOX de redes neuronales.

2.5.1 Metodología de Pronóstico

A- Definir variables de entrada relevantes en el proceso.

El tipo de entradas a considerar depende de las variables que se desea pronosticar. Para pronósticos de crecidas pluviales como de deshielo son necesarias las siguientes variables de entrada:

-Caudales horarios.

-Precipitación horaria.

-Indirectamente temperaturas horarias (Influyente en la ubicación de la línea de Nieve, separación de área pluvial y nival).

Para el pronóstico de caudales en el caso de crecidas de deshielo se requiere además contar con datos de temperaturas directamente para representar la influencia de la radiación.

B- Definir variables de salida.

La variable a pronosticar es el caudal en un tiempo posterior de tipo $(t+k)$, con $k \geq 1$.

C- Selección y Estandarización de datos

Considerando el tipo de funciones de transferencia que utilizan las neuronas de las capas ocultas, es necesario escalar los datos para evitar problemas de saturación de la función de transferencia a utilizar. Las transformaciones son las que siguen, de las cuales se utiliza mayormente la segunda, ya que permite conservar el signo de la variable, útil especialmente cuando se habla de gradientes.

$$x_i' = 0,9 * \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} + 0,05 \quad x_i' = \frac{x_i}{1.1 * \max(x_i)} \quad (2.65)$$

Donde:

x_i = Valor observado $i=1, \dots, n$.

$\min(x_i)$ = Mínimo valor de los datos observados.

$\max(x_i)$ = Máximo valor de los datos observados.

D- Elección de Red Neuronal.

Elección de la red neuronal de acuerdo al problema específico y métodos de ajuste de los pesos asociados a las neuronas de la red por medio de algoritmos de entrenamiento (redes recurrentes y prealimentadas) y condiciones que garanticen su estabilidad (red de Hopfield).

Las entradas y el número de neuronas de salida vienen determinados por la naturaleza del problema y por el tipo de secuencias a procesar. Otros valores como el número de neuronas de estado o el orden de entrada deben ser determinados tras experimentar con distintas combinaciones. Se construyen diversas arquitecturas con distintas configuraciones.

E- Entrenamiento de la red neuronal recurrente.

Este paso tiene como función determinar el valor de los pesos de la red que permiten resolver el modelo con éxito. Se considera como medida de error a minimizar, el error cuadrático medio o una variación de él (regularizar), que describe la adecuación de la salida proporcionada por la red al valor deseado. Cabe señalar que se utiliza un entrenamiento fuera de línea.

En el caso de una red prealimentada cualquier tipo de entrenamiento resulta válido, no necesariamente óptimo, pero se permite utilizar cualquier método de ajuste de los pesos asociados a las neuronas dada la independencia de un dato o patrón con respecto al anterior. En el caso de una red recurrente dicha aseveración no es completamente cierta, puesto que existe un almacenamiento de la información precedente entregada a la red, lo cual puede generar “ruido” en caso de que tal información ya no sea del set de datos que se desea pronosticar.

Nace por tanto un inconveniente, que es el hecho de permitir o no un entrenamiento que contenga una serie de crecidas entregadas a la red de forma continua, sin que dicha continuidad sea del todo cierta, ya que cada vez que se ingresa una crecida existe un salto en la serie de tiempo y por lo tanto la red almacena información no relevante para la crecida actual. Es fundamental entonces, un análisis de los pesos asociados a las variables retardadas y su convergencia a un valor nulo, de tal manera que el tiempo o iteración al que dicha matriz de pesos es anulada, indica la cantidad de variables retardadas que son relevantes para la red.

El análisis descrito, se considera como una condición necesaria y suficiente antes de llevar a cabo el proceso de pronóstico, razón por la cual es descrito dentro de este sub-capítulo indicado como “metodología”. Es necesario entonces, entrenar la red recurrente de Elman con las series continuas seleccionadas para cada una de las cuencas en que se pronosticaran caudales y deducir la relevancia de las variables retardadas.

F- Validación y elección de configuración y arquitectura de red recurrente.

Se seleccionan las arquitecturas que entreguen el menor error cuadrático medio en el conjunto de validación. Para escoger el modelo o configuración óptima se analiza el error cuadrático medio ponderado.

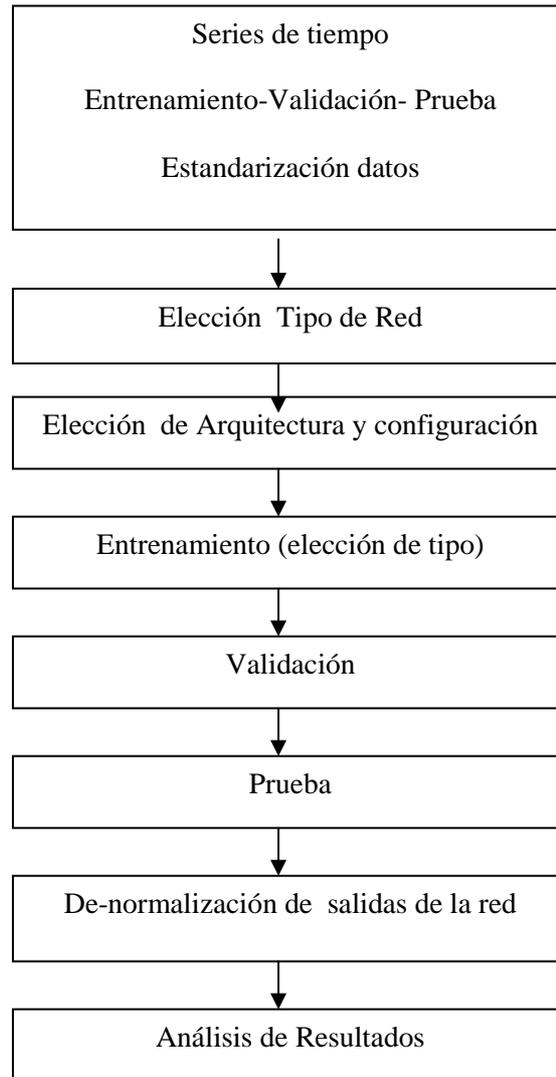
G- Pronóstico de caudales instantáneos.

Por medio del modelo elegido, se realiza pronósticos de caudal para distintos horizontes de tiempo ($t+k$; $k=1,2,\dots,T$).

H- Análisis de pronósticos.

Se analizan y comparan, tanto en calidad como en valor, los pronósticos obtenidos por medio del modelo seleccionado con modelos de redes neuronales prealimentadas y métodos de identificación paramétrica (ARMAX y RARMAX). Cabe señalar que cuando se realicen comparaciones de diferentes modelos formados a partir de distintos tipos de entrenamiento con modelos prealimentados con un entrenamiento de tipo continuo, se está comparando los diferentes análisis realizados con aquel estudio y pronóstico continuo de redes prealimentadas, planteado anteriormente por González, 2003.

Figura 2.14: Esquema metodológico.

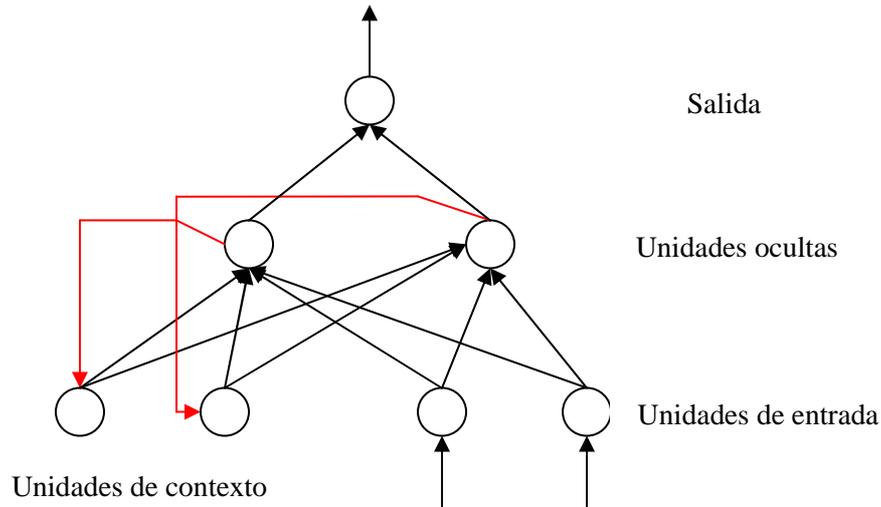


2.5.2 Condición de aplicación

Según se describió en el punto 2.5.1, punto E, correspondiente al entrenamiento de la red neuronal recurrente de Elman, se realiza el análisis de los pesos asociados a las variables retardadas, condición necesaria y suficiente para permitir su aplicación o no al pronóstico de crecidas en tiempo real.

A continuación en la figura 2.15, se presenta un esquema simplificado de la arquitectura de la red en la cual es posible apreciar sus unidades e interconexiones, las líneas rojas definen el almacenamiento de la salida de las unidades de la capa oculta en las unidades de contexto, el paso de la salida de la información contenida en estas últimas a la capa oculta queda ponderado por el correspondiente peso asociado a las variables retardadas.

Figura 2.15: Esquema Red Neuronal Recurrente de Elman.



Si bien, los fenómenos de precipitaciones, temperaturas, etc. que producen una determinada crecida no son relevantes para una crecida posterior, sí guardan relación con las condiciones en las cuales se encuentre la cuenca al momento de producirse la siguiente crecida, como son las características de humedad, grado de saturación y características físicas, que a su vez modifican las condiciones de infiltración y por ende de escorrentía directa, punto desde el cual es posible considerar una influencia antecesora a la crecida en un límite de tiempo que considere dichos efectos. Si bien es cierto que las redes recurrentes almacenan información de manera recursiva, es un punto relevante cuantificar la importancia que tendría para datos actuales lo ocurrido en pasados muy remotos de acuerdo a los pesos que asigne la red.

Dado que las crecidas seleccionadas, están separadas tanto por días, meses o años, en general no existe relación entre una y otra, en casi todos los casos ni siquiera en las condiciones iniciales en que se encuentra la cuenca, puesto que en general el tiempo de desfase entre crecidas seleccionadas es como se especifica de meses o años.

Por lo tanto, si es posible demostrar que la influencia de lo ocurrido con anterioridad, en general sólo tiene relevancia dentro del proceso “crecida”, es posible “pegar las crecidas” (serie completa) aunque ésta contenga saltos, ya que se estaría ocupando el beneficio de la red en el fenómeno individual, sin dar importancia a lo ocurrido en tiempos pasados o remotos a la crecida.

En pronósticos de crecidas pluviales es importante el tiempo de concentración de la cuenca, en el sentido de producir un retardo en la variable precipitación con respecto al caudal de salida de la cuenca, y en el caso de crecidas de deshielo es importante el tiempo de desfase

entre la ocurrencia de la temperatura máxima y el caudal máximo, de tal manera de retardar esta última variable, por ello en el caso de pronósticos mediante redes prealimentadas la incorporación de variables retardadas debe ser estrictamente necesaria dadas las características del problema, no así en el caso de una red neuronal recurrente de Elman, donde los retardos son implícitos y vienen otorgados por las ventajas de la red, pero pueden darse dos inconvenientes otorgar relevancia a lo ocurrido en pasados remotos ó que el límite en el cual otorga relevancia a dichas variables diste de una interpretación física.

Se realiza entonces, el análisis de los pesos asociados a las variables retardadas de la red de Elman, cabe destacar que los modelos de pronóstico de crecidas pluviales en las cuencas analizadas requieren de dos capas ocultas, por lo tanto existen dos operadores de retardo, sin embargo sólo importa el primero debido a que el retardo de la segunda capa oculta ya ha procesado las variables de salida de la primera capa de acuerdo a un cierto peso asignado y retronado dicha información a la primera capa oculta.

En la figura 2.16 se observa que tras 5 iteraciones la matriz de pesos anula sus valores, es decir si a la red entra un determinado vector por ejemplo en el caso de crecidas pluviales de la forma $(P(t), Q(t), \Delta Q(t))$ este vector es entregado a la capa oculta y posteriormente retardado con la ponderación de la matriz de pesos, para el siguiente pronóstico las variables de entrada son $(P(t+1), Q(t+1), \Delta Q(t+1))$, la red asigna pesos a esta información, la cual es sumada con el vector retardado ponderado por los respectivos pesos, y así para el siguiente el vector de entrada $(P(t+2), Q(t+2), \Delta Q(t+2))$, nuevamente la red asigna pesos a esta información y es sumada al retardo $(P(t+1), Q(t+1), \Delta Q(t+1))$ por la respectiva matriz de pesos, que a su vez ya incorpora el retardo de $(P(t), Q(t), \Delta Q(t))$ por la correspondiente matriz de pesos, entonces lo que va sucediendo es que la matriz de pesos se va multiplicando conforme los retardos se vayan considerando para las variables de salida, luego, de acuerdo al análisis realizado el umbral de truncamiento de importancia de las variables de entrada en este caso el volumen de precipitación, el caudal y el gradiente de este último es hasta 5 horas atrás, lo cual no dista mucho del valor del tiempo de concentración en las cuencas analizadas.

En conclusión, el umbral de truncamiento de importancia de las variables retardadas es óptimo en términos del número de datos de una determinada crecida y de los tiempos de concentración encontrados para ambas cuencas (aprox. 3 hrs.), luego el realizar entrenamientos continuos en la red de Elman tiene validez.

Cabe destacar que se realiza el mismo proceso para el caso de crecidas de deshielo, donde las variables de entrada relevantes del proceso son: caudal desfasado en 24 horas con respecto al

instante de pronósticos (para pronósticos a $t + 1$), caudal (t) y $T (t-D+1)$ (donde D es el desfase de ocurrencia entre los máximos de temperatura y caudal). Se obtiene que la matriz de pesos se anula aprox. a las 15 multiplicaciones, por lo tanto para un determinado pronóstico se está tomando relevancia a lo ocurrido aprox. hasta 15 horas antes, según se aprecia en la figura 2.18, lo cual es válido tanto en la cuenca del Estero Arrayán como en la cuenca del Mapocho en Los Almendros.

En la figura 2.17 se observa los pesos asociados a las variables de entrada mencionadas anteriormente relevantes para crecidas pluviales para el modelo NEWELM 3-7-7-1 analizado para la cuenca del Estero Arrayán en la Montosa; es una representación radial y cada uno de los ejes representa una neurona de la capa oculta, se observa que la mayor ponderación es otorgada a la variable caudal antecedente, luego a la variable volumen de precipitación y finalmente y variable en cada neurona al gradiente de caudal con respecto al instante anterior. La misma importancia otorgada a dichas variables es encontrada para crecidas pluviales en la cuenca del Mapocho en Los Almendros.

Figura 2.16: Análisis de pesos asociados a las variables de retardo. Crecidas Pluviales.

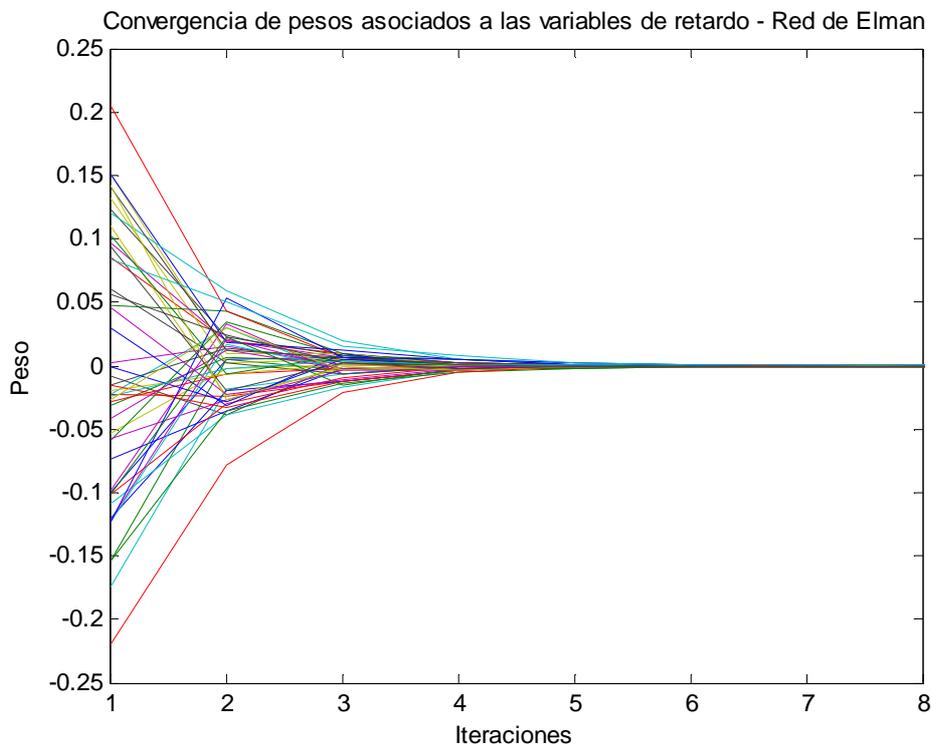
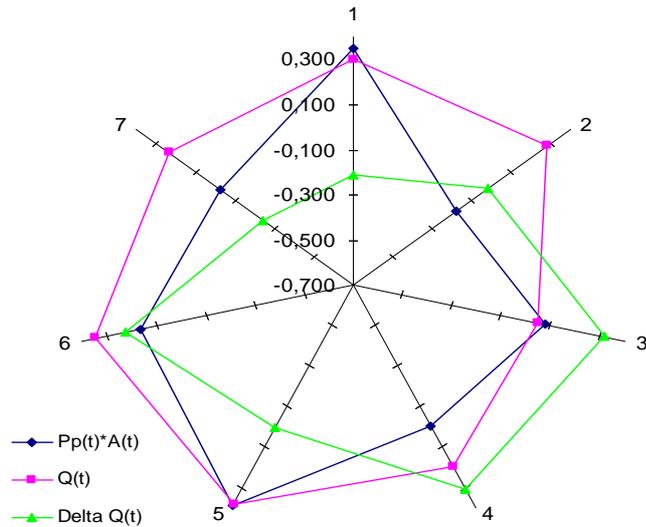


Figura 2.17: Pesos Asociados a las variables de entrada a la red. Crecidas Pluviales.



En la figura 2.19 se observa los pesos asociados a las variables de entrada mencionadas anteriormente, que son relevantes para crecidas deshielo para el modelo NEWELM 4-15-1 analizado para la cuenca del Estero Arrayán en la Montosa; es una representación radial y cada uno de los ejes representa una neurona de la capa oculta, no se observa alguna tendencia de ponderación clara de las variables de entrada a la red, lo mismo se obtiene para la cuenca del río Mapocho en Los Almendros.

Figura 2.18: Análisis de pesos asociados a las variables de retardo. Crecidas de Deshielo

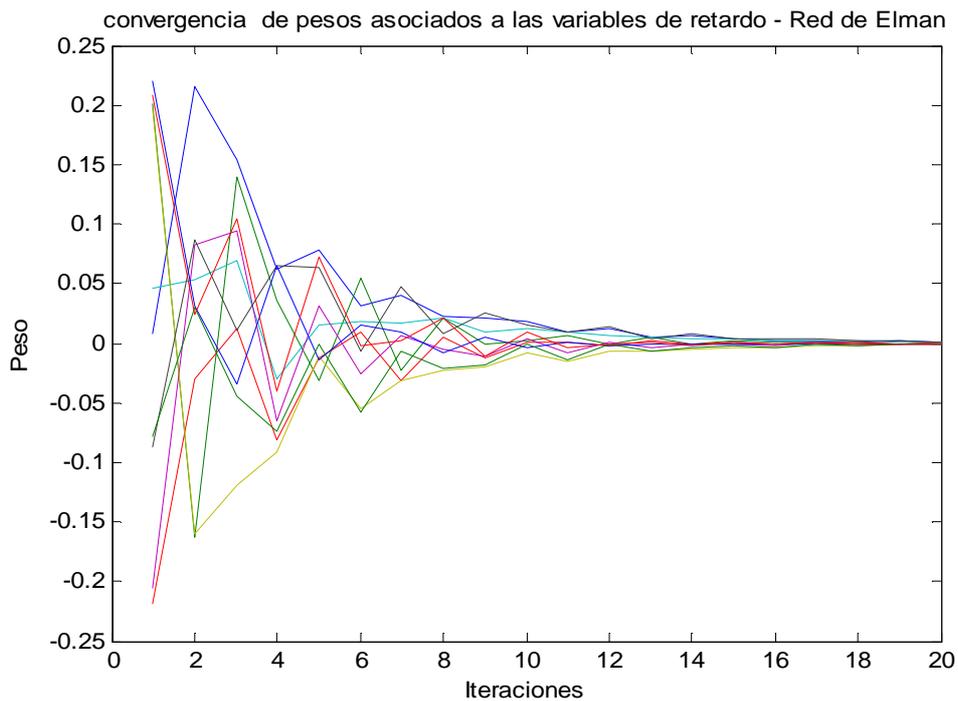
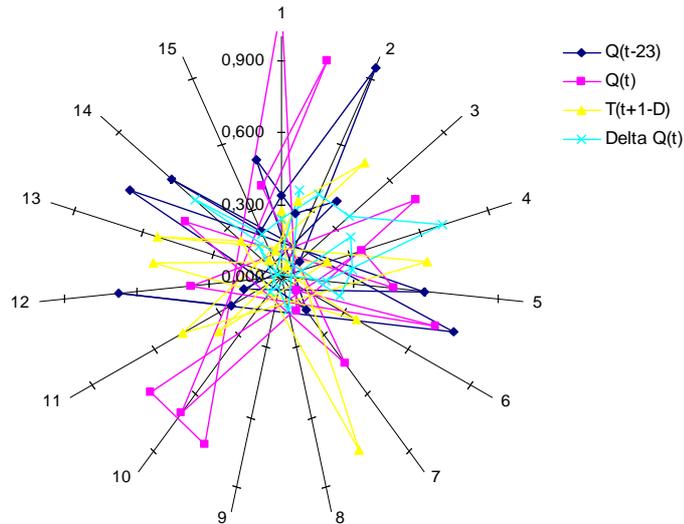


Figura 2.19: Pesos asociados a las variables de entrada a la red. Crecidas de Deshielo.



2.6 Análisis espectral de crecidas

Con el fin de establecer la variabilidad y amplitud de las frecuencias de la señal descrita por los datos de cada crecida y a modo de conocer si existe una relación directa entre la capacidad de generalización de una crecida y su variabilidad frecuencial, se hace necesario contar con alguna herramienta matemática que lleve la serie de caudales de un dominio de tiempo a un dominio de frecuencias. Para el cumplimiento de tal objetivo, se utiliza la transformada de Fourier, descrita a continuación.

2.6.1 Transformada de Fourier

Una transformada de Fourier es una operación matemática que transforma una señal de dominio de tiempo a dominio de frecuencia y viceversa.

Las condiciones para poder obtener la transformada de Fourier son (Condiciones de Dirichlet):

- Que la señal sea absolutamente integrable, es decir:

$$\int_{-\infty}^{\infty} |x(t)|^2 dt < \infty \quad (2.66)$$

- Que tenga un grado de oscilación finito.
- Que tenga un número máximo de discontinuidades.

La transformada de Fourier es una particularización de la transformada de Laplace con $S = j\omega$ (siendo $\omega = 2\pi f$), y en un dominio discreto se define como:

$$X(k) = \sum_{n=1}^N x(n) * e^{-j2\pi n(k-1)\left(\frac{n-1}{N}\right)} \quad 1 \leq k \leq N \quad (2.67)$$

Y su antitransformada se define como:

$$x(n) = \sum_{k=1}^N X(k) * e^{j2\pi n(k-1)\left(\frac{n-1}{N}\right)} \quad 1 \leq n \leq N \quad (2.68)$$

2.7 Índices de Evaluación de desempeño

2.7.1 Índices de desempeño general

Para comparar el desempeño de los modelos de pronóstico, se utilizan las siguientes expresiones, extensivas a cualquier método que se use.

A partir de ahora se entiende por:

d_k = Salida deseada.

y_k = Salida ajustada.

- **Error cuadrático Medio**

$$ECM = \frac{1}{n} * \sum_k (d_k - y_k)^2 \quad (2.69)$$

- **Error cuadrático Medio normalizado**

$$ECMN = \frac{\sum_k (d_k - y_k)^2}{\sum_k (d_k - \bar{y})^2} \quad (2.70)$$

Donde:

\bar{y} Denota el promedio de los datos. Si el valor del coeficiente ECMN se acerca más a cero, el modelo se ajusta en mayor medida a los datos originales.

Uno de los indicadores más importantes para evaluar el desempeño de un determinado modelo es el coeficiente de determinación r^2 el cual es función del ECMN y se define como:

$$r^2 = 1 - ECMN \quad (2.71)$$

Para un predictor perfecto este coeficiente debe ser 1, mientras que para un predictor trivial que siempre predice el valor medio $r^2 = 0$ (Pérez, 2002).

2.7.2 Índices para la Evaluación de la Calidad de un Modelo de Pronóstico

Fundamentalmente interesa evaluar el comportamiento del modelo de pronóstico en las cercanías de los valores máximos. Para evaluar la calidad del ajuste en la cercanía del Caudal máximo, se toman en cuenta las siguientes medidas de error:

- Promedio de la diferencia entre el caudal máximo simulado y el máximo observado.

$$EMEM = \frac{1}{N_c} \sum_i \frac{1}{N} \sum_j \frac{Max(\hat{Q}_j^{t_{max}}) - Q_{max\ i}}{Q_{max\ i}} \quad (2.72)$$

- Promedio de la diferencias entre caudal simulado y observado en el instante en que ocurre el máximo.

$$EMMOBS = \frac{1}{N_c} \sum_i \frac{1}{N} \sum_j \frac{\hat{Q}_j^{t_{max}} - Q_{max\ i}}{Q_{max\ i}} \quad (2.73)$$

- Tiempo promedio de la diferencia entre la ocurrencia del caudal máximo simulado y el máximo observado.

$$TMEM = \frac{1}{N_c} \sum_i \frac{1}{N} \sum_j t_j^{max} - t_{max\ i} \quad (2.74)$$

Donde:

N_c = Número de crecidas consideradas.

N = Número de veces en que se efectúan pronósticos para el instante en que ocurre el caudal máximo.

$\hat{Q}_j^{t_{max}}$ = Caudal simulado en el instante en que ocurre el máximo observado.

$Q_{\max i}$ =Caudal máximo observado.

\hat{t}_j^{\max} =Tiempo en que ocurre el caudal máximo simulado.

$t_{\max i}$ =Tiempo en que ocurre el caudal máximo observado.

-Valores positivos/negativos de los índices EMMOBS o EMEM indican que el modelo sobreestima/subestima el caudal máximo.

-Valores positivos/negativos del índice TMEM indica que el caudal máximo simulado ocurre después/antes que el máximo observado (desfase hacia delante/desfase hacia atrás).

2.7.3 Índices de Comparación entre Modelos de Pronóstico

Para comparar dos modelos de pronóstico planteados, se considera un índice que involucre todo el conjunto de datos utilizados.

- **Error cuadrático medio ponderado**

$$ECM = \frac{n_{ENT} * ECM_{ENT} + n_{val} * ECM_{val} + n_{PRU} * ECM_{PRU}}{n_{total}} \quad (2.75)$$

Donde:

$n_{ENT}, n_{VAL}, n_{PRU}$ = Número de datos utilizados en entrenamiento, validación y prueba respectivamente.

ECM = Errores cuadráticos medios cometidos en cada uno de estos procesos.

Capítulo 3

3 Caracterización de la Cuenca

3.1 Generalidades

La cuenca del Río Mapocho perteneciente a la hoya hidrográfica del río Maipo drena las aguas cordilleranas del sector del cerro El Plomo y tiene como afluentes principales al río Molina, al río San Francisco y a los esteros Arrayán y Lampa, cruzando toda la ciudad de Santiago.

Presenta en su parte alta un régimen de carácter nival, siendo los deshielos en los inicios de la primavera la fuente de su caudal más significativo. El caudal medio anual es de $6,1 \text{ m}^3/\text{s}$ (Estación Los Almendros). En su parte baja su caudal se ve aumentado por los aportes de los afluentes intermedios y el régimen es pluvionival (Mapocho en Rinconada de Maipú). Con respecto a la variación estacional, el caudal medio mensual presenta un máximo de $13,6 \text{ m}^3/\text{s}$ en noviembre y un mínimo en abril del orden de $2,3 \text{ m}^3/\text{s}$ (Mapocho en Los Almendros).

3.1.1 Subcuenca Alta del Mapocho

Corresponde al área drenada por la parte alta del río Mapocho, desde el nacimiento de sus afluentes cordilleranos: río Molina, río San Francisco, estero Yerba Loca, hasta la junta del río Mapocho con el estero Arrayán, incluyendo este último.

En esta subcuenca se observa un régimen nivo – pluvial, ya que aunque la influencia nival es la más importante es posible advertir cierta influencia pluvial, salvo en el caso del estero Yerba Loca, el que es puramente nival. Los mayores caudales mensuales se observan entre noviembre y enero, mientras que los menores ocurren entre marzo y mayo.

3.1.2 Subcuenca Baja del Mapocho

Corresponde al área drenada por el río Mapocho, desde la junta del estero Arrayán hasta la confluencia del Mapocho con el río Maipo, incluyendo al estero Colina.

En esta subcuenca se observa una considerable influencia pluvial y tiene una gran intervención antrópica. Los mayores caudales se observan en julio y octubre, mientras que el período de menores caudales ocurre entre febrero y abril.

3.2 Control de estaciones

3.2.1 Estación Fluviográfica Estero Arrayán en la Montosa

Estación de tipo fluviográfica y limnigráfica controlada por la DGA. Controla la mayor parte del Estero Arrayán, afluente cordillerano del río Mapocho y presenta un tiempo de respuesta bastante rápido de aproximadamente 3 horas.

Sus principales antecedentes se presentan en la tabla 3.1. Las características de la cuenca definida por los aportes a la estación se observan en la tabla 3.2 y gráficamente en la figura 3.1, donde también se puede apreciar la cuenca de Mapocho en Los Almendros (delimitación gris).

Tiene un claro comportamiento nivopluvial, registrando caudales máximos en los meses de diciembre para años húmedos y en noviembre para años secos, como se aprecia en la figura 3.2.

Tabla 3.1: Estación Arrayán en La Montosa.

Código BNA	05722001-5
Latitud S	33° 21'
Longitud O	70° 29'
Altitud [msnm]	880

Tabla 3.2: Características Cuenca Estero Arrayán en La Montosa.

Pendiente Media [m/m]	0,415
Máxima distancia de flujo [Km.]	34
Largo Canal Principal [Km.]	28
Área [km ²]	219
Tiempo de Concentración [Hrs.]	3

Figura 3.1: Cuenca del Estero Arrayán en La Montosa.

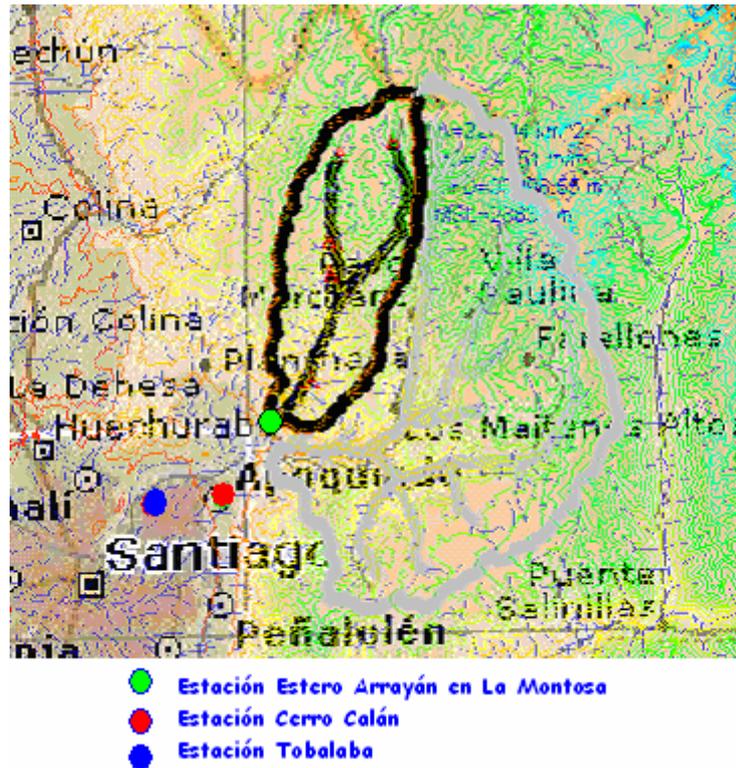
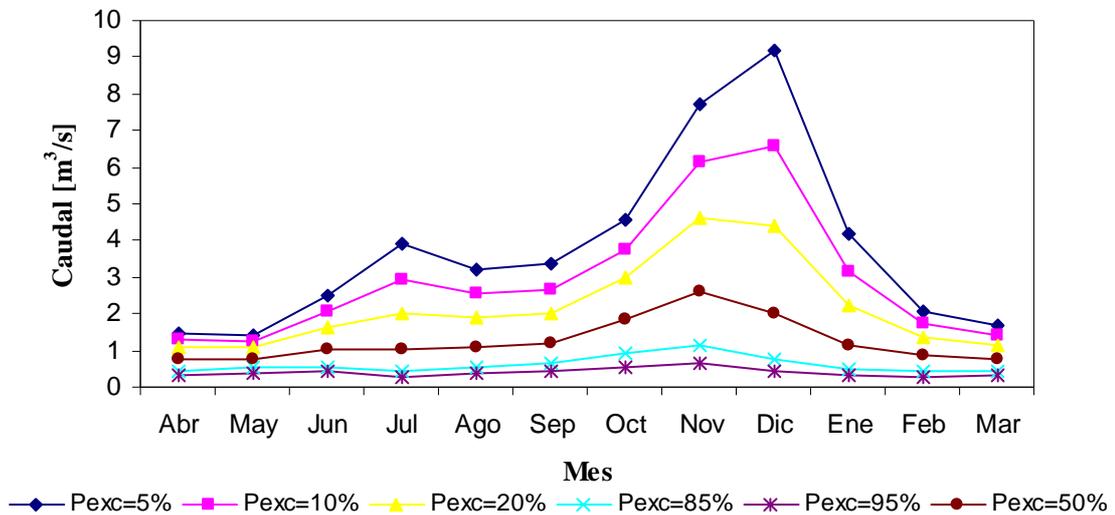


Figura 3.2: Curva de Variación Estacional –Arrayán en La Montosa.



3.2.2 Estación Fluviográfica Mapocho en Los Almendros

Estación de tipo fluviográfica y limnigráfica, controlada por la DGA. Ubicada sobre el río Mapocho antes de la confluencia con el estero Arrayán. Mide los aportes provenientes de la parte alta de la cuenca, cuyos principales afluentes son el río San Francisco y el río Molina, presenta tiempos de respuesta de 2 a 3 horas.

Sus principales antecedentes se presentan en la tabla 3.3. Las características de la cuenca definida por los aportes a la estación se observan en la tabla 3.4 y gráficamente en la figura 3.4, en la cual se puede observar también la cuenca del Estero Arrayán (delimitación gris).

Tiene un régimen caracterizado por el aporte menor, pero no despreciable de las precipitaciones invernales en los meses de junio a agosto, para presentar una marcada crecida en el mes de noviembre con los aportes provenientes de los deshielos, como se aprecia en la figura 3.3.

Tabla 3.3: Estación Mapocho en Los Almendros.

Código BNA	05733003-3
Latitud S	33° 22'
Longitud O	70° 28'
Altitud[m.s.n.m.]	1024

Tabla 3.4: Características Cuenca Mapocho en Los Almendros.

Pendiente Media [m/m]	0,446
Máxima distancia de flujo [Km.]	43
Largo Canal Principal [Km.]	38
Área [km ² .]	620
Tiempo de Concentración [Hrs.]	3

Figura 3.3: Curva de Variación Estacional - Mapocho en Los Almendros.

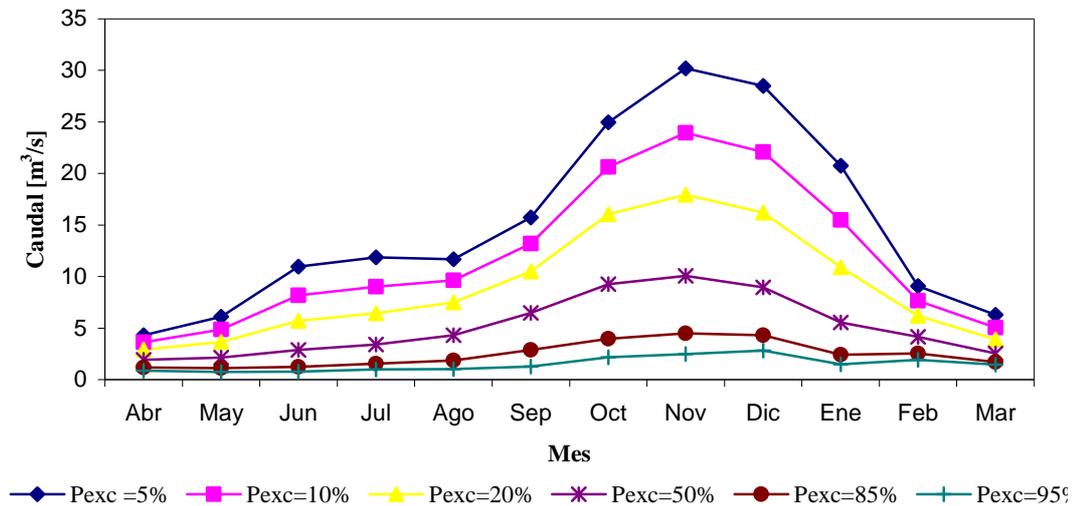
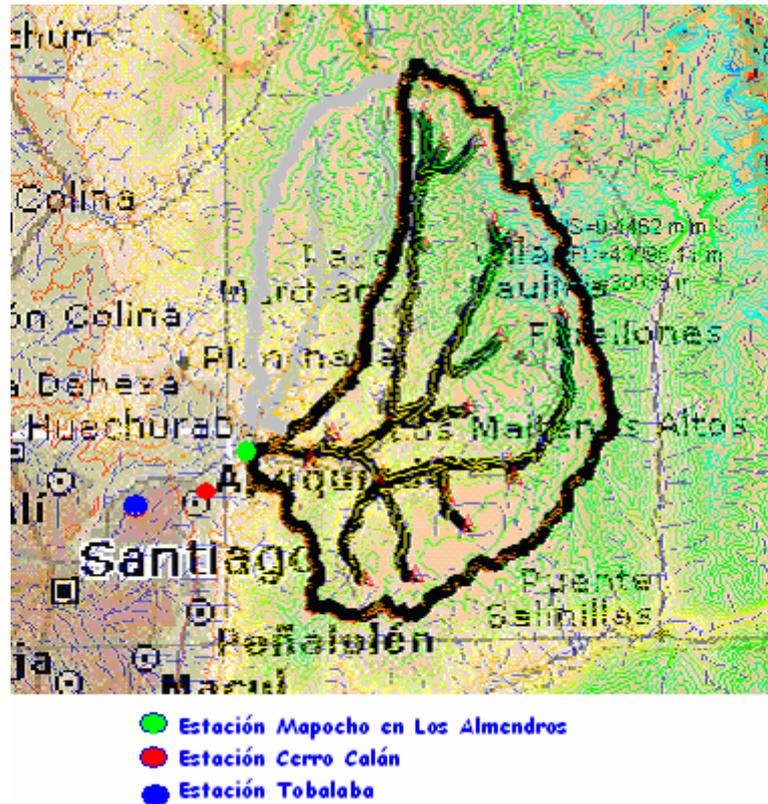


Figura 3.4: Cuenca del Mapocho en Los Almendros.



3.2.3 Estación Pluviográfica Cerro Calán

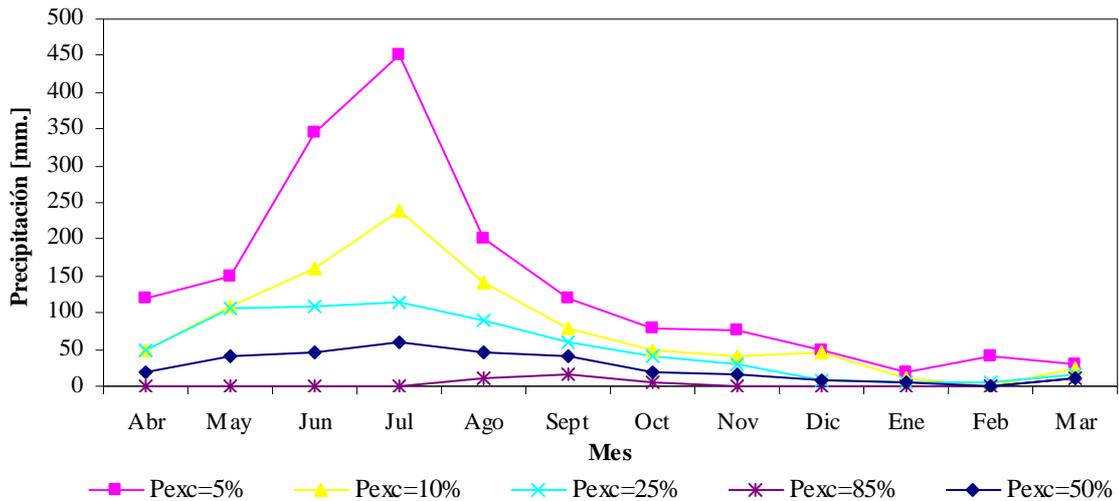
Estación de tipo Meteorológico, controlada por la DGA, sus principales antecedentes se presentan en la tabla 3.5. Son de interés los registros pluviográficos y los de temperaturas en horas sinópticas y las extremas diarias.

Se aprecia en la figura 3.5 que la mayor parte de las precipitaciones se produce en el período comprendido entre los meses de abril a septiembre.

Tabla 3.5: Estación Cerro Calán.

Código BNA	05730050-7
Latitud S	33° 24'
Longitud O	70° 32'
Altitud[m.s.n.m.]	900

Figura 3.5: Curva de Variación Estacional –Cerro Calán.



3.2.4 Estación Meteorológica Tobalaba

La estación Tobalaba es de tipo meteorológico, depende de la Dirección Meteorológica de Chile (DMC) y posee registros de temperaturas horarias entre los años 1961 y la actualidad. Sus principales antecedentes se presentan en la tabla 3.4.

Tabla 3.6: Estación Tobalaba.

Código DMC	05730019
Latitud S	33° 27'
Longitud O	70° 33'
Altitud[m.s.n.m.]	650

3.2.5 Curvas Hipsométricas

Para analizar los volúmenes de entrada a la cuenca producidos por efecto de las precipitaciones (líquidas) es necesario encontrar el área de influencia asociada a dicha precipitación en función del tiempo.

La estimación del área pluvial está fuertemente relacionada con la estimación de la línea de nieve (LN), la cual separa a los sectores con precipitación líquida y sólida. Garreaud, (1992) y Escobar y Vidal, (1992) han demostrado que la temperatura del aire en superficie es un buen índice para definir la forma de precipitación.

La ubicación de la línea de nieve es posible de estimar mediante un gradiente térmico fijo, registros termográficos en una estación cercana a la cuenca y la definición de una temperatura umbral.

Mediante registros de temperatura horaria en estaciones meteorológicas, la definición de un gradiente adiabático saturado de 6,5°C/Km. (Garreaud, 1992) y de una temperatura umbral de 1,5°C (Seguel y Stowhas, 1985) típico para la zona central de Chile, se tiene:

$$H_{LN} = H_{REF} + \frac{T_{REF} - T_U}{|\alpha|} \quad (2.76)$$

Donde:

H_{LN} = Altura de nieve (m.s.n.m).

H_{REF} = Altura Estación (m.s.n.m).

T_{REF} = Temperatura medida en estación base (°C)

T_U = Temperatura Umbral (°C).

α = Gradiente de Temperatura (°C/m).

Una vez que se conoce la posición de dicha LN se separan las áreas aportantes instantáneas considerando como punto de salida de la sub-cuenca la cota correspondiente a la estación limnigráfica y encontrando un modelo que se ajuste al comportamiento de los datos altura v/s área, de tal manera que queda determinada el área asociada a la altura de la LN y por lo tanto el volumen de precipitación aportante horario ($P_t * A_t$).

A continuación se presentan las ecuaciones de los modelos polinómicos ajustados a los datos que definen las curvas hipsométricas (figuras 3.6 y 3.7).

Mapocho en Los Almendros:

$$\text{ÁREA} = -3,54 * 10^{-18} * LN^6 + 5,79 * 10^{-14} * LN^5 - 3,73 * 10^{-10} * LN^4 + 1,17 * 10^{-6} * LN^3 - 1,78 * 10^{-3} * LN^2 + 1,27 * LN - 3,46^2$$

Arrayán en La Montosa:

$$\text{ÁREA} = 2,39 * 10^{-18} * LN^6 - 3,89 * 10^{-14} * LN^5 + 2,49 * 10^{-10} * LN^4 - 8,18 * 10^{-7} * LN^3 + 1,46 * 10^{-3} * LN^2 - 1,25 * LN + 4,04 * 10^2$$

Figura 3.6: Curva de Hipsométrica –Mapocho en Los Almendros.

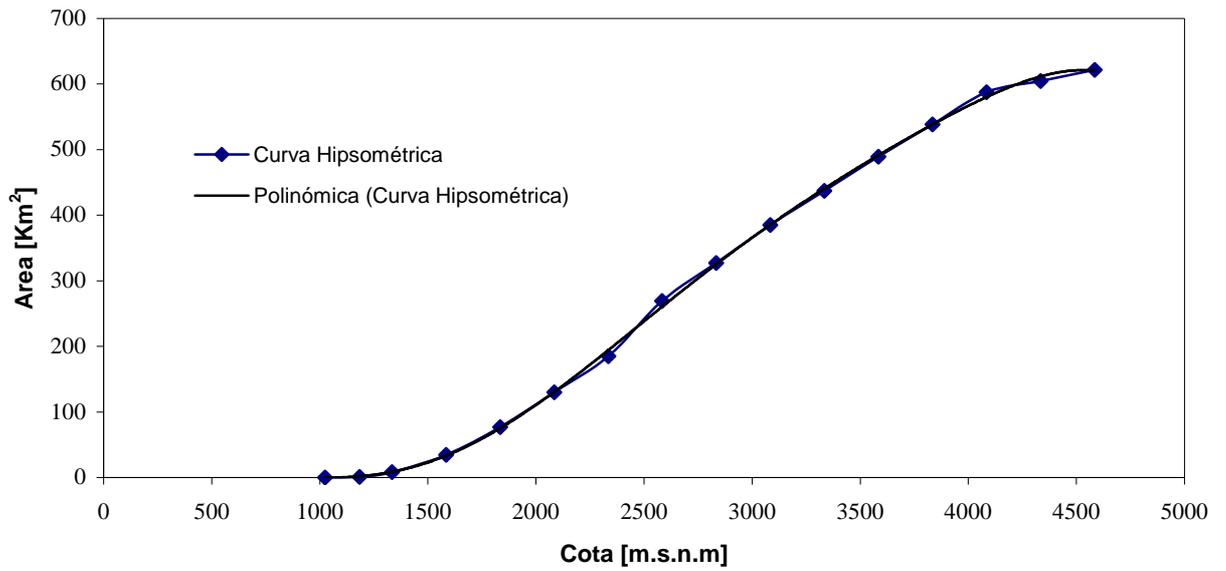
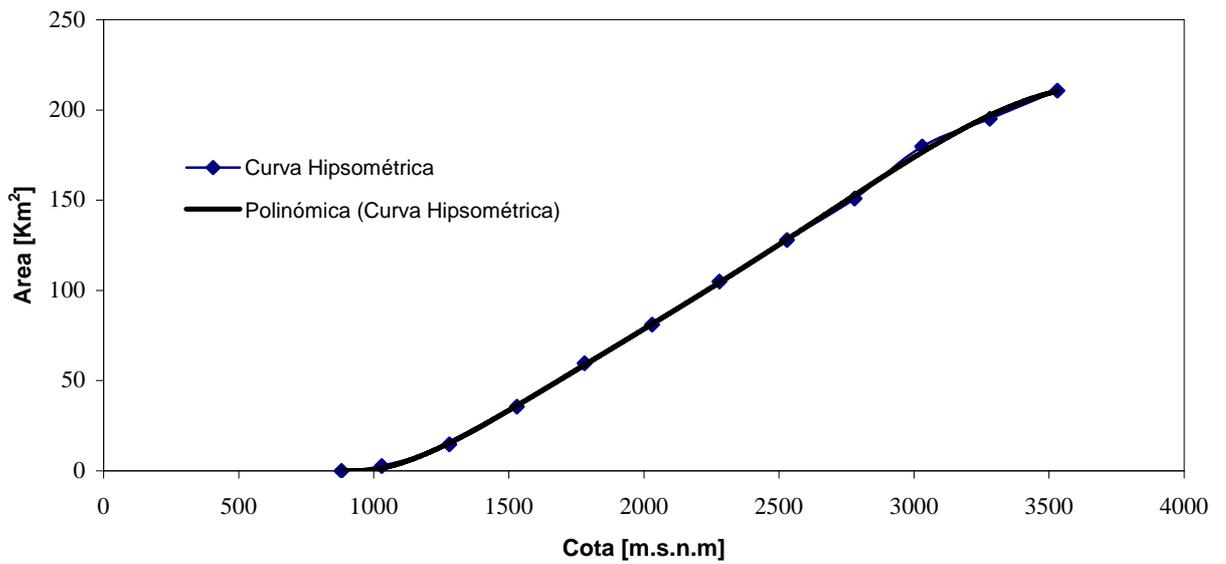


Figura 3.7: Curva de Hipsométrica –Arrayán en La Montosa.



Capítulo 4

4 Pronóstico de Crecidas

4.1 Introducción

El objetivo de este capítulo es realizar pronósticos en tiempo real de eventos de origen pluvial y de deshielo.

Las crecidas pluviales se producen por efecto de la caída de lluvia en una determinada zona de la cuenca, la cantidad de lluvia caída experimenta variaciones debido a factores como la evaporación producto de fenómenos meteorológicos y a características físicas de la cuenca como son, la pendiente, la vegetación, el tipo de suelo, drenaje, condiciones iniciales de saturación entre otros, sin presentar necesariamente una distribución uniforme.

Luego, las variables de interés desde el punto de vista hidrológico en crecidas pluviales son los caudales medidos en el punto de salida de la cuenca, las precipitaciones y su área de influencia y además podría considerarse una tercera variable, la radiación, que si bien afecta en alguna medida, se considera de mayor relevancia en épocas de deshielo.

Las crecidas de deshielo puede ser explicadas como aquel evento ocurrido en la época de deshielo (septiembre – abril para la zona central de Chile) que presente un comportamiento cíclico diario y que sobrepase un cierto caudal umbral predefinido, pueden eventualmente formar parte de una crecida, algunos caudales menores al caudal umbral, lo cual se define para otorgar continuidad temporal a los eventos seleccionados. La duración de una crecida puede ser incluso de meses, teniéndose por lo tanto una gran cantidad de registros.

Se producen por variados y diferentes procesos de transferencias de calor (radiación solar, radiación neta de onda larga, transferencias de calor sensible, calor latente, de condensación, etc.), cuya determinación depende además de la condición física y térmica del manto de nieve.

Para poder simular escorrentía nival con cierta precisión, se debe considerar el mecanismo de acumulación de nieve de tal manera que una vez llegada la época de deshielo, las condiciones del manto de nieve sean representativas de la realidad (Martínez, 1993).

Una vez que el manto de nieve está apto para su derretimiento, la variable más representativa del fenómeno viene a ser la radiación, sin embargo dado que no existen registros continuos (a nivel horario) de dicha variable, la temperatura se considera un buen índice de la generación de deshielos.

4.2 Planteamiento General

Se realizan pronósticos en tiempo real de caudales a nivel horarios en las sub-cuencas definidas por la estación Arrayán en La Montosa y Mapocho en Los Almendros.

Los datos de la estación metereológica Tobalaba permiten estimar la posición de la LN horaria en base a la temperatura y los registros de la estación pluviográfica Cerro Calán permiten caracterizar las precipitaciones horarias. Ambas son consideradas representativas de cada cuenca a pesar de encontrarse fuera del área controlada por las estaciones fluviográficas correspondientes, debido a la inexistencia de otras estaciones de medición en la zona.

Cabe destacar, que los registros en las estaciones de interés han sido obtenidos de: González, 2003, con el fin de realizar una comparación de los pronósticos obtenidos. Dichos datos constituyen registros de la DGA y la DMC.

Dos de los supuestos más importantes a considerar para el pronóstico de crecidas pluviales en cuanto a las variables de entrada, son el suponer la precipitación uniformemente distribuida en el área aportante y que no se hace uso de la variable temperatura de manera directa, sino que a través de la estimación de la línea de nieve horaria.

En el caso de pronósticos de crecidas de deshielo, en cuanto a las variables de entrada es importante tener en cuenta que como se explicó con anterioridad el fenómeno es cíclico y por lo tanto variables relevantes para modelar el fenómeno son los caudales que se dan con 24 horas de anticipación. En el mismo ámbito, horas después de ocurrida la temperatura máxima se produce el caudal máximo, dicho número de horas define otra variable relevante para modelar el fenómeno. Además, se considera que la precipitación no es una variable relevante en esta época frente a la variable temperatura, lo cual se basa en análisis anteriores (González, 2003).

Considerando que las redes se encargan de pronosticar el comportamiento del fenómeno precipitación-escorrentía y derretimiento o Temperatura-escorrentía en la cuenca, es importante destacar que lo que la red aprende y modela no es el fenómeno en sí, sino que crea una representación del comportamiento de la cuenca, es decir intenta modelar la respuesta de la “cuenca” frente a diversas entradas (precipitaciones o Temperaturas).

Se contemplan diferentes modelos y arquitecturas neuronales de una o más capas ocultas, donde el número de neuronas en dichas capas debe ser el suficiente como para dar la representatividad requerida a la variable de salida y no redundante como para provocar ruidos o perturbaciones en el modelo de pronóstico.

Tomando en cuenta que se tienen registros de series no continuas de tiempo para realizar el entrenamiento de la red, y dado que el tipo de redes que se utiliza, “recurrentes”, son del tipo que almacenan información entre la presentación de un patrón y el siguiente, si se realiza un entrenamiento continuo en el sentido de “crecidas pegadas”, cada vez que se ingrese una crecida existirá un salto en la serie de tiempo (paso de una crecida a otra), con lo que la red estaría almacenando información de la crecida pasada no relevante para la actual. Según el análisis realizado en el punto 2.5.2, el almacenamiento de información corresponde aprox. a 5 datos precedentes en el caso de crecidas pluviales y a aprox. 15 en el caso de crecidas de deshielo, lo que autoriza su aplicación, dado que el límite de tiempo no es extenso, pero igualmente puede provocar errores en el aprendizaje de la red del fenómeno modelado. Luego, se realiza dos tipos de entrenamiento, cada uno de los cuales queda descrito como sigue.

- El entrenamiento de la red considerando los datos de cada crecida en particular, con la mejor arquitectura encontrada mediante pruebas y para diferentes configuraciones de entrada, de tal manera de minimizar el error cuadrático medio y con el objetivo de obtener un caudal máximo representativo.

Cabe señalar, según la teoría y metodología expuesta en el capítulo 1, que se considera el estudio de la arquitectura óptima para 1 configuración, en el sentido de encontrar el número óptimo de capas ocultas, así como también, la cantidad de épocas de entrenamiento necesarias. Esta arquitectura se hace extensiva a todos los modelos del mismo tipo identificados por la crecida de entrenamiento, considerando como única variación las variables de entrada.

Posteriormente, se realiza la validación y prueba con las crecidas restantes, permitiendo tener un amplio espectro de decisión para pronosticar en base al mínimo error cuadrático medio ponderado que presenta cada modelo formado a partir del entrenamiento con los datos de cada crecida.

Paralelamente, se realiza un análisis de frecuencia para cada crecida, de tal manera de observar el espectro de frecuencias de la crecida y establecer una relación entre la amplitud de las componentes de frecuencia y la capacidad de generalización de la

crecida. Capacidad medida en el sentido de que una vez entrenada la red con el conjunto de datos que la describen sea capaz de generalizar y presentar buenas validaciones, es decir sea capaz de representar adecuadamente comportamientos suaves, bruscos y diferentes a los presentados durante el entrenamiento.

- El entrenamiento de la red considerando un determinado número de crecidas y entregándolas a la red como datos continuos. Se realiza una división de todo el conjunto de crecidas, en un set de entrenamiento, validación y prueba.

Para ambos tipos de entrenamiento, se programan **redes recurrentes de Elman** y **redes prealimentadas**, ambas son comparadas entre sí y posteriormente con **modelos paramétricos**, para finalizar con un **modelo neuronal recurrente de Hopfield**. Cabe destacar que el modelo de Hopfield sólo es aplicado a crecidas pluviales, debido a consideraciones que se explican conforme se desarrollen los modelos.

Los índices de comparación de errores usados son: el error cuadrático medio ponderado de los conjuntos de entrenamiento, validación y prueba, el pronóstico de caudales máximos y los errores de pronóstico obtenidos de acuerdo a una validación por crecidas.

El entrenamiento de la red recurrente de Elman se realiza en base al algoritmo de retropropagación de errores, y el entrenamiento de la red prealimentada en base a dicho algoritmo o Levenberg-Marquardt, según existan o no mejoras.

Antes de continuar, es importante destacar que para el pronóstico mediante modelos neuronales recurrentes no se hace necesaria la incorporación de variables retardadas puesto que la red presenta precisamente ese beneficio, producir un retardo de la información y asignarle cierta relevancia. En el caso de las redes prealimentadas esto si se hace necesario, ya que no presentan retardos de la información y por ende tampoco almacenamiento de ella, de manera que se debe entregar a la red un vector de entrada con más parámetros para poder representar el modelo de manera similar y hacerlos comparables.

Primeramente se realizan pronósticos a un paso de adelanto utilizando los diferentes tipos de modelos descritos, se incluyen análisis detallados, para posteriormente realizar pronósticos con diferentes horas de antelación considerando cada uno de los modelos planteados.

4.3 Aplicación a Crecidas Pluviales

4.3.1 Cuenca del Estero Arrayán en La Montosa

Se cuenta con un total de 21 crecidas, 1587 datos cuyas características se presentan en la tabla 4.1. Dado el rápido tiempo de respuesta de la cuenca (aprox. 3 hrs.) se considera adecuado utilizar modelos de dos capas ocultas, debido a que son capaces de modelar fenómenos más complejos, por el mayor procesamiento de la información.

Tabla 4.1: Caracterización de Crecidas.

CRECIDA	INICIO	TERMINO	DURACIÓN [Hr.]	MEDIA [m ³ /s]	DESV ESTAND [m ³ /s]	Q Máx [m ³ /s]	QMin [m ³ /s]	Pp [mm.]	T ^o C
1	27-5-86 6:00	29-5-86 4:00	43	1,5	1,0	4,9	0,5	92,1	8,5
2	15-6-86 4:00	20-6-86 12:00	125	13,4	6,6	44,8	2,5	104,3	11,3
3	2-6-87 16:00	5-6-87 9:00	62	6,0	2,8	13,1	1,9	40,9	11,1
4	11-7-87 8:00	18-7-87 4:00	161	23,2	13,9	64,2	8,0	312,2	8,7
5	9-8-87 0:00	16-8-87 9:00	174	10,0	6,0	27,0	3,2	183,2	7,7
6	22-8-89 16:00	27-8-89 8:00	112	5,5	4,8	22,0	1,6	77,3	9,8
7	26-5-91 19:00	30-5-91 10:00	84	5,5	1,8	10,9	2,6	57,2	13,8
8	15-9-91 3:00	19-9-91 13:00	103	4,6	1,4	8,1	2,5	72,8	11,7
9	26-5-92 0:00	29-5-92 4:00	73	3,4	2,0	9,9	1,7	82,8	6,2
10	18-4-93 23:00	20-4-93 0:00	22	3,9	2,4	9,9	1,3	48,9	12,1
11	3-5-93 3:00	4-5-93 6:00	24	9,8	7,1	27,5	1,5	35,2	13,0
12	5-5-93 14:00	7-5-93 20:00	51	9,1	3,3	20,1	4,0	58,1	9,6
13	17-6-97 18:00	26-6-97 15:00	210	4,3	1,9	8,9	1,5	169,1	9,0
14	15-8-97 15:00	20-8-97 12:00	114	3,6	0,8	4,9	2,1	100,0	8,3
15	30-8-97 12:00	1-9-97 11:00	44	5,6	1,1	7,1	2,9	7,1	12,5
16	3-9-97 2:00	5-9-97 13:00	56	4,7	0,5	6,2	4,0	26,2	11,2
17	6-9-97 11:00	8-9-97 23:00	57	4,0	0,3	4,8	3,6	24,9	10,8
18	18-9-97 17:00	19-9-97 20:00	24	3,9	0,3	4,5	3,5	15,5	12,7
19	25-5-02 16:00	26-5-02 9:00	14	8,6	5,7	19,8	2,0	60,3	11,5
20	3-6-02 6:00	3-6-02 22:00	13	8,8	3,1	13,6	5,2	126,5	7,1
21	4-6-02 1:00	5-6-02 1:00	21	7,9	3,1	15,6	4,4	66,5	6,1

Se realiza un análisis espectral de las crecidas, aplicando la transformada de Fourier a la serie de caudales que compone cada crecida con el objetivo de determinar cuales de ellas presentan mayores amplitudes de frecuencia, o lo que es lo mismo que presentan mayor variabilidad temporal, lo cual queda representado en la figura 4.1 y 4.2.

Se observa que las crecidas que mayor amplitud y variabilidad presentan son las crecidas número 4, 2, 5, 12, 6 y 7, y que las que presentan menores amplitudes de frecuencia son las crecidas número 16, 17, 18, 19, 10 y 20.

Cabe destacar que el comportamiento en torno a la frecuencia cero no tiene relevancia, puesto que es simplemente el caudal total acumulado durante la crecida medido en el punto de salida de la cuenca. Es de relevancia notar lo que pasa en los alrededores de la frecuencia cero, como muestra la figura 4.2, donde se puede apreciar claramente que la crecida 4 tiene amplitudes destacadamente mayores que las demás.

Otro punto importante de mencionar es que si se trata de analizar la variabilidad de la crecida, podría pensarse en observar el comportamiento de la serie de tiempo de la crecida, considerar aquella que sea de mayor duración o cuantificar su energía mediante fórmulas conocidas que consideran simplemente el cuadrado de la variable y el número de datos. Ésto tomaría una menor cantidad de tiempo, pero lo que se estaría haciendo en esos casos no es encontrar la variabilidad o dicho de otro modo, su posible relación con la capacidad de generalización de la crecida en el caso de ser usada como set de datos de entrenamiento del modelo de pronóstico, sino que se estaría cuantificando su valor energético, lo cual no necesariamente es representativo de la variabilidad y amplitud frecuencial.

Figura 4.1: Transformada de Fourier de la serie de caudales –tiempo.

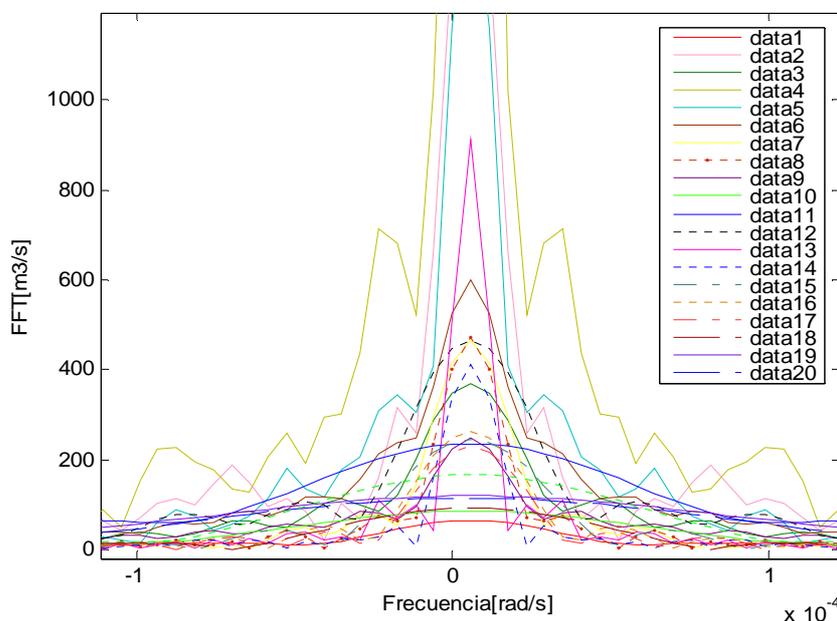
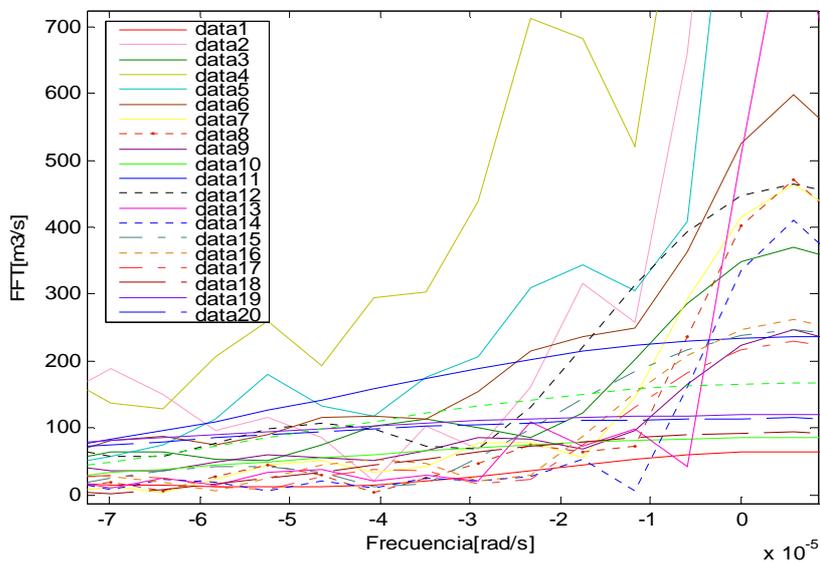


Figura 4.2: Transformada de Fourier de la serie de caudales –tiempo.



4.3.1.1 Pronósticos mediante Redes Neuronales de pronóstico - Entrenamiento Unitario

Primeramente y en este apartado se realizan pronósticos a un paso de adelanto.

De acuerdo al planteamiento general, cada crecida es usada como set de datos de entrenamiento, formando un modelo de pronóstico que es validado con las crecidas restantes. Se crean por lo tanto, 21 modelos independientes cuyos errores de entrenamiento y validación se aprecian en una matriz de errores, en que el error cuadrático medio de validación de la crecida de entrenamiento es el mismo error que se obtiene de su entrenamiento.

4.3.1.1.1 Pronósticos mediante Redes Neuronales Recurrentes de Elman.

De acuerdo al planteamiento general, se consideran como parámetros de entrada las variables Q_t , $(P^*A)_t$ y \dot{Q}_t . Su uso se justifica a continuación:

- Pronóstico de caudales => Comportamiento de la variable Q_t .
- $(P^*A)_t$ representa el volumen de lluvia en el punto de salida de la cuenca.
- \dot{Q}_t (Gradiente de caudales), contribuye a dar representatividad al efecto producido por las curvas de ascenso, máximo y descenso de caudal en el sentido de derivada positiva (ascenso), nula (peak), negativa (descenso).

Se realizan dos configuraciones de entrada:

- $(P^*A)_t$ y Q_t (configuración 0).
- $(P^*A)_t$, Q_t y \dot{Q}_t (configuración 1).

Dado que el tiempo de respuesta de la cuenca ante eventos de precipitación es “**rápido**”, de aprox. 3hrs. (capítulo 3), se consideran modelos de dos capas ocultas.

Se prueban distintos números de neuronas en cada una de las capas ocultas, de tal manera de encontrar la arquitectura que entregue menor error cuadrático medio ponderado (entrenamiento, validación y prueba) sin sobrestimar el número de neuronas útiles, escogiéndose el mejor modelo o la (as) crecida (as) óptima (s) de entrenamiento para realizar pronósticos. Se compara la capacidad de generalización obtenida por medio de la RNR con el análisis realizado por medio de la transformada de Fourier.

A. Configuración de entrada 0

Se modelan diferentes arquitecturas neuronales, que involucran 2 neuronas en la capa de entrada definidas por las variables, $(P \cdot A)_t$ y Q_t , el número de neuronas de cada capa oculta se elige mediante pruebas, variando tal cantidad de 4 a 9, elección basada en modelos neuronales realizados con anterioridad (Dogan, 2007; Marra, 2006; González, 2003). Las referencias se hacen extensivas a todas las pruebas realizadas para determinar el número de neuronas en la capa oculta, independientemente de la cuenca analizada y el tipo de red utilizada, y finalmente, una neurona de salida (pronóstico).

Se analiza el número de neuronas y de épocas de entrenamiento tal que el error de validación no aumente y que minimiza el error cuadrático medio ponderado para cada modelo formado a partir del entrenamiento utilizando una crecida en particular. El error ponderado es entendido aquí como el promedio entre el error cometido al entrenar con una determinada crecida y validar con las 20 restantes, por lo tanto es un error que está cuantificando en mayor parte el error de validación que se comete para cada una de las crecidas. Luego, si este valor es pequeño no necesariamente indica un buen entrenamiento, sino que indica una buena capacidad de generalización de la crecida de entrenamiento, que es lo que se busca para realizar pronósticos.

La tabla 4.2 muestra la comparación de errores ponderados en función del número de neuronas en las capas ocultas para cada modelo definido por la crecida utilizada en el entrenamiento. Entendida como la diferencia de los errores ponderados cometidos para diferentes números de neuronas en cada capa oculta, en particular, se presentan las diferencias al tener 4, 7 ó 9 neuronas en cada capa oculta. Luego, por ejemplo un valor negativo de 4-7 implica que para esa crecida el error ponderado es mayor para un modelo que contiene 7 neuronas en cada capa.

Finalmente, se observa que se producen mejoras de aproximadamente un 71 % para un modelo que contiene 7 neuronas en cada capa oculta en relación a uno que contenga 4 ó 9, entre otros. Finalmente, se define 7 neuronas como el número óptimo de neuronas en cada capa oculta.

Tabla 4.2: Comparación de ECM ponderado de acuerdo al n° de neuronas de las capas ocultas para un modelo de Elman tipo 2-x-x-1.

N° de datos	crecida	Comparación de Errores[m ³ /s] ²	
		9-7	4-7
43	1	1,67	0,02
125	2	0,33	0,88
62	3	0,23	0,03
161	4	0,49	0,70
174	5	0,18	0,75
112	6	0,08	1,05
84	7	2,38	0,02
103	8	0,51	0,61
73	9	6,37	2,27
22	10	2,07	33,22
24	11	-0,87	-2,91
51	12	-0,12	0,15
210	13	0,99	1,25
114	14	1,39	0,14
44	15	-0,13	-3,73
56	16	2,81	0,28
57	17	0,35	0,12
24	18	6,91	-1,55
14	19	-39,93	-12,75
13	20	-30,67	35,47
21	21	-23,35	-21,68

En la tabla 4.3 se presenta el porcentaje de mejoras en términos de la medida de error de validación que se produce al variar el número de épocas de entrenamiento desde 500 a 2500 épocas. Porcentaje entendido de tal manera que si es menor que un 50 % para un determinado paso de épocas, existe un porcentaje mayor al 50 % de crecidas que tiene menor o igual error de validación en la época anterior. Por lo tanto, dicho número de épocas de entrenamiento no es el “óptimo” o más satisfactorio, sino que el anterior. Se observa, que en el paso de 1000 a 2000 épocas casi todas las crecidas mejoran su error de validación y que a las 2500 épocas casi no hay crecidas (modelos) que minimicen el error, por lo tanto el número satisfactorio de épocas de entrenamiento es 2000.

Tabla 4.3: Porcentajes de mejora de ECM de validación de un modelo NEWELM 2-7-7-1 según el número de épocas de entrenamiento.

crecida	500 ->1000 épocas	1000->2000 épocas	2000->2500 épocas
	% de crecidas mejoradas	% de crecidas mejoradas	% de crecidas mejoradas
1	81	71	24
2	76	62	19
3	90	29	19
4	81	81	19
5	100	48	0
6	71	76	52
7	90	43	57
8	81	90	10
9	33	90	67
10	43	33	0
11	38	71	19
12	71	86	86
13	24	90	5
14	100	90	24
15	19	81	95
16	90	90	33
17	29	100	24
18	33	81	81
19	10	5	90
20	90	0	100
21	24	0	100

De ahora en adelante, para los análisis que siguen, los análisis presentados en las tablas 4.2 y 4.3 con motivo de ser explicados sencillamente, se presentan mediante curvas en donde se puede apreciar una envolvente mínima de los errores de validación en términos del número de épocas de entrenamiento y número de neuronas en la (s) capa (s) oculta (s). Cabe destacar a modo de situar al lector en el análisis de errores que según lo descrito en el capítulo 2.4, “metodología”, se hablará de errores y error cuadrático medio indistintamente.

En las tablas 4.4, 4.5 y 4.6 se presenta la matriz de errores obtenida para el modelo elegido de 3 neuronas de entrada, 2 capas ocultas de 7 neuronas y 1 neurona de salida, para 2000 épocas de entrenamiento. Cabe señalar que el número indicado bajo “Valid x”, con x el número de la crecida validada equivale a la duración de la crecida x , lo cual se hace extensivo a todas las matrices de error que se presenten.

Se observa que los menores errores (entrenamiento y validación de crecidas), se obtienen para las crecidas 4 ,8 y 14 (resaltadas en amarillo) y los mayores errores para las crecidas 10, 19, 20,21.

Los menores valores de los ECM ponderados de los modelos formados a partir de cada crecida de entrenamiento, son del orden de 1 a 2 $[m^3/s]^2$, lo que representa menos del 20 % del mínimo caudal medio de todas las crecidas.

El valor de los ECM de entrenamiento son pequeños del orden de $10^{-1} [m^3/s]^2$, exceptuando las crecidas 2 y 4 que presentan errores de entrenamiento mayores, pero que ponderadamente son mejores que el resto.

Los más altos errores de validación se obtienen en las crecidas 4,11 y 5, las cuales de acuerdo a Fourier, son las que presentan mayor amplitud de frecuencia y por lo tanto son más difíciles de modelar a partir de entrenamientos realizados con otras crecidas. Por la misma razón, se encuentran entre las crecidas que como datos de entrenamiento generan modelos que presentan un menor error de validación en el resto de las crecidas.

A su vez llama la atención que las crecidas con las cuales son entrenados los modelos que peor pronostican los fenómenos (crecidas 19, 20, 21, resaltadas en verde), sean crecidas que no se encuentran en los últimos puestos en relación a su amplitud de frecuencia de acuerdo al análisis de Fourier.

Tabla 4.4: Errores de un modelo NEWELM 2-7-7-1.

crecida	Nº de datos	Error de Entrenamiento[m ³ /s] ²	Errores de Validación[m ³ /s] ²								
			Entrenamiento								
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9
			43	125	62	161	174	112	84	103	73
1	43	0,014	0,014	5,648	0,433	12,828	2,692	0,579	1,057	0,446	0,156
2	125	3,957	0,092	3,957	0,489	18,308	3,734	0,879	0,938	0,335	0,194
3	62	0,096	0,069	6,457	0,096	12,331	1,202	0,627	0,925	0,256	0,287
4	161	4,270	0,149	6,406	0,347	4,270	0,901	0,634	0,854	0,093	0,286
5	174	0,751	0,191	7,426	0,443	5,335	0,751	0,960	0,876	0,093	0,394
6	112	0,453	0,074	4,876	0,169	8,086	1,766	0,453	0,865	0,176	0,185
7	84	0,493	0,057	5,569	0,503	11,315	2,531	0,769	0,493	0,197	0,322
8	103	0,067	0,134	6,115	0,406	4,239	0,677	0,587	0,827	0,067	0,240
9	73	0,100	0,109	7,843	0,877	29,061	6,316	1,211	1,335	0,540	0,100
10	22	0,075	0,418	19,792	1,764	85,768	17,677	6,842	4,635	0,959	1,208
11	24	1,060	0,060	5,826	0,864	36,211	6,919	0,995	1,027	0,585	0,290
12	51	0,445	0,071	5,987	0,324	7,275	1,608	0,548	1,072	0,115	0,297
13	210	0,033	0,206	7,144	0,637	5,971	0,917	0,790	0,899	0,086	0,352
14	114	0,011	0,088	7,063	0,368	4,283	0,917	0,356	0,876	0,070	0,293
15	44	0,043	0,208	14,568	1,157	42,402	8,720	3,737	1,169	0,412	0,681
16	56	0,011	0,040	4,863	0,254	6,499	1,465	0,323	0,855	0,114	0,173
17	57	0,002	0,073	7,511	0,279	5,376	1,335	0,501	0,818	0,115	0,313
18	24	0,002	0,100	10,407	0,443	15,039	3,217	1,221	0,948	0,199	0,425
19	14	0,232	1,407	106,826	12,944	352,063	71,312	15,777	8,471	5,156	2,592
20	13	0,082	6,797	84,765	43,078	132,173	24,518	159,261	37,757	16,439	20,068
21	21	0,408	0,903	93,501	7,753	163,436	36,787	11,079	7,320	3,122	4,276

Tabla 4.5: Errores de un modelo NEWELM 2-7-7-1.

crecida	Nº de datos	Errores de Validación[m ³ /s] ²											
		Valid 10	Valid 11	Valid 12	Valid 13	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21
		22	24	51	210	114	44	56	57	24	14	13	21
1	43	2,413	1,154	5,204	0,311	0,122	0,280	0,025	0,020	0,014	10,021	3,483	7,124
2	125	2,001	5,438	3,591	0,316	0,132	0,311	0,031	0,017	0,021	13,339	3,687	5,905
3	62	1,555	4,031	3,856	0,301	0,123	0,260	0,037	0,016	0,016	13,293	3,079	6,336
4	161	2,623	9,079	1,385	0,054	0,027	0,096	0,031	0,004	0,015	13,759	3,931	5,553
5	174	2,409	11,765	2,070	0,059	0,038	0,138	0,040	0,006	0,020	17,974	5,224	5,924
6	112	2,041	3,898	2,054	0,152	0,063	0,142	0,025	0,008	0,013	11,280	2,995	5,841
7	84	2,065	3,750	2,241	0,169	0,072	0,211	0,024	0,013	0,016	9,512	4,213	5,435
8	103	2,643	8,748	1,343	0,031	0,022	0,093	0,026	0,003	0,013	12,347	4,343	5,567
9	73	2,715	6,429	5,956	0,543	0,189	0,400	0,035	0,023	0,017	14,700	4,632	8,414
10	22	0,075	19,991	36,580	1,376	0,225	0,737	0,168	0,144	0,080	55,411	15,490	28,085
11	24	1,931	1,060	5,090	0,638	0,253	0,643	0,028	0,029	0,018	9,968	2,345	6,199
12	51	3,169	4,167	0,445	0,103	0,045	0,163	0,016	0,003	0,007	8,535	4,210	5,691
13	210	3,000	13,516	1,538	0,033	0,028	0,123	0,032	0,003	0,020	14,181	5,531	5,357
14	114	3,098	6,166	1,046	0,023	0,011	0,070	0,017	0,003	0,006	11,044	3,700	5,269
15	44	3,045	9,445	7,182	0,467	0,046	0,043	0,047	0,029	0,016	17,744	2,817	7,732
16	56	2,649	2,620	0,849	0,086	0,037	0,109	0,011	0,004	0,006	6,847	2,995	5,074
17	57	2,025	4,603	0,892	0,062	0,027	0,097	0,025	0,002	0,011	9,183	3,488	4,811
18	24	3,292	5,718	2,372	0,189	0,052	0,152	0,019	0,008	0,002	9,878	3,615	6,719
19	14	18,873	80,833	31,200	7,557	2,394	5,379	0,297	0,182	0,146	0,232	22,372	17,330
20	13	17,174	161,769	48,561	31,517	6,281	13,590	2,424	0,333	0,315	44,377	0,082	10,666
21	21	7,616	56,673	13,862	3,959	0,812	2,986	0,379	0,092	0,141	54,864	17,858	0,408

Tabla 4.6: Errores Ponderados de un modelo NEWELM 2-7-7-1.

crecida	Nº de datos	Error Ponderado[m ³ /s] ²
1	43	2,684
2	125	3,260
3	62	2,514
4	161	1,624
5	174	1,945
6	112	1,892
7	84	2,379
8	103	1,547
9	73	5,203
10	22	15,839
11	24	5,629
12	51	1,848
13	210	1,975
14	114	1,574
15	44	7,614
16	56	1,573
17	57	1,713
18	24	3,339
19	14	58,832
20	13	49,540
21	21	32,643

B. Configuración de entrada 1

A continuación se prueba otro modelo neuronal, cuya única variación con respecto al anterior es la incorporación de la variable gradiente de caudal ($Q_t - Q_{t-1}$) como variable de entrada a la red.

Se genera la matriz de errores presentada en las figuras 4.7, 4.8 y 4.9, en la cual se observan considerables disminuciones tanto en los errores de entrenamiento como en los de validación. En los primeros las disminuciones son de aproximadamente 1 orden de magnitud en todas las crecidas y en los segundos, en las crecidas que mejor generalizan, por ejemplo 2 y 4, dichos errores disminuyen también en aproximadamente 1 orden de magnitud, por lo cual para las crecidas que se encuentran en los primeros “puestos” del análisis de Fourier el error ponderado es del orden de $10^{-1} [\text{m}^3/\text{s}]^2$.

Al observar los errores ponderados de la tabla 4.9 se hace aún más notoria la capacidad de generalización de las crecidas que presentan mayor amplitud de frecuencia según el análisis de Fourier, así como también la poca capacidad de generalización dada por las crecidas 16, 17, 18, que en este caso presentan peores errores que las crecidas 19 y 20, lo cual refuerza el análisis de Fourier, ya que los resultados encontrados mediante él arrojan que las crecidas 16, 17 y 18 son las que presentan menor amplitud en las componentes de frecuencia características.

Nuevamente, se observan los mejores modelos de pronóstico resaltados en amarillo y los peores en verde, lo cual se hace extensivo a todas las tablas que presenten matrices de errores como la indicada en este apartado.

Tabla 4.7: Errores de un modelo NEWELM 3-7-7-1.

crecida	Nº de datos	Error de Entrenamiento[m3/s] ²	Errores de Validación[m³/s] ²								
		Entrenamiento	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9
			43	125	62	161	174	112	84	103	73
1	43	0,001	0,001	0,768	0,027	7,071	0,759	1,995	0,096	0,152	0,117
2	125	0,138	0,018	0,138	0,080	3,074	0,263	0,489	0,181	0,103	0,160
3	62	0,013	0,027	1,462	0,013	2,452	1,051	0,663	0,418	0,236	0,113
4	161	0,573	0,075	1,469	0,271	0,573	0,150	0,140	0,325	0,016	0,172
5	174	0,049	0,046	0,750	0,192	0,160	0,049	0,094	0,237	0,014	0,111
6	112	0,012	0,113	1,518	0,580	0,662	0,103	0,012	0,343	0,014	0,293
7	84	0,008	0,025	1,564	0,110	2,445	1,236	1,933	0,008	0,188	0,046
8	103	0,005	0,067	1,565	0,239	0,520	0,142	0,073	0,340	0,005	0,087
9	73	0,013	0,088	2,231	0,564	14,619	1,928	0,869	0,403	0,336	0,013
10	22	0,001	0,386	29,576	1,529	36,773	9,956	13,588	1,210	1,062	0,817
11	24	0,092	0,004	0,561	0,047	5,192	0,534	1,353	0,097	0,116	0,123
12	51	0,057	0,122	1,976	0,589	1,699	0,606	0,757	0,409	0,018	0,287
13	210	0,004	0,234	5,995	1,045	4,646	0,860	0,642	0,730	0,044	0,453
14	114	0,001	0,067	1,323	0,273	0,519	0,136	0,173	0,294	0,004	0,046
15	44	0,001	0,130	4,011	0,362	2,372	1,285	1,532	0,430	0,020	0,209
16	56	0,009	0,429	29,113	1,241	52,588	9,111	10,085	0,975	0,224	1,301
17	57	0,004	0,621	36,860	2,205	59,198	9,932	11,093	1,563	0,359	1,638
18	24	0,001	1,088	73,779	3,466	129,630	22,275	26,233	1,976	0,500	3,288
19	14	0,055	0,085	8,339	0,928	24,469	4,561	2,236	1,088	0,860	0,284
20	13	0,006	0,268	18,664	1,040	29,314	5,216	5,048	0,840	0,268	0,827
21	21	0,056	0,174	16,673	3,144	5,863	5,745	6,379	3,167	3,694	0,897

Tabla 4.8: Errores de un modelo NEWELM 3-7-7-1.

crecida	Nº de datos	Errores de Validación[m ³ /s] ²											
		Valid 10	Valid 11	Valid 12	Valid 13	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21
		22	24	51	210	114	44	56	57	24	14	13	21
1	43	1,283	0,152	6,840	0,498	0,066	0,330	0,311	1,768	1,212	1,088	2,067	0,682
2	125	1,329	0,983	2,568	0,223	0,066	0,196	0,185	0,912	0,645	2,133	0,522	1,472
3	62	1,011	1,650	2,919	0,105	0,137	0,472	0,276	0,644	0,457	2,419	0,538	1,230
4	161	2,049	2,626	0,485	0,063	0,022	0,053	0,073	0,399	0,277	3,316	0,413	2,045
5	174	1,753	2,818	0,544	0,057	0,011	0,035	0,057	0,326	0,225	2,598	0,264	1,558
6	112	2,619	8,482	0,223	0,066	0,014	0,041	0,076	0,539	0,364	5,180	1,145	2,772
7	84	1,025	1,235	1,982	0,619	0,005	0,374	0,285	0,924	1,417	1,066	3,515	0,124
8	103	2,983	5,250	0,171	0,024	0,005	0,014	0,025	0,196	0,118	2,605	0,411	2,008
9	73	1,724	9,021	11,122	0,160	0,072	0,347	0,235	0,555	0,368	2,080	5,296	2,768
10	22	0,001	27,638	20,557	3,513	0,232	1,830	1,005	7,000	4,501	31,703	25,561	2,937
11	24	1,251	0,092	5,129	0,372	0,045	0,233	0,216	1,294	0,977	1,132	1,651	0,767
12	51	2,355	3,789	0,057	0,027	0,017	0,042	0,054	0,305	0,202	6,688	1,223	2,919
13	210	3,077	15,781	0,519	0,004	0,033	0,112	0,088	0,165	0,160	13,617	3,746	4,365
14	114	2,091	4,217	0,452	0,050	0,001	0,009	0,026	0,287	0,190	1,700	0,154	1,883
15	44	2,107	5,331	0,532	0,010	0,003	0,001	0,017	0,221	0,116	6,826	1,096	2,817
16	56	2,848	15,354	3,443	0,517	0,050	0,089	0,009	0,115	0,063	13,965	2,663	5,299
17	57	4,046	27,458	4,922	0,628	0,061	0,148	0,055	0,004	0,017	20,602	5,717	6,916
18	24	5,064	38,710	6,912	1,280	0,061	0,155	0,028	0,014	0,001	24,047	5,080	7,234
19	14	1,214	4,141	6,918	0,917	0,442	0,861	0,733	1,721	1,341	0,055	1,542	0,991
20	13	2,582	10,008	2,136	0,470	0,070	0,140	0,201	0,690	0,446	2,246	0,006	2,193
21	21	1,114	9,428	10,965	1,458	3,255	5,153	3,703	2,838	1,599	5,319	2,006	0,056

Tabla 4.9: Errores Ponderados modelo NEWELM 3-7-7-1.

crecida	Nº de datos	Error Ponderado[m ³ /s] ²
1	43	1,470
2	125	0,661
3	62	0,822
4	161	0,415
5	174	0,271
6	112	0,560
7	84	0,955
8	103	0,424
9	73	2,689
10	22	10,840
11	24	1,090
12	51	0,731
13	210	1,719
14	114	0,384
15	44	1,101
16	56	10,202
17	57	12,110
18	24	25,113
19	14	4,579
20	13	5,936
21	21	4,963

C. Análisis y Comparación de Modelos Recurrentes de Elman

En la tabla 4.10 se presentan los coeficientes de determinación de entrenamiento y validación para cada crecida, cada una de las cuales representa un modelo neuronal. No se ha presentado el coeficiente ponderado (entrenamiento y validación), por ser prácticamente idéntico al de validación, dado el número de datos que posee cada conjunto.

Se realiza la comparación de dichos valores para los dos modelos planteados, de acuerdo a la configuración de entrada, encontrándose mejoras considerables en un modelo que incorpora el gradiente de caudal como variable de entrada, especialmente en las crecidas 2, 4, 5, 6 y 8 que

presentan los errores ponderados más bajos. Es notable mejoría en las crecidas que presentan baja capacidad de generalización (10, 19, 20, 21).

Tabla 4.10: Coeficientes de determinación. Modelo NEWELM 2-7-7-1 y 3-7-7-1.

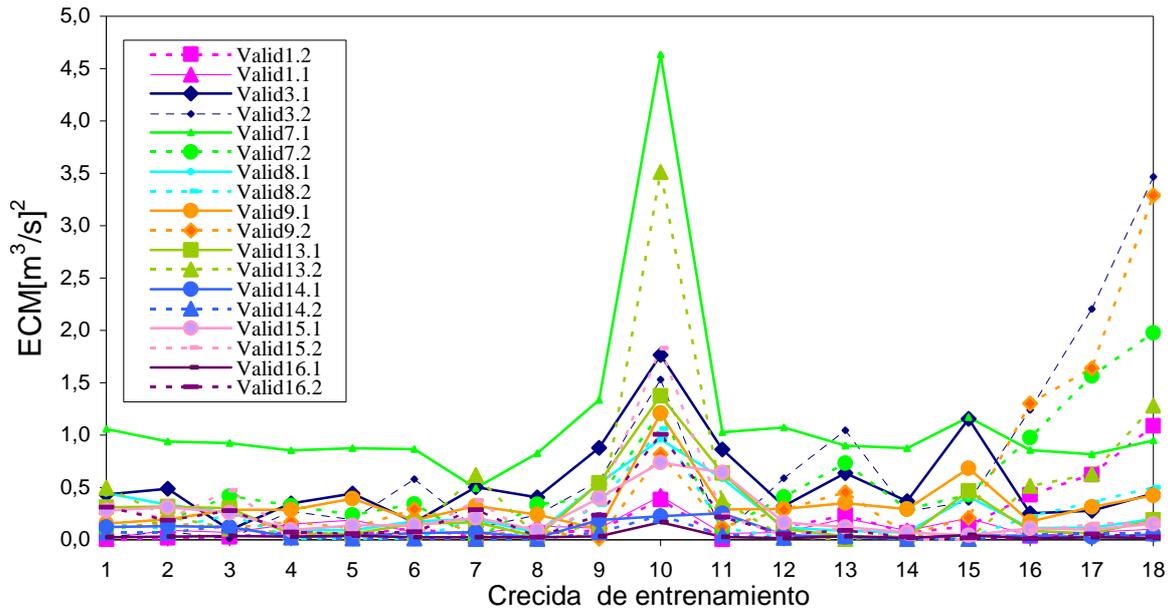
crecida	r ² entrenamiento		r ² Validación	
	Newelm2-7-7-1	Newelm 3-7-7-1	Newelm 2-7-7-1	Newelm 3-7-7-1
1	0,984	0,999	0,757	0,865
2	0,876	0,997	0,714	0,886
3	0,924	0,998	0,789	0,796
4	0,842	0,997	0,882	0,912
5	0,799	0,999	0,858	0,926
6	0,916	0,999	0,854	0,904
7	0,941	0,998	0,845	0,851
8	0,864	0,997	0,892	0,933
9	0,852	0,997	0,571	0,850
10	0,728	1,000	0,565	0,572
11	0,925	0,998	0,541	0,881
12	0,913	0,995	0,859	0,892
13	0,778	0,999	0,862	0,849
14	0,915	0,999	0,910	0,931
15	0,785	0,999	0,743	0,893
16	0,953	0,957	0,881	0,492
17	0,922	0,944	0,894	0,393
18	0,898	0,987	0,844	-0,830
19	-0,250	0,998	-2,588	0,615
20	-3,084	0,999	-7,391	0,800
21	0,299	0,994	-0,328	-0,561

En la figura 4.3 se puede observar el comportamiento de los errores de validación para cada crecida de entrenamiento según el tipo de modelo. Una serie x.1 indica los errores de validación de crecida x, para distintas crecidas de entrenamiento con un modelo que no incorpora la variable gradiente de caudal y de la misma forma un modelo x.2 indica los errores de validación de la crecida x con un modelo que si incorpora la dicha variable.

Se observa claramente, la disminución de errores para el segundo modelo, sobre todo para la crecida de entrenamiento o modelo número 10.

Cabe destacar que la curva en si no tiene ningún significado físico, es sólo esquemática, a modo de mostrar el mejor modelo como la envolvente mínima de errores de validación y sólo importan los valores punto a punto. No se presentan las crecidas 19, 20, 21 debido a que para el primer modelo los errores salen del rango en que se permita apreciar el comportamiento de las demás (ver errores de validación tabla 4.7).

Figura 4.3: Errores de validación según crecida de entrenamiento. Modelos 2-7-7-1 y 3-7-7-1.



En conclusión, un 86 % de los modelos presenta menores errores ponderados cuando involucra el gradiente de caudal como variable de entrada y un 14 % mantiene sus errores o es mejor con un modelo que no involucra dicha variable.

Luego, y de acuerdo a lo anterior, la metodología planteada consiste entonces en encontrar el mínimo error de ponderación para una determinada crecida validada; error tomado como promedio entre su validación y el correspondiente entrenamiento de la red que otorga dicha validación, de tal manera que el número de datos queda equitativamente separado, encontrándose el mejor modelo en términos de pronóstico para cada crecida.

Cuando se desee pronosticar se puede realizar un rápido análisis de las salidas (errores) de entrenamiento, validación y prueba (pronóstico) del programa creado en función de las entradas (Q_t), de tal manera de encontrar la crecida que sirva como mejor modelo de pronóstico en términos de ser ocupada para entrenar el modelo neuronal y generar un modelo de pronóstico.

De antemano, son recomendables las crecidas 2, 4, 5, 6, 8, 14.

4.3.1.1.2 Pronóstico mediante Redes Neuronales Prealimentadas

Se plantean modelos neuronales prealimentados con el objetivo de comparar su eficiencia al realizar pronósticos con respecto a modelos neuronales recurrentes.

Analizando las características de la cuenca se observa un tiempo de concentración de aproximadamente 3 hrs., por lo tanto, la variable caudal interesa retardada hasta ese tiempo, en

cuanto al volumen de precipitación se hará lo mismo. Posteriormente, al igual que para la RRS de Elman se evalúa un modelo que además incorpora la variable \dot{Q}_t .

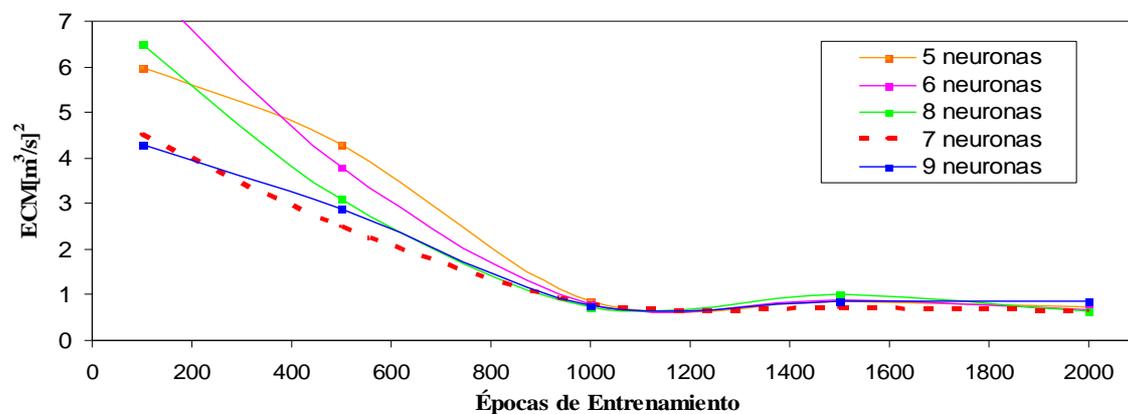
Luego, se evalúan dos tipos de modelos que poseen la siguiente configuración de entrada:

- 8 neuronas de entrada $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$. (Conf. 0)
- 9 neuronas de entrada $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, \dot{Q}_t$. (Conf. 1)

Dada la naturaleza del problema también se emplean modelos de dos capas ocultas.

Se varía el número de neuronas y épocas de entrenamiento de la red con el fin de encontrar las cantidades satisfactorias desde el punto de vista de los errores. En la figura 4.4 se observa que el número óptimo de épocas y de neuronas en las capas ocultas es de 2000 y 7 respectivamente, al igual que para el caso de la red neuronal de Elman.

Figura 4.4: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.



A. Configuración de entrada 0

La matriz de resultados se presenta en las tablas 4.11, 4.12 y 4.13, en términos generales los errores de entrenamiento son del mismo orden que los obtenidos para un modelo de Elman equivalente o de variables de entrada $(P * A)_t$ y Q_t , pero si se observan los errores de validación, se puede ver que son bastante mayores, especialmente los que se obtienen al validar la crecida 2, 4, 10, 11, 12, 19, 20, 21.

En cuanto a los errores ponderados, para cada crecida de entrenamiento se observa un comportamiento consecuente con el aumento de los errores de validación en relación a los obtenidos para una red de Elman, los más bajos varían de 2 a 5 [m³/s]².

Tabla 4.11: Errores de un modelo NEWFF 8-7-7-1.

crecida	Nº de datos	Error de Entrenamiento[m3/s] ²	Errores de Validación[m ³ /s] ²								
		Entrenamiento	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9
			43	125	62	161	174	112	84	103	73
1	43	0,013	0,013	10,592	1,105	49,728	9,602	1,136	1,334	0,875	0,137
2	125	3,546	0,164	3,546	0,507	15,305	2,908	1,173	1,262	0,296	0,295
3	62	0,155	0,112	7,797	0,155	12,730	2,331	0,982	1,218	0,266	0,327
4	161	5,093	0,156	6,093	0,463	5,093	0,862	0,788	0,874	0,094	0,237
5	174	0,848	0,205	7,238	0,620	6,212	0,848	1,113	0,942	0,106	0,278
6	112	0,350	0,085	6,865	0,230	11,220	2,381	0,350	1,054	0,227	0,223
7	84	0,463	0,140	7,981	0,698	10,140	1,391	0,845	0,463	0,097	0,247
8	103	0,074	0,172	8,068	0,548	5,875	0,922	0,934	0,743	0,074	0,237
9	73	0,085	0,078	15,712	1,132	46,369	9,439	1,660	1,604	0,886	0,085
10	22	0,153	0,255	33,456	1,977	67,444	13,522	6,574	3,160	1,441	0,381
11	24	1,036	0,040	11,731	1,451	54,844	11,410	1,472	1,660	1,036	0,224
12	51	0,525	0,083	5,408	0,381	9,411	1,877	0,573	0,855	0,163	0,225
13	210	0,044	0,240	8,422	0,813	6,920	0,963	1,079	0,985	0,097	0,309
14	114	0,018	0,165	8,706	0,639	9,444	1,981	1,009	1,091	0,123	0,338
15	44	0,048	0,185	12,425	1,579	48,380	10,242	2,465	1,358	0,556	0,419
16	56	0,011	0,072	7,053	0,533	22,881	4,764	0,925	0,931	0,320	0,236
17	57	0,001	0,191	9,282	0,344	6,562	0,986	0,408	1,237	0,102	0,342
18	24	0,001	0,222	16,033	1,173	51,874	9,210	4,545	1,335	0,458	1,149
19	14	0,864	0,180	13,221	2,042	45,035	10,156	4,236	2,071	0,744	0,867
20	13	0,053	1,542	96,715	4,932	367,848	59,162	43,097	5,598	2,469	5,003
21	21	0,370	0,246	18,113	1,445	22,704	5,545	4,239	1,847	0,233	0,657

Tabla 4.12: Errores de un modelo NEWFF 8-7-7-1.

crecida	Nº de datos	Errores de Validación[m ³ /s] ²											
		Valid 10	Valid 11	Valid 12	Valid 13	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21
		22	24	51	210	114	44	56	57	24	14	13	21
1	43	1,937	2,305	8,285	1,019	0,391	0,848	0,040	0,039	0,021	22,004	4,879	10,258
2	125	2,433	9,657	3,373	0,261	0,091	0,263	0,050	0,015	0,037	15,665	5,374	7,897
3	62	2,049	4,800	2,532	0,271	0,100	0,223	0,044	0,009	0,020	14,370	4,409	8,070
4	161	2,342	9,628	1,884	0,053	0,031	0,109	0,033	0,006	0,019	13,759	4,499	5,834
5	174	2,356	13,243	2,168	0,066	0,043	0,148	0,044	0,007	0,027	17,375	5,692	6,217
6	112	2,272	4,091	2,708	0,197	0,089	0,179	0,024	0,009	0,012	12,782	3,126	7,176
7	84	2,960	12,373	2,007	0,109	0,056	0,176	0,028	0,006	0,011	11,334	4,109	4,615
8	103	2,568	12,221	1,739	0,052	0,031	0,134	0,034	0,004	0,016	12,225	4,558	4,763
9	73	1,586	5,341	9,652	1,029	0,359	0,719	0,066	0,038	0,027	35,951	6,795	11,011
10	22	0,153	16,268	17,858	1,829	0,518	1,479	0,231	0,051	0,060	70,070	13,736	27,337
11	24	2,262	1,036	7,271	1,235	0,464	0,971	0,038	0,036	0,019	18,287	4,560	8,900
12	51	2,874	4,724	0,525	0,191	0,087	0,249	0,019	0,004	0,013	7,213	3,593	5,674
13	210	2,898	15,903	2,028	0,044	0,035	0,145	0,040	0,005	0,025	16,556	6,250	5,840
14	114	3,199	10,460	2,294	0,110	0,018	0,104	0,027	0,007	0,012	10,720	3,288	5,366
15	44	3,618	10,507	11,381	0,522	0,058	0,048	0,046	0,041	0,017	24,772	4,912	11,064
16	56	2,645	2,817	3,466	0,304	0,109	0,225	0,011	0,014	0,008	10,080	2,357	6,871
17	57	3,367	12,835	1,272	0,049	0,037	0,148	0,032	0,001	0,016	14,352	5,446	6,386
18	24	4,094	8,686	7,090	0,516	0,095	0,171	0,031	0,026	0,001	21,614	3,923	8,959
19	14	3,505	9,011	2,236	1,151	0,417	1,326	0,067	0,023	0,029	0,864	3,541	4,107
20	13	6,879	49,849	17,315	4,824	1,211	2,721	0,264	0,111	0,038	46,464	0,053	11,696
21	21	2,754	24,760	2,195	0,381	0,053	0,661	0,088	0,022	0,046	9,347	7,235	0,370

Tabla 4.13: Errores Ponderados Modelo NEWFF 8-7-7-1.

crecida	Nº de datos	Error Ponderado[m ³ /s] ²
1	43	8,085
2	125	2,985
3	62	2,842
4	161	1,721
5	174	2,075
6	112	2,512
7	84	2,456
8	103	1,979
9	73	8,411
10	22	14,037
11	24	8,898
12	51	2,047
13	210	2,294
14	114	2,533
15	44	8,442
16	56	4,016
17	57	2,194
18	24	8,903
19	14	7,877
20	13	58,163
21	21	5,549

B. Configuración de entrada 1

En las figuras 4.14, 4.15 y 4.16 se observa la matriz de errores, la cual presenta claras disminuciones en los errores de entrenamiento y validación con respecto a una red de este tipo que no involucra la variable \dot{Q}_t . Los errores de entrenamiento se reducen aprox. en un orden de magnitud. Un comportamiento similar presentan los errores de validación exceptuando las crecidas 2, 4, 10, 19, 20 y 21, en las cuales si bien se reduce considerablemente el error en validarlas, no es lo suficiente como para descender en un orden de magnitud. Consecuentemente se aprecia una clara disminución de los errores ponderados.

Al comparar los errores obtenidos con una red de Elman equivalente, en el sentido de involucrar la variable gradiente de caudal, se observa que los errores de entrenamiento en todas las crecidas son menores para una red de tipo Elman, lo mismo se obtiene en cuanto a los errores de validación. Se observa, que en su mayoría todas las crecidas presentan errores de validación más altos en modelos de red prealimentada exceptuando la crecida 10 y la crecida 18, las cuales se ven mejor pronosticadas con un modelo neuronal de este último tipo.

En cuanto a los errores ponderados, la tendencia de los errores de validación a ser mayores que en el caso de usar una red de Elman como modelo de pronóstico se mantiene, exceptuando las crecidas 10, 17 y 19 las cuales como set de datos de entrenamiento definen mejores modelos en una red prealimentada.

Tabla 4.14: Errores de un modelo NEWFF 9-7-7-1.

crecida	Nº de datos	Error de Entrenamiento[m ³ /s] ²	Errores de Validación[m ³ /s] ²																							
			Valid 1			Valid 2			Valid 3			Valid 4			Valid 5			Valid 6			Valid 7			Valid 8		
1	43	0,006	43	125	62	161	174	112	84	103	73	0,006	1,728	0,311	13,602	2,421	0,730	0,373	0,457	0,057						
2	125	0,951	0,049	0,951	0,186	6,589	1,228	0,312	0,459	0,240	0,080															
3	62	0,043	0,025	0,821	0,043	4,285	0,788	0,305	0,260	0,123	0,049															
4	161	1,252	0,088	2,043	0,278	1,252	0,207	0,142	0,355	0,036	0,086															
5	174	0,290	0,116	2,819	0,374	1,946	0,290	0,268	0,417	0,054	0,124															
6	112	0,153	0,067	1,958	0,174	3,312	0,603	0,153	0,445	0,142	0,075															
7	84	0,033	0,005	1,332	0,044	4,964	0,605	0,928	0,033	0,124	0,050															
8	103	0,021	0,127	3,182	0,464	1,764	0,287	0,232	0,415	0,021	0,130															
9	73	0,035	0,064	4,895	0,652	26,337	5,296	1,104	0,926	0,738	0,035															
10	22	0,027	0,186	14,963	0,634	41,179	5,943	5,719	1,020	1,008	0,715															
11	24	0,248	0,009	1,550	0,336	13,929	2,592	0,654	0,438	0,478	0,068															
12	51	0,093	0,082	2,032	0,251	1,019	0,187	0,070	0,394	0,019	0,082															
13	210	0,022	0,253	6,253	0,993	5,331	0,762	0,752	0,732	0,061	0,318															
14	114	0,005	0,165	6,347	0,390	8,049	1,554	2,371	0,464	0,028	0,260															
15	44	0,014	0,077	2,735	0,253	5,221	0,754	0,915	0,387	0,059	0,111															
16	56	0,017	0,454	32,268	1,791	65,329	11,275	9,502	1,347	0,421	1,456															
17	57	0,007	0,581	30,664	2,333	50,897	7,992	8,031	1,672	0,362	1,544															
18	24	0,008	0,972	72,844	4,059	136,672	23,224	21,340	2,695	0,900	3,377															
19	14	0,055	0,007	0,357	0,036	2,149	0,292	0,425	0,083	0,043	0,013															
20	13	0,079	0,517	44,432	2,629	70,005	12,031	11,805	1,706	0,718	2,110															
21	21	0,048	0,010	20,388	0,044	22,918	1,835	4,316	0,032	0,487	0,056															

Tabla 4.15: Errores de un modelo NEWFF 9-7-7-1.

crecida	Nº de datos	Errores de Validación[m ³ /s] ²											
		Valid 10	Valid 11	Valid 12	Valid 13	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21
1	43	0,702	0,387	3,674	0,416	0,282	0,595	0,352	0,739	0,690	4,818	2,457	2,061
2	125	1,020	2,935	1,935	0,197	0,142	0,351	0,172	0,342	0,323	4,327	1,278	2,646
3	62	1,006	1,231	2,148	0,139	0,045	0,141	0,086	0,396	0,334	2,484	1,257	2,242
4	161	1,645	5,577	0,439	0,034	0,020	0,051	0,023	0,164	0,116	5,694	1,032	2,415
5	174	1,745	7,596	0,585	0,039	0,034	0,084	0,032	0,160	0,115	7,930	1,540	2,698
6	112	1,375	3,899	0,869	0,113	0,097	0,204	0,102	0,242	0,196	5,814	1,223	2,610
7	84	0,838	0,351	3,334	0,217	0,034	0,188	0,152	0,631	0,581	1,254	2,247	0,159
8	103	2,105	8,458	0,337	0,014	0,012	0,037	0,011	0,103	0,061	6,978	1,361	2,548
9	73	1,098	4,223	6,097	0,546	0,463	0,836	0,444	0,602	0,542	21,053	5,289	6,909
10	22	0,027	11,985	18,631	1,144	0,274	1,222	0,821	2,507	2,423	20,564	18,256	4,606
11	24	0,779	0,248	3,216	0,417	0,324	0,682	0,390	0,686	0,644	4,084	2,092	2,086
12	51	2,001	5,087	0,093	0,016	0,017	0,042	0,013	0,100	0,066	4,557	0,744	2,572
13	210	2,785	17,147	1,151	0,022	0,030	0,101	0,022	0,059	0,037	15,207	3,409	4,009
14	114	2,328	8,579	0,383	0,086	0,005	0,023	0,027	0,203	0,141	6,694	0,870	3,050
15	44	1,610	4,546	1,214	0,096	0,013	0,014	0,012	0,135	0,078	6,680	1,480	3,179
16	56	2,928	15,945	6,109	0,713	0,065	0,070	0,017	0,080	0,043	18,543	4,228	7,142
17	57	3,713	29,017	5,676	0,539	0,077	0,180	0,073	0,007	0,030	28,651	7,133	7,147
18	24	4,735	32,744	11,054	1,583	0,119	0,143	0,105	0,032	0,008	30,014	6,677	10,268
19	14	1,167	0,305	1,747	0,107	0,013	0,097	0,071	0,345	0,298	0,055	0,528	0,702
20	13	3,290	14,629	4,195	1,142	0,138	0,246	0,214	0,210	0,168	4,532	0,079	2,622
21	21	0,713	0,917	6,868	1,450	0,025	0,240	0,178	0,733	0,680	4,346	6,235	0,048

Tabla 4.16: Errores Ponderados de un modelo NEWFF 9-7-7-1.

crecida	Nº de datos	Error Ponderado[m ³ /s] ²
1	43	2,264
2	125	1,227
3	62	0,845
4	161	0,587
5	174	0,815
6	112	0,857
7	84	0,983
8	103	0,812
9	73	4,615
10	22	8,126
11	24	2,282
12	51	0,525
13	210	1,835
14	114	2,014
15	44	1,197
16	56	12,197
17	57	10,474
18	24	25,788
19	14	0,445
20	13	13,762
21	21	5,051

4.3.1.1.3 Análisis y Comparación de Modelos Neuronales

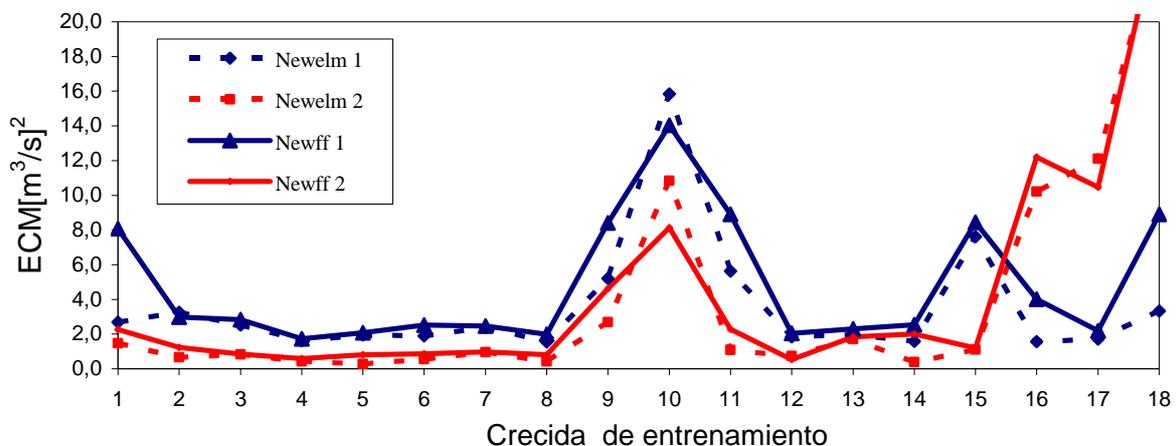
A continuación, se esquematiza el comportamiento de los errores de validación y ponderados, así como también de ciertos índices que cuantifican la eficiencia de un modelo de pronóstico en el valor máximo. De ahora en adelante se entiende por Newelm 1. (Newff 1.), el modelo de red recurrente (prealimentada) que involucra caudales y precipitaciones en el tiempo t como variables de entrada y por Newelm 2. (Newff 2.) el modelo que además de las variables anteriores involucra el gradiente.

En la figura 4.5 se observan los errores ponderados de los 4 modelos analizados (los 2 de Elman y los 2 prealimentados), la serie en sí unida y representada como una curva no tiene ningún significado físico y sólo es una esquematización de los errores obtenidos de tal manera de facilitar al lector el entendimiento del comportamiento de dichos errores en todo el conjunto de crecidas. No se presentan como crecidas de entrenamiento las crecidas 19,20 y 21, por lo altos errores obtenidos en un modelo Newelm y Newff 1, que no permiten apreciar el comportamiento de las demás (ver por ejemplo tabla 4.11, crecida 20, valid. 4).

Se observa, según se desprendió unitariamente que la crecida 10 es una crecida que no serviría para entrenar un modelo de pronóstico, puesto que en todos los modelos planteados presenta validaciones con errores significativamente grandes, al contrario un modelo entrenado con los datos de la crecida 4 presenta bajos errores de validación en todas las crecidas.

Se deduce que un modelo de Elman del tipo 3-7-7-1, entrenado con los datos de la crecida 4 ó 5, con las variables $(P \cdot A)_t$, Q_t y \dot{Q}_t como variables de entrada es el mejor modelo de pronóstico.

Figura 4.5: Errores Ponderados Modelos NEWELM y NEWFF-Validación.



En las figuras 4.6.a y 4.6.b se presenta el comportamiento de los errores ponderados de dos modelos formados a partir del entrenamiento con los datos de las crecidas 8 y 12 en función del

número de épocas de entrenamiento para una configuración tipo Newelm 2. y Newff 2. Se aprecia claramente que independiente del número de épocas de entrenamiento un modelo Newelm presenta errores ponderados más bajos y por lo tanto es un mejor modelo de pronóstico.

Figura 4.6.a: Errores Ponderados Crecida 8 Modelos NEWELM y NEWFF.

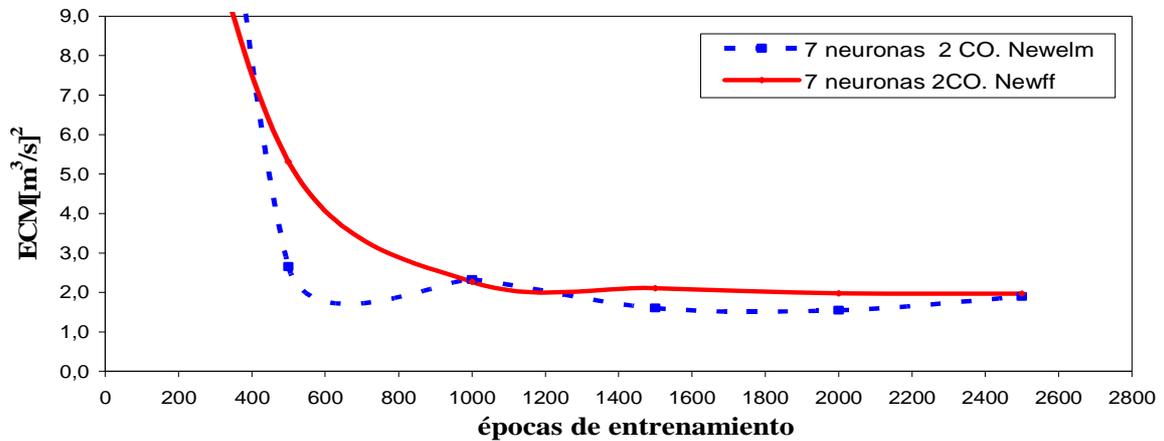
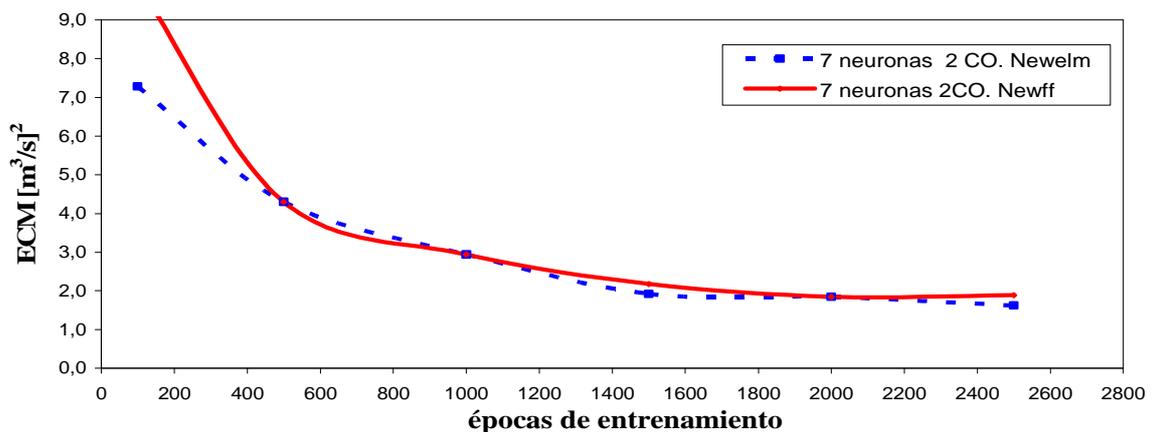


Figura 4.6.b: Errores Ponderados Crecida 12 Modelos NEWELM y NEWFF.



En las figuras 4.7 y 4.8 se observa el comportamiento de los errores de validación para cada crecida de entrenamiento de acuerdo a un modelo Newelm 2.y a un modelo Newff 2., la línea continua se entiende como el comportamiento que se obtiene para una red prealimentada y la línea punteada para una red recurrente, los colores asocian la crecida validada (indicado por el número x en Newelm x.2). De estas figuras se desprende la mejor capacidad de validación de la RRS de Elman, ya que las curvas de la red prealimentada siempre se encuentran sobre las curvas de Elman.

Nuevamente, Las crecidas 19,20 y 21 no se presentan como crecidas de entrenamiento debido a que los errores que se obtienen al validar crecidas mediante dichos modelos se encuentran en un rango que impide apreciar las demás. Ver tablas matriciales, por ejemplo tabla 4.14.

Figura 4.7: Errores de Validación Modelos NEWELM y NEWFF.

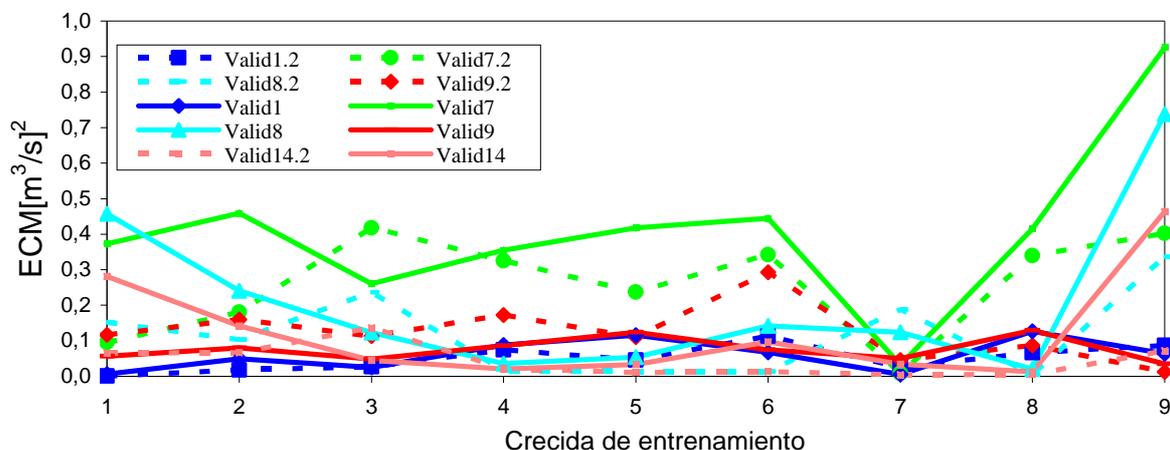
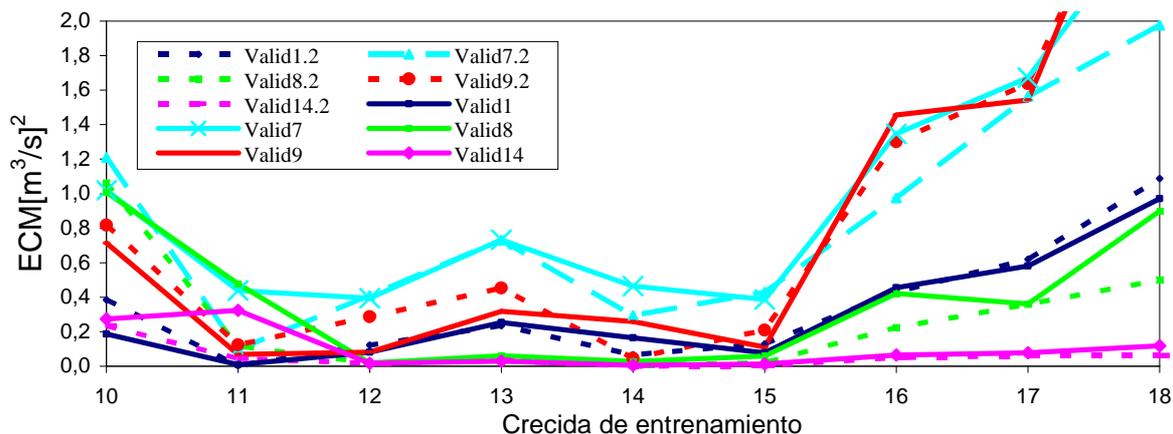


Figura 4.8: Errores de Validación Modelos NEWELM y NEWFF.



A continuación en las figuras 4.9 a 4.11 se presentan los índices especificados en el capítulo 2.6.2, que guardan relación con el pronóstico de caudales máximos en el entrenamiento. Se grafica el comportamiento de dichos índices en los 4 modelos planteados, encontrándose que un modelo de RRS tipo Elman 3-7-7-1 es mejor también en torno al pronóstico de máximos.

Cabe destacar el hecho de que algunos valores de los índices EMEM y EMRMOBS sean negativos indica que el máximo observado es mayor que el simulado, por lo tanto se está subestimando y que un valor de TMEM negativo quiere decir que el pronóstico se encuentra retrasado, lo cual es extensivo a todos los análisis de este tipo que se realizan.

Un resultado muy importante es que si se observa la figura 4.11, se desprende que para el modelo Newelm 3-7-7-1 prácticamente no existen desfases entre la ocurrencia del caudal

máximo observado y pronosticado, por lo que EMEM y EMRMOBS, definen prácticamente lo mismo, “son lo mismo”, es decir, la diferencia porcentual entre el caudal máximo pronosticado y observado en el instante en que se producen. Dicho comportamiento no se observa para ningún otro modelo y es de mucha utilidad a la hora de realizar pronósticos.

Se observa que EMEM y EMRMOBS son inferiores al 5% para modelos Newelm 3-7-7-1, valores que crecen a un rango de 1% hasta 10% en modelos del tipo Neweff 9-7-7-1 y significativamente más grandes (hasta un 20%) en modelos Newelm y Neweff que no involucran como parámetro de entrada la variable \dot{Q}_t .

Claramente para modelos entrenados con la crecida 4 se obtiene el mínimo error en el pronóstico de máximos de entrenamiento, con mayor notoriedad en un modelo Newelm 3-7-7-1. Destacándose también los modelos entrenados con la crecida 18, que no presenta una alta capacidad de generalización por lo que los errores cometidos en el entrenamiento son bajos, no así los de validación como se observa más adelante.

Figura 4.9: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.

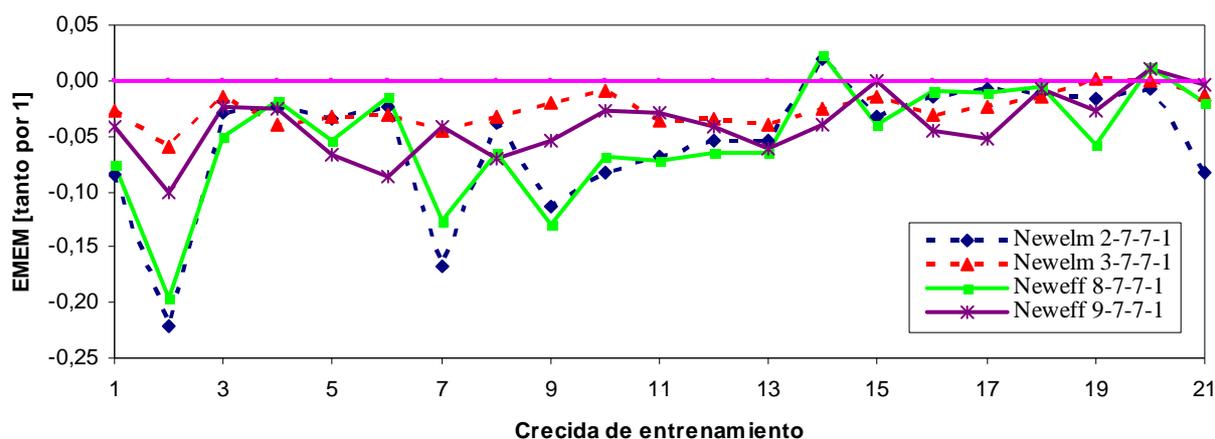


Figura 4.10: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.

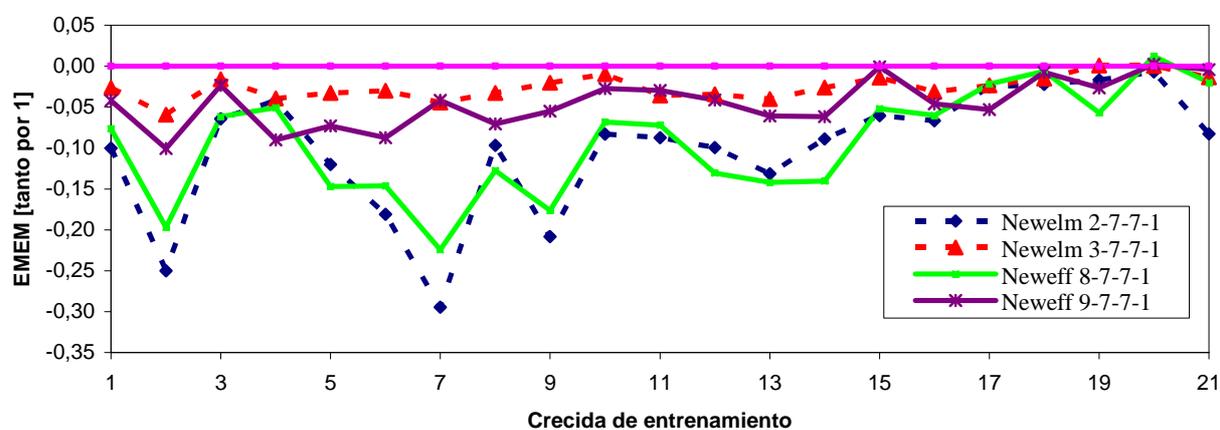
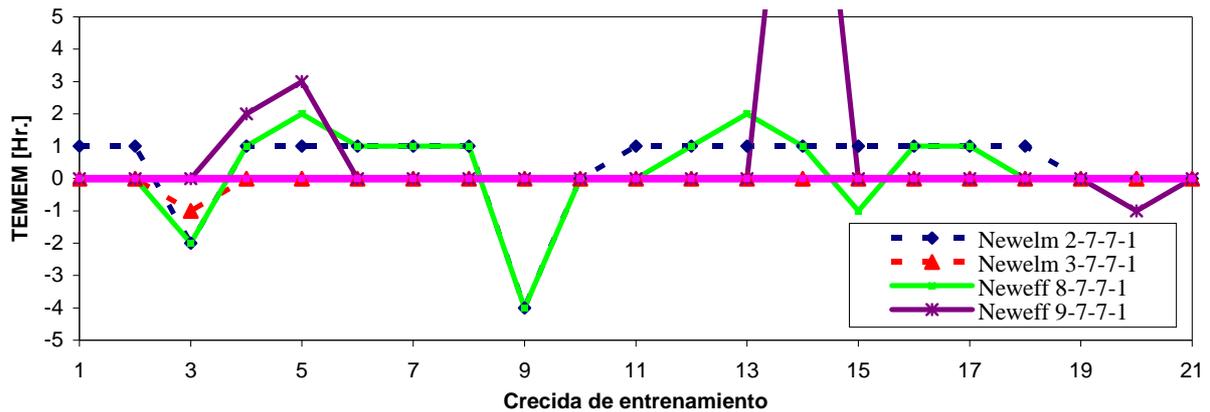


Figura 4.11: Índice TMEM entrenamiento. Modelos NEWELM y NEWFF.



En las figuras 4.12, 4.13 y 4.14 se presentan los mismos índices de evaluación de pronósticos del caudal máximo para una crecida perteneciente al conjunto de validación, a modo de ejemplo de lo que sucede con dichos índices en tal conjunto. La crecida evaluada es la número 5, a partir de modelos formados del entrenamiento con el resto de las crecidas. Se observa que los índices EMEM y EMRMOBS son en promedio de un 15 % y un 20% del caudal máximo observado respectivamente en todos los modelos, exceptuando aquellos formados a partir de la crecida 10 de entrenamiento, si se observa dichos índices para un modelo Newelm 3-7-7-1 ambos valores disminuyen a aprox. un 10% en promedio.

El índice TMEM para el modelo Newelm 3-7-7-1 presenta el menor desfase en la ocurrencia de máximos, siendo nulo para dicho modelo entrenado con los datos de la crecida 4. En particular en modelos formados por el entrenamiento de las crecidas 7, 10, 17 y 20 es igual o superior a 1 hora, lo cual resulta ineficiente, desde el punto de vista de pronósticos a un paso.

Figura 4.12: Índice EMEM validación crecida 5. Modelos NEWELM y NEWFF.

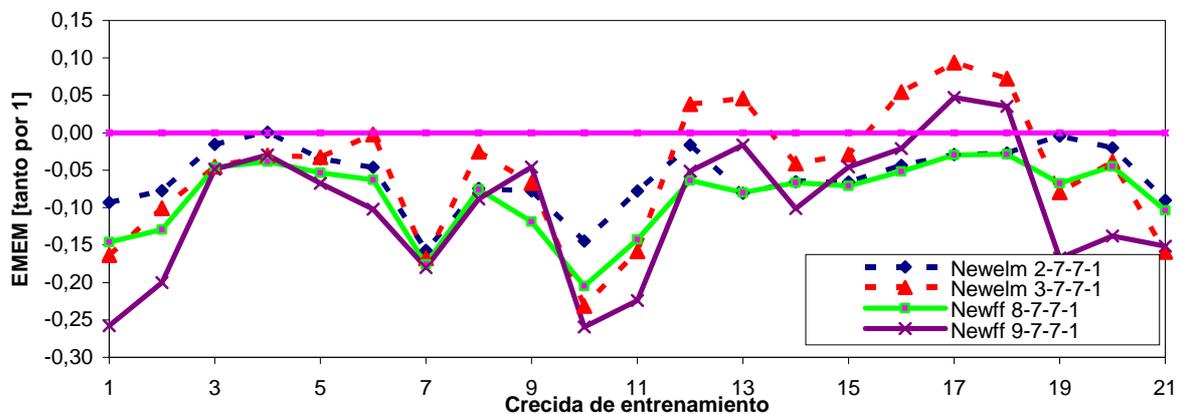


Figura 4.13: Índice EMRMOBS validación crecida 5. Modelos NEWELM y NEWFF.

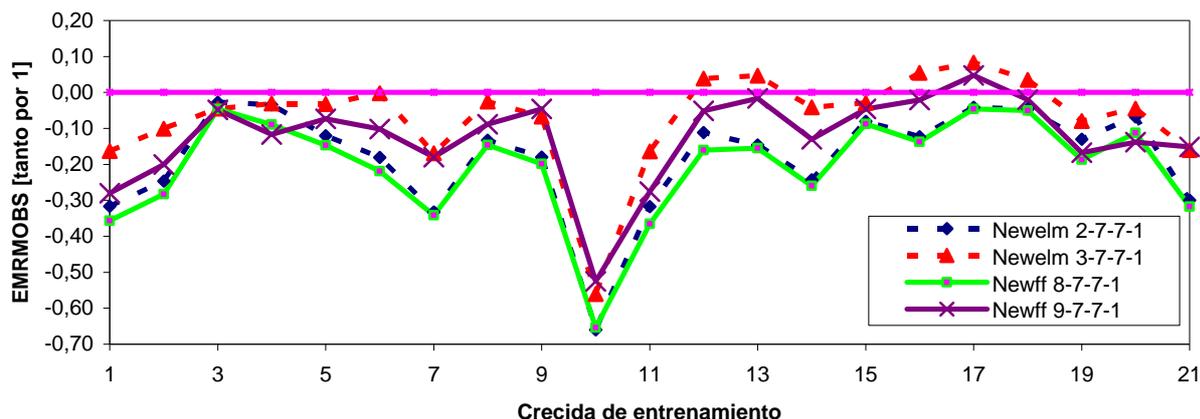
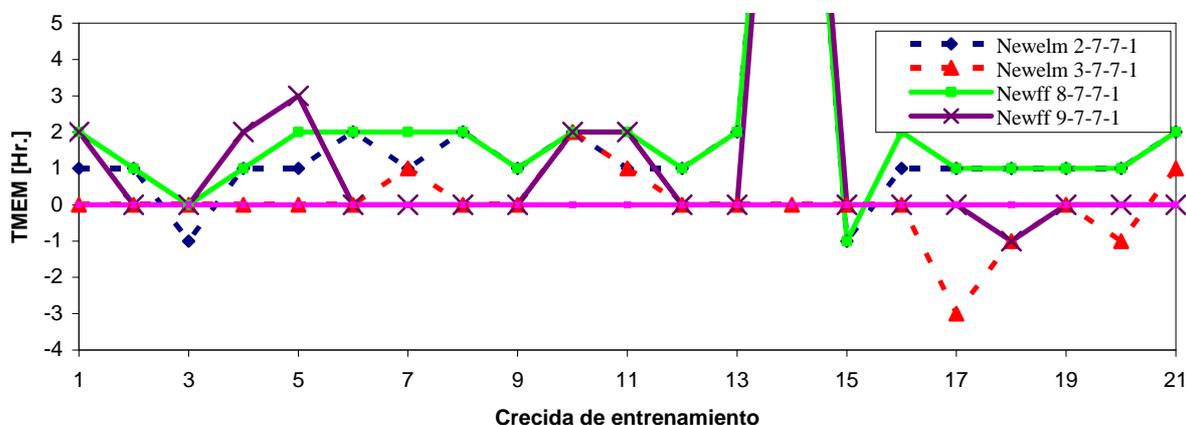


Figura 4.14: Índice TEMEM validación crecida 5. Modelos NEWELM y NEWFF.



4.3.1.2 Pronóstico mediante Redes Neuronales de pronóstico -Entrenamiento Continuo

4.3.1.2.1 Pronóstico mediante Redes Neuronales Recurrentes de Elman

De acuerdo al planteamiento general, se forman tres conjuntos, uno de entrenamiento, otro de validación y un último de prueba, los cuales contienen 853 (crecidas 1 ,2, 3, 4, 5, 6,7 y 8), 531 (crecidas 9, 10, 11, 12, 13, 14, 15) y 179 (crecidas 16, 17, 18, 19, 20, 21) datos respectivamente.

Las características de cada uno de los conjuntos de entrenamiento se presentan en la tabla 4.17.

Tabla 4.17: Caracterización conjuntos de datos.

CONJUNTO	MEDIA [m ³ /s]	DESV ESTANDAR [m ³ /s]	Q Máx [m ³ /s]	QMín [m ³ /s]
ENTRENAMIENTO	10,6	10,0	64,2	0,6
VALIDACIÓN	4,9	3,1	27,5	1,3
PRUEBA	5,3	2,8	19,8	2,0

Se realizan modelos de redes neuronales que involucran dos capas ocultas, nuevamente debido a la naturaleza del fenómeno. Se realizan configuraciones, considerando las mismas entradas para un entrenamiento unitario del mismo tipo de modelo, las cuales son:

- P_t^*A y Q_t (Configuración 0).
- P_t^*A , Q_t y \dot{Q}_t (configuración 1).

Se prueban arquitecturas que consideran de 1 a 8 neuronas en cada una de las capas ocultas, de manera de encontrar aquella que entregue el menor error cuadrático medio ponderado, definiendo la mejor arquitectura de un modelo de pronóstico de este tipo.

A. Configuración de entrada 0 y 1

Las figuras 4.15 y 4.16 muestran el análisis de los errores de validación en función del número de épocas y neuronas de las capas ocultas. Se observa que existen dos tipos de “buenos” modelos, uno que involucra 4 neuronas en la capa oculta y un óptimo de 750 épocas de entrenamiento y otro que considera 7 neuronas en cada capa oculta para un óptimo de 2000 épocas de entrenamiento. Comparando los errores obtenidos (Tabla 4.18) se deduce que el modelo óptimo considera 7 neuronas en la capa oculta.

Las tablas 4.19 y 4.20 muestran los valores de los errores obtenidos en cada uno de los set de datos considerados para distintas arquitecturas de RNR y ambos tipos de configuraciones de entrada. Se observa que los errores obtenidos varían entre 1 y 2 $[\text{m}^3/\text{s}]^2$, encontrándose que la mejor configuración involucra 7 neuronas en cada capa oculta.

El valor del error (recordar, ECM) en un modelo que posee una configuración de entrada que no involucra la variable \dot{Q}_t es mayor aprox. en un 50% del valor obtenido para una configuración que si involucra dicha variable, se desprende al igual que para modelos unitarios la importancia de la variable gradiente de caudal para realizar pronósticos con modelos de redes neuronales.

Figura 4.15: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.

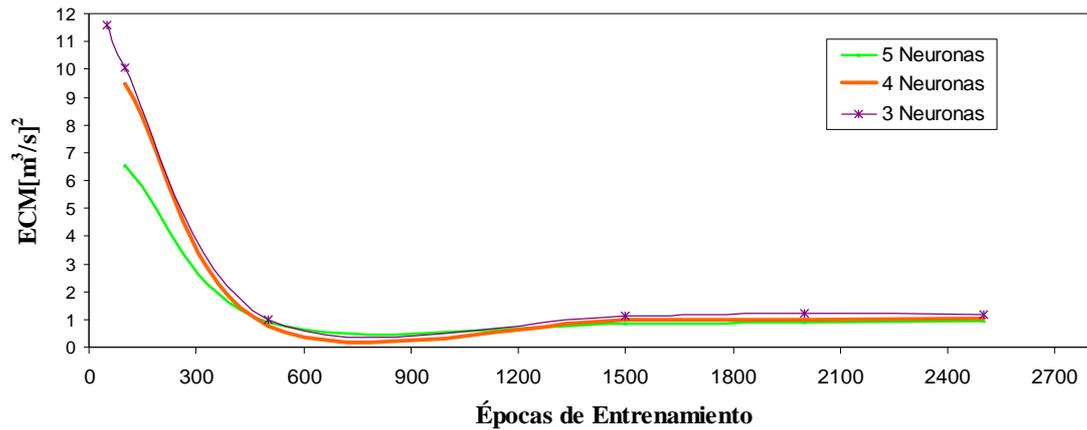


Figura 4.16: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.

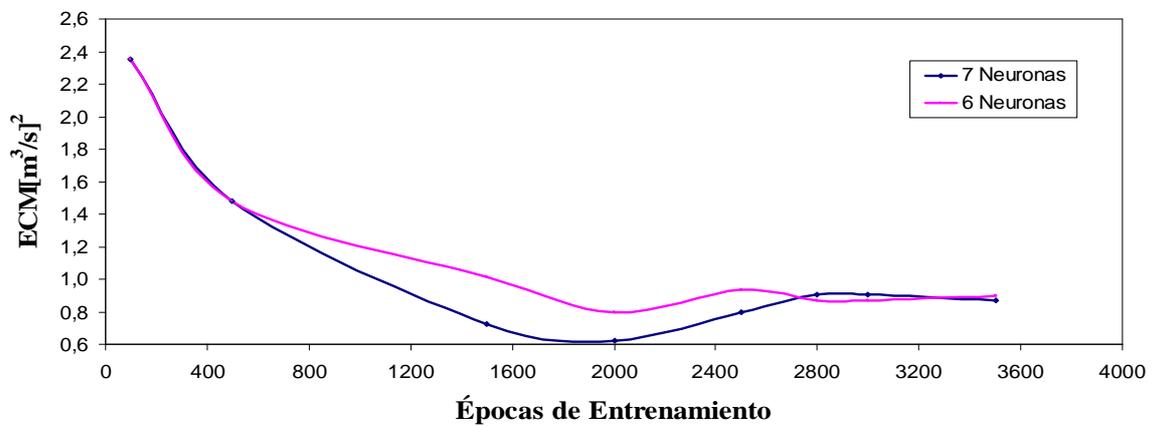


Tabla 4.18: Comparación de errores de un modelo NEWELM 3-4-4-1 y un modelo NEWELM 3-7-7-1.

Nº de neuronas	ERRORES [m³/s]²				r²		
	entrenamiento	Validación	Prueba	Ponderado	r² entrenamiento	r² validación	r² prueba
4	3,319	0,961	2,036	2,371	0,961	0,880	0,645

Nº de neuronas	ERRORES [m³/s]²				r²		
	entrenamiento	Validación	Prueba	Ponderado	r² entrenamiento	r² validación	r² prueba
7	2,275	0,623	1,735	1,652	0,976	0,918	0,744

Tabla 4.19: Errores de un modelo NEWELM 2-7-7-1.

Configuración	ERRORES [m³/s]²				r²		
	2000 épocas de entrenamiento			Ponderado	r² entrenamiento	r² validación	r² prueba
	Entrenamiento	Validación	Prueba				
	853	531	179				
2-4-4-1	5,305	1,630	3,399	3,838	0,936	0,795	0,480
2-5-5-1	3,376	1,440	3,151	2,692	0,950	0,829	0,551
2-6-6-1	2,393	1,141	2,587	1,990	0,963	0,869	0,649
2-7-7-1	2,054	1,057	2,471	1,763	0,967	0,881	0,682
2-8-8-1	2,720	1,111	2,776	2,180	0,951	0,874	0,689
2-7-6-1	3,060	1,058	2,222	2,284	0,967	0,878	0,690
2-7-8-1	2,900	1,003	2,326	2,190	0,969	0,888	0,714
2-6-7-1	3,250	1,258	2,577	2,496	0,965	0,855	0,654
2-5-7-1	3,759	1,265	2,858	2,808	0,958	0,854	0,612
2-7-4-1	4,693	1,493	3,019	3,414	0,948	0,825	0,581

Tabla 4.20: Errores de un modelo NEWELM 3-7-7-1.

Configuración	ERRORES [m ³ /s] ²				r ²		
	2000 épocas de entrenamiento			Ponderado	r ² entrenamiento	r ² validación	r ² prueba
	Entrenamiento	Validación	Prueba				
	853	531	179				
3-4-4-1	3,468	0,996	2,280	2,492	0,959	0,874	0,599
3-5-5-1	1,983	0,916	2,162	1,641	0,966	0,890	0,644
3-6-6-1	1,572	0,796	1,910	1,347	0,972	0,908	0,713
3-7-7-1	1,275	0,623	1,735	1,106	0,976	0,918	0,744
3-8-8-1	1,351	1,119	2,315	1,382	0,975	0,871	0,713
3-7-6-1	2,657	0,861	1,974	1,969	0,972	0,903	0,715
3-7-8-1	2,699	0,985	2,701	2,117	0,972	0,890	0,674
3-6-7-1	2,802	1,080	2,341	2,164	0,970	0,877	0,686
3-5-7-1	2,552	0,829	2,055	1,910	0,972	0,904	0,699
3-7-4-1	2,812	0,854	2,061	2,061	0,969	0,900	0,678

En la tabla 4.21 se pueden apreciar los índices de error que se cometen al pronosticar el caudal máximo con respecto al valor observado para cada conjunto de datos.

En general, se observan subestimaciones del caudal máximo para las dos configuraciones de entrada. El modelo NEWELM 3-7-7-1, que entrega los menores errores, pronostica el caudal máximo en aprox. un 7% menos que el valor observado durante el entrenamiento, valor que aumenta aprox. a un 9% en promedio entre el conjunto de validación y prueba; en cuanto a los errores cometidos al tiempo en que ocurre el máximo real dichos valores aumentan para estos últimos conjuntos a un promedio del 23% del caudal observado, valores que son aún mayores en el caso de considerar un modelo Newelm 2-7-7-1.

A nivel de entrenamiento continuo, la variable gradiente de caudal no causa el mismo efecto que a nivel unitario, puesto que el caudal máximo se estima con 1 hora de desfase en promedio, lo que puede deberse a la inestabilidad que provocan las series pegadas..

Tabla 4.21: Índices de error en el pronóstico de caudal máximo de modelos NEWELM.

Datos	ENTRENAMIENTO-VALIDACIÓN-PRUEBA - 2000 épocas					
	Modelo 2-7-7-1 NEWELM			Modelo 3-7-7-1 NEWELM		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]
Entrenamiento	-0,105	-0,140	1	-0,065	-0,066	1
Validación	-0,197	-0,388	1	-0,107	-0,290	1
Prueba	-0,100	-0,203	1	-0,087	-0,180	1

4.3.1.2.2 Pronóstico mediante Redes Neuronales Prealimentadas

Se consideran dos configuraciones de entrada equivalentes a las que se utilizan en la red de Elman, en el sentido de que se considera las mismas variables sólo que además retardadas de acuerdo al tiempo de concentración de la cuenca, con el fin de hacer los modelos comparables.

La figura 4.17 muestra el análisis de los errores de validación en función del número de épocas y neuronas de las capas ocultas. Se observa que la envolvente mínima de errores considera 7 neuronas en cada capa oculta para un óptimo de 2000 épocas de entrenamiento, al igual que para una red recurrente de Elman.

En la tabla 4.22 y 4.23 se muestran los valores de los errores obtenidos en cada uno de los set de datos considerados para ambos tipos de configuraciones de entrada y distintas arquitecturas de RNR.

Se desprende, al igual que para modelos unitarios que la mejor configuración es la que involucra 9 neuronas de entrada y 7 neuronas en cada capa oculta.

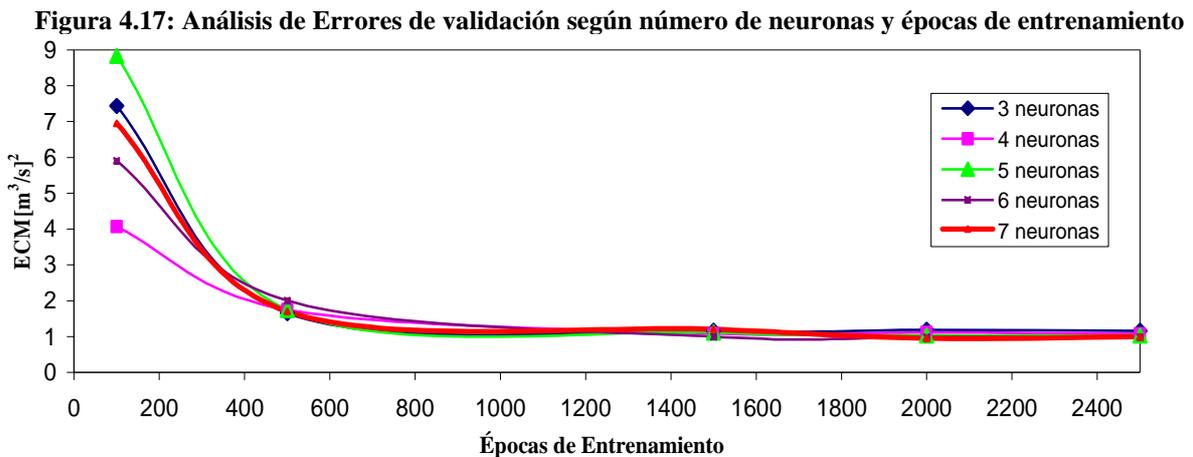


Tabla 4.22: Errores de un modelo NEWFF 8-7-7-1.

Configuración	ERRORES [m³/s]²				r²		
	2000 épocas de entrenamiento			Ponderado	r² entrenamiento	r² validación	r² prueba
	Entrenamiento	Validación	Prueba				
	853	531	179				
2-4-4-1	4,736	1,530	3,049	3,454	0,943	0,805	0,522
2-5-5-1	4,244	1,445	2,957	3,145	0,950	0,822	0,554
2-6-6-1	3,841	1,378	2,895	2,896	0,956	0,834	0,579
2-7-7-1	3,306	1,275	2,800	2,558	0,964	0,851	0,617
2-8-8-1	3,476	1,265	2,738	2,640	0,961	0,850	0,614
2-7-6-1	3,604	1,277	2,712	2,711	0,960	0,848	0,606
2-7-8-1	3,648	1,303	2,738	2,747	0,960	0,845	0,610
2-6-7-1	3,736	1,329	2,810	2,812	0,958	0,841	0,591
2-5-7-1	4,647	1,500	2,891	3,377	0,947	0,818	0,573
2-7-4-1	4,019	1,381	2,845	2,988	0,954	0,833	0,580

Tabla 4.23: Errores de un modelo NEWFF 9-7-7-1.

Configuración	ERRORES [m ³ /s] ²				r ²		
	2000 épocas de entrenamiento			Ponderado	r ² entrenamiento	r ² validación	r ² prueba
	Entrenamiento	Validación	Prueba				
	853	531	179				
3-4-4-1	3,405	1,113	2,168	2,485	0,960	0,859	0,641
3-5-5-1	2,922	1,032	2,098	2,186	0,967	0,873	0,663
3-6-6-1	2,668	0,992	2,050	2,028	0,970	0,880	0,680
3-7-7-1	2,394	0,910	1,922	1,836	0,974	0,893	0,714
3-8-8-1	2,480	0,949	1,987	1,904	0,973	0,887	0,697
3-7-6-1	2,591	0,962	2,003	1,970	0,971	0,885	0,691
3-7-8-1	3,146	1,106	2,159	2,340	0,966	0,873	0,720
3-6-7-1	2,586	0,976	2,022	1,975	0,971	0,883	0,687
3-5-7-1	2,743	0,998	2,054	2,071	0,969	0,879	0,676
3-7-4-1	2,716	0,994	2,053	2,055	0,970	0,880	0,679

Al comparar los errores cometidos en las mejores arquitecturas neuronales para modelos de Elman y prealimentados con entrenamientos de tipo continuo y configuraciones de entrada equivalentes, se observa que un modelo de Elman presenta errores más bajos que un modelo prealimentado, diferencia de aprox. un 40% con respecto al valor del modelo prealimentado. Ventaja que se complementa además, con la cantidad de variables de entrada necesarias para realizar pronósticos.

En la tabla 4.24 se presentan los índices de error de pronosticar el caudal máximo con ambos tipos de configuraciones para cada conjunto de datos.

Se observa claramente que un modelo que involucra la variable \dot{Q}_i arroja menores diferencias entre el máximo pronosticado y observado, pero igualmente subestima dicho valor. Individualmente, para el modelo prealimentado elegido, NEWFF 9-7-7-1, la diferencia entre el caudal pronosticado y observado en el conjunto de entrenamiento es de aprox. un 7 % del valor de este último y de un 16 % en promedio en el conjunto de validación y prueba, porcentaje que crece aprox. a un 26 % para el pronóstico de caudal en el tiempo en que se produce el máximo observado.

En términos generales, comparando los porcentajes de estos últimos modelos en relación a modelos recurrentes de Elman equivalentes, se desprende que para modelos neuronales prealimentados entrenados continuamente se presentan subestimaciones de caudal máximo bastante superiores a los obtenidos para modelos recurrentes de Elman.

En cuanto al desfase que se produce en el pronóstico de máximos, la variable gradiente de caudal no tiene el mismo efecto en un análisis de tipo continuo que en un análisis de tipo unitario, ya que como se observa, los desfases son en promedio de 1 hora.

Tabla 4.24: Índices de eficiencia de pronósticos de caudal máximo.

ENTRENAMIENTO-VALIDACIÓN-PRUEBA - 2000 épocas						
Datos	Modelo 8-7-7-1 NEWFF			Modelo 9-7-7-1 NEWFF		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]
Entrenamiento	-0,127	-0,183	1	-0,065	-0,066	1
Validación	-0,233	-0,451	1	-0,215	-0,394	1
Prueba	-0,117	-0,236	1	-0,117	-0,179	1

De acuerdo a los índices precedentes calculados para modelos recurrentes de Elman y prealimentados, se desprende que los primeros presentan mejores pronósticos que los segundos, independientemente del tipo de entrenamiento, tanto en valor como en calidad.

La eficiencia del pronóstico con respecto al desfase entre el caudal máximo observado y pronosticado, se debe a la incorporación de la variable gradiente de caudal con respecto al instante anterior de pronóstico, pero no tiene el mismo efecto en ambos tipos de entrenamiento, siendo mejor en un análisis de tipo unitario, la razón de ello puede ser la inestabilidad que crea en el pronóstico la serie “continua” o pegada de crecidas, mayormente en las RNR.

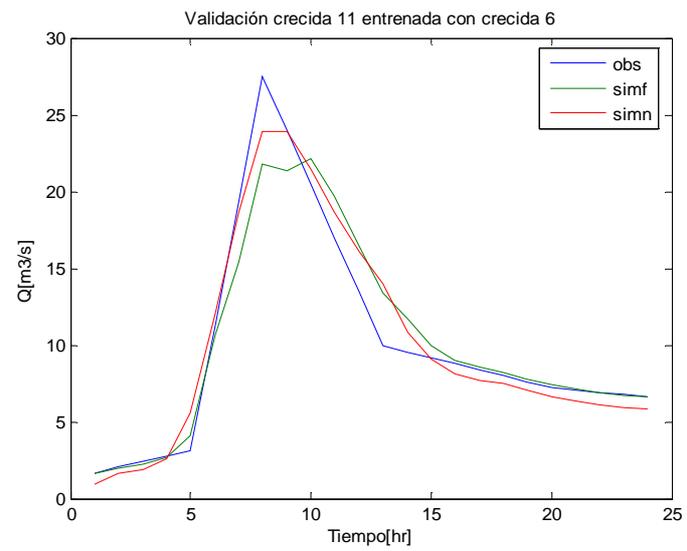
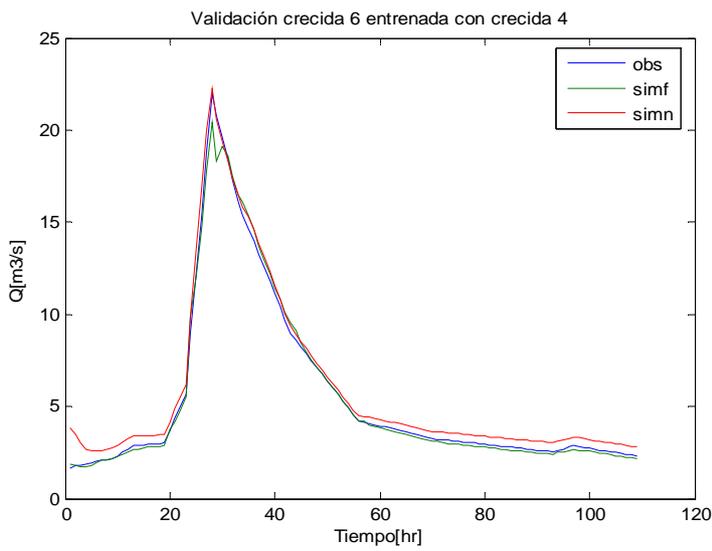
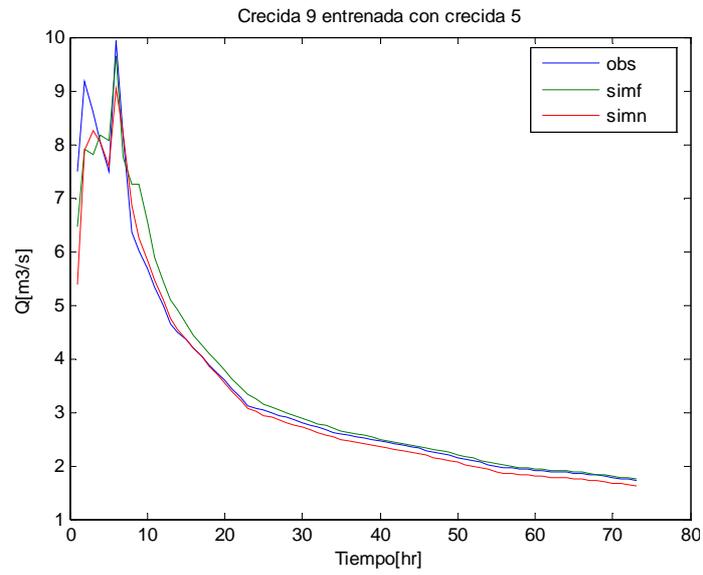
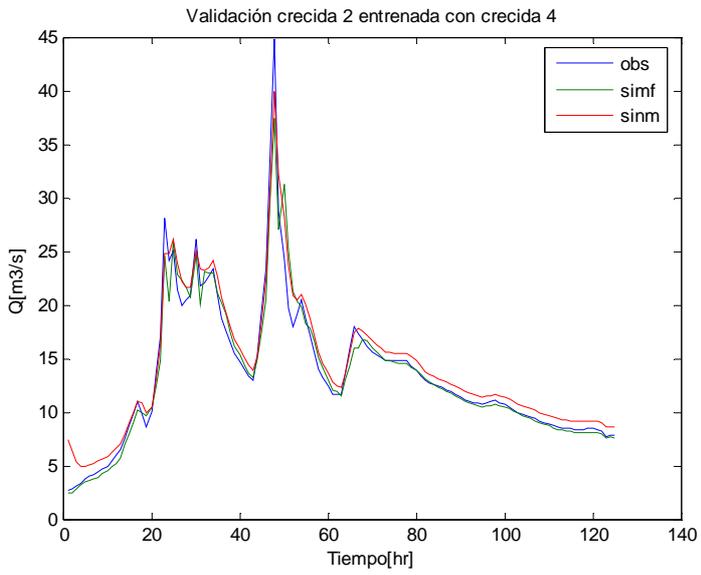
4.3.1.3 Comparación de modelos neuronales de pronóstico por crecidas

A continuación en la figura 4.18 se presentan una serie de crecidas validadas con los modelos neuronales Newelm 3-7-7-1 y Newff 9-7-7-1 formados de acuerdo a un entrenamiento de tipo unitario. Se puede apreciar gráficamente las ventajas de realizar pronósticos mediante redes recurrentes.

En la figura 4.19 se aprecian validaciones realizadas con diferentes modelos Newelm, de acuerdo a la crecida de entrenamiento, y con las variables de entrada definidas por la configuración 0 (sin gradiente (s/d)) y por la configuración 1 (con gradiente(c/d)). Se aprecian las buenas validaciones obtenidas al entrenar con las crecidas 4, 5, 6 y la poca capacidad de generalización de las crecidas que menores amplitudes de frecuencia presentan, 16, 17, 18, 19, 20 y 21.

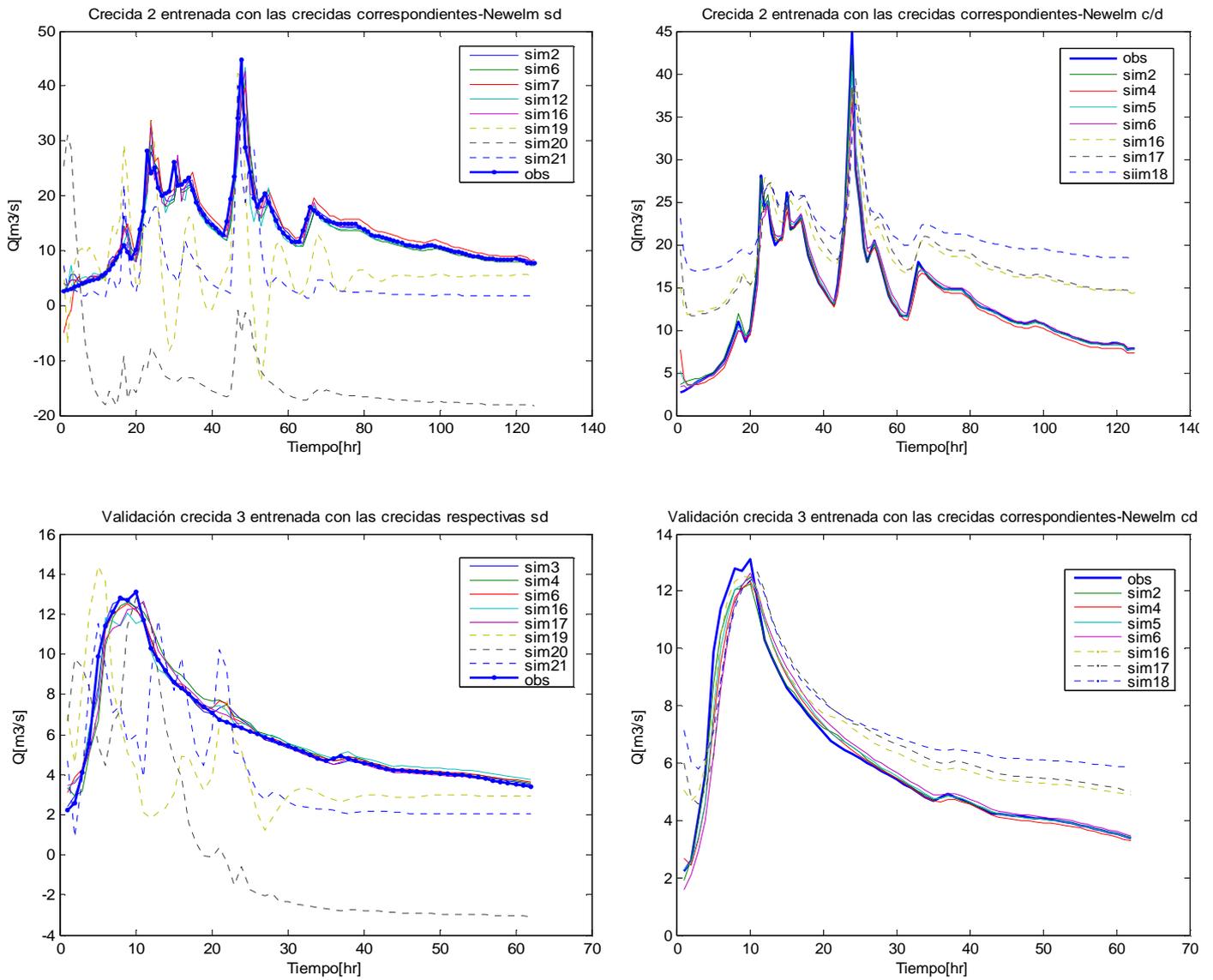
Con respecto a las crecidas número 16, 17 y 18, al ser usadas como datos de entrenamiento, las validaciones que realizan son mejores en modelos que no involucran \dot{Q} como variable de entrada.

Figura 4.18: Pronósticos de crecidas Modelos NEWELM y NEWFF³.



³ En los gráficos, crecida x entrenada con las crecidas correspondientes, quiere decir que la crecida x es validada para modelos entrenados con las crecidas correspondientes.

Figura 4.19: Pronósticos de crecidas Modelos NEWELM cd. y NEWELM sd.⁴



⁴ En los gráficos, crecida x entrenada con las crecidas correspondientes, quiere decir que la crecida x es validada para modelos entrenados con las crecidas correspondientes.

4.3.1.4 Pronóstico mediante Modelos Paramétricos

Los tres conjuntos empleados para realizar pronósticos mediante modelos neuronales entrenados continuamente, se utilizan para realizar un análisis de pronóstico mediante modelos paramétricos tipo ARMAX y RARMAX. En este apartado a un paso de adelanto y con el objetivo de compararse con modelos neuronales.

Las variables de entrada son las consideradas para modelos neuronales de Elman, P_t^*A , Q_t y \dot{Q}_t , y por supuesto la variable de salida es Q_{t+1}

Los retardos de caudal (na), volumen de precipitación (nb) y gradiente de caudal se varían de 1 a 6 hrs. y el desfase inicial de precipitación desde 1 a 5 hrs. (nk), de manera de encontrar la mejor configuración para cada modelo paramétrico, en el sentido de encontrar cual es el límite de influencia de cada una de las variables en cuestión.

Al realizar el análisis de los errores se observa que un modelo que involucra la variable gradiente de caudal como parámetro de entrada no es más eficiente que uno que no lo hace, por lo tanto en un modelo paramétrico dicha variable no aporta información adicional al modelo de pronóstico por lo cual resulta innecesaria y no se presenta en este apartado.

En la tabla 4.25 se presentan las configuraciones de los 7 mejores modelos tipo ARMAX y RARMAX respectivamente, obtenidos de acuerdo a una comparación del error cuadrático medio ponderado, cuyos valores se observan en la tabla 4.26.

Los valores de error y coeficiente de determinación asociado a cada conjunto de datos para cada modelo se pueden observar en la tabla 4.27 y 4.28 respectivamente.

Tabla 4.25: Mejores Modelos.

Tabla 4.25.1: Modelos ARMAX

Modelo	na	nb	nk
82	5	5	1
98	7	7	1
77	7	4	1
94	3	7	1
97	6	7	1
95	4	7	1
91	7	6	1

Tabla 4.25.2: Modelos RARMAX

Modelo	na	nb	nk
15	1	3	1
57	1	2	2
22	1	4	1
8	1	2	1
64	1	3	2
50	1	1	2
71	1	4	2

Tabla 4.26: Error Cuadrático Medio Ponderado Modelo ARMAX.

Tabla 4.26.1: Modelos ARMAX

Modelo	R ² ponderado	ECM ponderado [m ³ /s]
82	0,934	1,655
98	0,936	1,658
77	0,936	1,658
94	0,936	1,661
97	0,928	1,662
95	0,937	1,662
91	0,936	1,668

Tabla 4.26.2: Modelos RARMAX

Modelo	R ² ponderado	ECM ponderado [m ³ /s]
15	0,907	1,979
57	0,908	1,989
22	0,904	2,031
8	0,913	2,034
64	0,905	2,038
50	0,913	2,090
71	0,896	2,134

Tabla 4.27: Error cuadrático medio Modelos ARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento[m ³ /s]	ECM Validación [m ³ /s]	ECM Prueba [m ³ /s]
82	0,934	0,940	0,707	2,222	0,555	2,219
98	0,936	0,944	0,715	2,260	0,524	2,153
77	0,936	0,944	0,716	2,262	0,523	2,148
94	0,936	0,943	0,721	2,273	0,527	2,112
97	0,928	0,929	0,681	2,125	0,665	2,414
95	0,937	0,944	0,721	2,280	0,519	2,110
91	0,936	0,944	0,718	2,282	0,525	2,134

Tabla 4.28: Error cuadrático medio Modelos RARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento[m ³ /s]	ECM Validación [m ³ /s]	ECM Prueba [m ³ /s]
15	0,976	0,921	0,739	2,392	1,104	2,610
57	0,976	0,926	0,718	2,426	1,049	2,696
22	0,976	0,921	0,743	2,461	1,042	2,920
8	0,974	0,921	0,740	2,607	0,963	2,480
64	0,975	0,926	0,720	2,495	0,984	2,986
50	0,973	0,925	0,723	2,716	0,928	2,556
71	0,975	0,932	0,719	2,546	1,036	3,430

Se desprende, que un modelo ARMAX presenta mejores pronósticos que un modelo RARMAX, la diferencia se debe principalmente al aumento de los errores de validación que se comenten en el segundo modelo con respecto al primero.

Realizando una comparación con modelos de RNR, se observa que para entrenamientos de este tipo (continuo), un modelo NEWELM 3-7-7-1 presenta un error aprox. un 33% menor que el mejor modelo ARMAX (82) c/r al valor del error de este último y un 44 % menor que un modelo RARMAX.

Posteriormente, al igual que para modelos neuronales, se realiza un análisis unitario, donde cada modelo tiene la mejor configuración obtenida para ARMAX y RARMAX “continuo” y es entrenado de acuerdo a los datos de cada crecida, de tal manera que se forman 21 modelos.

En la tabla 4.29 se presentan los errores ponderados obtenidos para cada modelo formado a partir de los datos cada crecida. Se observa que el modelo RARMAX entrega un desempeño mucho más equilibrado para cada uno de los modelos (crecidas) en comparación a un modelo ARMAX. La razón radica en que el modelo RARMAX ajusta los parámetros del modelo y los corrige punto a punto tanto en entrenamiento, validación y prueba (con la condición inicial de

los parámetros de entrenamiento), de tal manera que para los primeros datos de la crecida, el modelo comienza estimando muy mal y por lo tanto suma un gran error, pero a medida que procesa los datos de la crecida, el ajuste va mejorando. Un modelo ARMAX, en cambio ajusta los parámetros característicos del modelo una sola vez, durante el proceso de entrenamiento, y dicho modelo se hace extensivo para validar y probar.

Por ejemplo, un modelo ARMAX entrenado con los datos de la crecida 10, presenta un error muy alto, al ajustar los parámetros del modelo a esta crecida se pierde la capacidad de generalización, ya que su comportamiento no representa el de las demás crecidas. En cambio un modelo RARMAX presenta un error bastante menor dentro de su estabilidad, ya que una vez ajustados los parámetros durante el entrenamiento de la crecida 10, el modelo toma dichos parámetros como condiciones iniciales y vuelve a ajustarlos durante cada validación.

Observando cada modelo (crecida), en general, se desprende que las crecidas de entrenamiento 4 y 2 definen los mejores modelos ARMAX para realizar pronósticos y que las crecidas 16 y 17 definen los mejores modelos tipo RARMAX.

El hecho de que en el caso de modelos RARMAX las peores crecidas desde el punto de vista de amplitud de frecuencia presenten los menores errores ponderados, se debe igualmente al último punto señalado anteriormente, que indica que el ajuste se realiza punto a punto tanto en entrenamiento como en validación de crecidas. Luego, para una crecida que contenga muchas componentes de frecuencia y sea más variable el ajuste se hace más difícil por medio de un modelo polinómico o paramétrico recursivo, lo cual fija el valor inicial de los parámetros característicos del modelo y por lo tanto es menos extensivo a crecidas más simples, entregando de esta manera un mayor error. En cambio, para modelos ARMAX entrenados con los datos de la crecida con mayor amplitud de frecuencia los parámetros se ajustan una sola vez en función del comportamiento general de la crecida, permitiendo una mayor capacidad de generalización.

Las diferencias porcentuales de error indican que un modelo de RNR con un entrenamiento unitario es aprox. un 80 % mejor que un modelo ARMAX en cuanto al valor del ECM y por lo tanto significativamente mejor que un modelo RARMAX, (valor comparado sobre la base del modelo que arroja menores errores según la crecida utilizada en el entrenamiento).

Tabla 4.29: Errores Ponderados Modelo ARMAX y RARMAX.

Crecida	Error Ponderado [m ³ /s] ²	
	ARMAX	RARMAX
1	2,176	2,279
2	1,524	2,465
3	1,699	2,265
4	1,508	2,280
5	1,875	2,268
6	2,345	2,251
7	1,937	2,262
8	1,655	2,250
9	2,072	2,249
10	13,949	2,309
11	7,702	2,328
12	2,310	2,254
13	2,178	2,248
14	2,091	2,248
15	1,835	2,250
16	2,374	2,247
17	2,101	2,248
18	5,709	2,249
21	2,098	2,524

Observando los mejores modelos ARMAX y RARMAX desde el punto de vista de los errores cuadráticos medios ponderados, se desprende que un modelo ARMAX entrenado con los datos de la crecida que presenta mayor amplitud de frecuencia puede ser un mejor modelo de pronóstico en relación a un modelo del mismo tipo entrenado con una serie de crecidas “pegadas”. En cambio, un modelo RARMAX, constituye un mejor modelo de pronóstico cuando es entrenado a nivel continuo, ello se debe a que posee una mayor cantidad y variabilidad de datos, y dado que ajusta los parámetros punto a punto, de esta manera posee un mayor base para pronosticar.

Hay que tener en cuenta que un modelo de redes neuronales recurrentes es un modelo mucho más complejo de modelar que evidentemente presenta mejores resultados, tanto con entrenamientos a nivel continuo y sobre todo a nivel unitario, donde los errores son de 1 a 2 ordenes de magnitud menor. Por otra parte, es un modelo que demora mucho más en entregar resultados y más complejo de programar, razón por la cual, no se debe desmerecer la capacidad de modelos de tipo paramétrico para realizar pronósticos, primeramente en esta cuenca.

4.3.1.5 Pronóstico mediante Red Neuronal Recurrente de Hopfield-Identificación.

Dada la teoría expuesta en el capítulo 2, la red de Hopfield es una red de tipo recurrente que se utiliza para el reconocimiento de patrones y en la cual para garantizar la convergencia de las variables de entrada a los patrones almacenados se requiere que estos últimos estén codificados con 1s y -1s, lo que se convierte en una de las principales dificultades de utilizar esta red.

Se consideran 160 patrones constituidos por la información de la crecida 4, crecida que posee la mayor capacidad de generalización y que forma el mismo conjunto que se utiliza para realizar un análisis considerando un entrenamiento unitario con resultados eficientes en una red de Elman.

Cada patrón es formado de tal manera de contener la información relevante para pronosticar el caudal a 1 paso de adelanto y con el objetivo de representar a un determinado “punto” (tiempo, caudal (t)), aunque el modelo en sí no es directamente de pronóstico de caudal, según se explica más adelante.

Las variables contenidas en cada uno de los patrones son: $P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y \dot{Q}_{t+1} . Cabe recordar que la red es de interconexión total entre neuronas y que cada componente del patrón define una neurona con diferentes estados que caracterizan las trayectorias posibles a dicho patrón.

Para garantizar el correcto funcionamiento de la red cada uno de los patrones es codificado binariamente, dicho de otro modo, un determinado valor se representa por medio de 0s y 1s que multiplican potencias de 2. Se codifica el valor entero de cada variable de cada patrón y vector de entrada, se incluyen 2 decimales representativos bajo la misma codificación y un “bit” de signo, 0 en caso de que el valor sea negativo y 1 en el caso de que el valor sea positivo. Posteriormente los 0s son reemplazados por -1, quedando todos los valores representados por 1s y -1s.

La cantidad de componentes del vector que representa cada número no es igual y por lo tanto se realiza un relleno de aquellos vectores con menos componentes con valores nulos (representados por -1), para ello se considera como número de componentes de dichos vectores la cantidad que se tiene para el valor máximo de cada variable. Todas las componentes, las del vector que representa el valor entero, los decimales y el *bit* de signo se agrupan para constituir un solo vector que representa sólo un número, esta agrupación se realiza para cada dato de cada variable, las cuales posteriormente también se agrupan y constituyen un solo vector característico ($P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y \dot{Q}_{t+1}). Luego, lo que se conoce como vector patrón de la red caracterizado por 5 variables, para un modelo de Hopfield de este tipo, queda definido por un vector que posee una gran cantidad de componentes, 21 en total, por lo tanto la red está formada por 21 neuronas y una totalidad de 210 ($21 * 20 / 2$) interconexiones entre ellas.

La red reconoce dichos patrones, y los almacena creando un hipercubo con diferentes zonas de atracción. Un vector de entrada, de acuerdo al valor de sus componentes “cae” en alguna de las zonas y sigue la trayectoria hacia un patrón, la convergencia hacia este último queda garantizada de acuerdo al número de iteraciones que la red realice (equivalente a las épocas de

entrenamiento de una red supervisada), no existe de antemano un número de ellas necesarias para la convergencia, de tal manera que si el número de épocas elegidas es insuficiente, el vector de entrada no alcanza a completar su trayectoria hacia el patrón almacenado y la respuesta de la red ante dicha entrada es un estado espúreo o memoria falsa que se encuentra dentro de la zona de atracción de la red y que como vector no posee 1s y -1s, sino que es un estado intermedio con valores decimales en dicho rango, por lo tanto imposible de decodificar.

Una vez creada la red, en base a pruebas en que se varía el número de iteraciones necesarias, se decide que una cantidad de 10 es suficiente para alcanzar la convergencia a los patrones.

La información de las crecidas es ingresada de tal manera de rescatar el gradiente posterior (\dot{Q}_{t+1}). Se ingresa el vector de entrada con las variables descritas: $P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y no el gradiente con respecto a $t+1$, en reemplazo de esta última variable se repite la tendencia precedente (\dot{Q}_t), de tal manera que la red realiza una lectura de todas las variables que componen este vector y busca el patrón que se le parezca más (identificación), entregando por lo tanto \dot{Q}_{t+1} , que es lo que se quiere rescatar de la información contenida en los patrones. Durante el proceso la red trata de identificar a que zona de atracción pertenece el vector de entrada, encontrando distintas respuestas dadas por los estados que se encuentran en su trayectoria hacia la convergencia del patrón almacenado, entonces, cuando la alcanza devuelve el gradiente en el instante de pronóstico. Conocido el gradiente en dicho instante y el caudal antecesor, se estima el caudal a la siguiente hora (pronosticado). Por lo tanto, el modelo de pronóstico está basado en dos etapas, una de identificación del gradiente con respecto al tiempo de pronóstico y otra de estimación del caudal en base a una sencilla ecuación ($Q_{t+1} = \dot{Q}_{t+1} + Q_t$).

En síntesis, la red crea una base de datos de condiciones iniciales que permiten tener un espectro de decisión amplio de comportamientos de salida. El patrón podría contener más información antecesora, por ejemplo, el caudal hasta con 3 hrs. de retardo (tiempo de concentración) a partir de t o $t+1$, pero esto añade una complicación en la codificación de cada vector, debido a la cantidad de componentes necesarias para representarlos.

En la tabla 4.30 se presentan los errores estándar y cuadráticos medios para cada crecida validada, así como también los índices de eficiencia de pronóstico de caudal máximo. El error estándar es entendido como un promedio de la diferencia entre los valores simulados y los observados. Se observa que los errores son del mismo orden que los obtenidos para modelos neuronales de pronóstico tanto prealimentados como recurrentes de Elman, aunque mayores que los obtenidos para modelos unitarios y en términos generales bien similares a los obtenidos para un modelo paramétrico tipo ARMAX y pero inferiores a los de un modelo RARMAX.

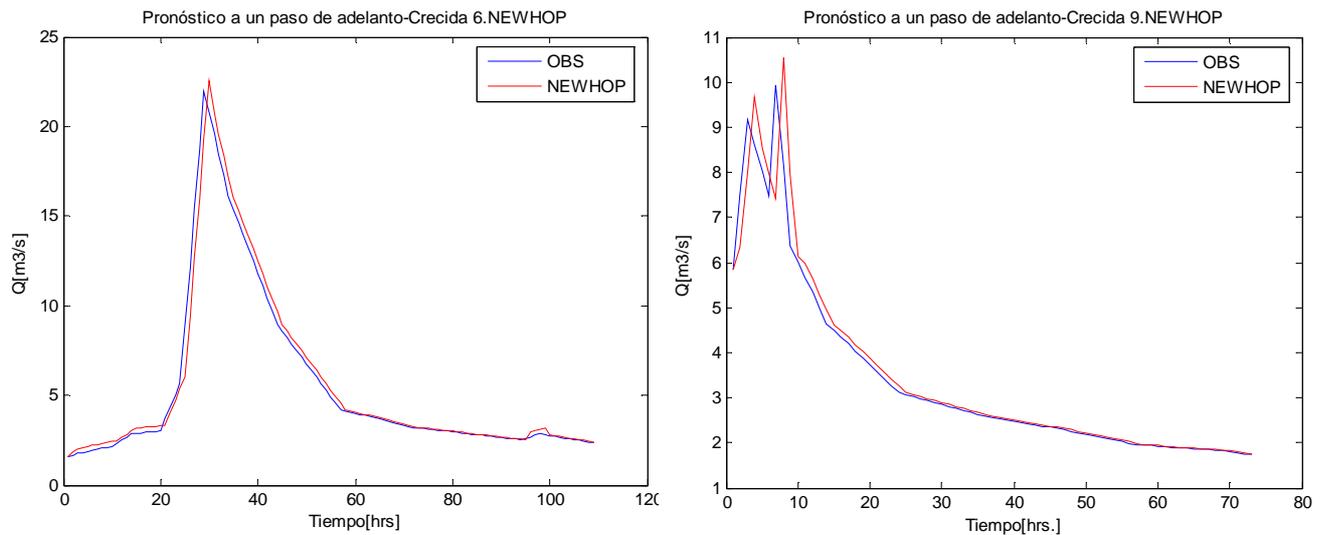
Se observa que en general existe sobre-estimación del caudal máximo, pero, en el momento de ocurrencia real de dicho valor se presenta más bien una subestimación con respecto al valor observado. El hecho de que existan mayormente sobre-estimaciones de los valores máximos de caudal se debe a que el modelo de pronóstico trabaja en base a la identificación del gradiente en el instante de pronóstico conocida la tendencia precedente. Luego, cuando se está en la primera parte de la curva, la tendencia es ascendente y por lo tanto el modelo de pronóstico se basa en ello, con notoriedad cuando llega a la zona de inflexión, donde es importante la diferencia entre el gradiente antecesor al tiempo de pronóstico y el valor nulo de este que se produce en el peak, hasta que encuentra un valor negativo y bruscamente pasa a la zona de receso.

Tabla 4.30: Errores estándar de validación. Modelo de Hopfield.

Crecida	Duración	HOPFIELD				
		E. Estandar [m ³ /s]	ECM [m ³ /s] ²	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]
1	43	0,097	0,130	0,096	-0,151	1
2	125	0,087	6,663	0,011	-0,228	1
3	62	0,043	0,415	0,026	-0,031	1
4	161	0,115	4,950	-0,002	-0,024	1
5	174	0,109	0,841	-0,014	-0,076	1
6	112	0,091	0,500	0,024	-0,125	1
7	84	0,100	0,816	0,050	-0,350	1
8	103	0,098	0,081	0,060	-0,023	1
9	73	0,073	0,267	0,062	-0,253	1
10	22	0,096	2,941	0,055	-0,722	1
11	24	-0,089	9,970	0,021	-0,275	1
12	51	0,078	1,041	0,023	-0,103	1
13	210	0,101	0,030	0,057	-0,029	1
14	114	0,092	0,031	0,108	-0,064	1
15	44	0,120	0,091	0,055	0,009	1
16	56	0,134	0,047	0,096	0,004	1
17	57	0,164	0,053	0,086	0,067	1
18	24	0,222	0,082	0,093	0,092	1
19	14	0,175	7,303	0,027	0,022	1
20	13	0,198	8,036	-0,025	0,041	1
21	21	0,223	4,977	0,030	-0,306	1
Ponderado		0,102	1,670	0,041	-0,100	1

En la figura 4.20 se presenta, a modo de ejemplo, el caudal observado y simulado para la crecida 6 y 9. Al igual que se aprecia en la tabla 4.30 existe un desfase entre máximos de 1 hora lo que no lo hace eficiente al nivel de redes de pronóstico con entrenamientos de tipo unitario, pero si se observan las diferencias entre los valores de los máximos observados y pronosticados a su tiempo de ocurrencia las diferencias son en promedio de un 4 % y para el tiempo de ocurrencia real alcanzan el 10 %, valores que se encuentra al nivel de redes neuronales recurrentes de Elman entrenadas continuamente.

Figura 4.20: Pronósticos a un paso de adelanto. Modelo de Hopfield.



4.3.1.6 Resumen y análisis de modelos de pronóstico a un paso de adelanto

Esta sección tiene por objetivo, situar al lector en los desarrollos de pronósticos realizados hasta el momento, a fin de continuar con pronósticos a varios pasos de adelanto.

Se realizan análisis entre los mejores modelos de pronóstico de cada tipo analizado. Para comparar modelos neuronales entrenados unitariamente se considera como crecida de entrenamiento la crecida 4 (recordar, crecida con mayor amplitud frecuencial y demostrada su mayor capacidad de generalización). Se comparan modelos de redes neuronales recurrentes de Elman y de redes prealimentadas a nivel de entrenamiento unitario y continuo, como también el modelo neuronal de Hopfield y modelos paramétricos, ARMAX y RARMAX [na, nb, nk].

Para estos últimos modelos se incluyen sólo el modelo ARMAX unitario y el modelo RARMAX continuo, debido a las razones indicadas precedentemente (punto 4.3.1.4) que los señalan como los mejores modelos de pronóstico de este tipo.

Los parámetros de comparación utilizados son el ECM ponderado, el ECM de validación por crecida y los índices de error en el pronóstico de máximos. Los cuales se presentan en la tabla 4.31.a ,4.31.b y 4.31.c respectivamente.

En la tabla 4.31.a se observan las ventajas a nivel de errores ponderados de realizar entrenamientos de tipo unitario en modelos neuronales. Para este tipo de entrenamiento y además, para el de tipo continuo el menor error se obtiene para un modelo Newelm 3-7-7-1.

En general, los modelos paramétricos presentan errores superiores a los de redes neuronales de pronóstico y sin una clara tendencia “común” a entregar menores errores con entrenamientos de

tipo continuo o unitario. Cabe recordar lo expuesto en el punto 4.3.1.4, en que se destacan las razones de mejores pronósticos a nivel continuo para modelos RARMAX y a nivel unitario para modelos ARMAX, este último no es netamente concluyente, dado que la diferencia de error entre ambos tipos de entrenamiento es pequeña. Además, en cuanto al modelo de Hopfield se observa que entrega un error ponderado al nivel de un modelo ARMAX.

Tabla 4.31.a: Errores ponderados según modelo de pronóstico.

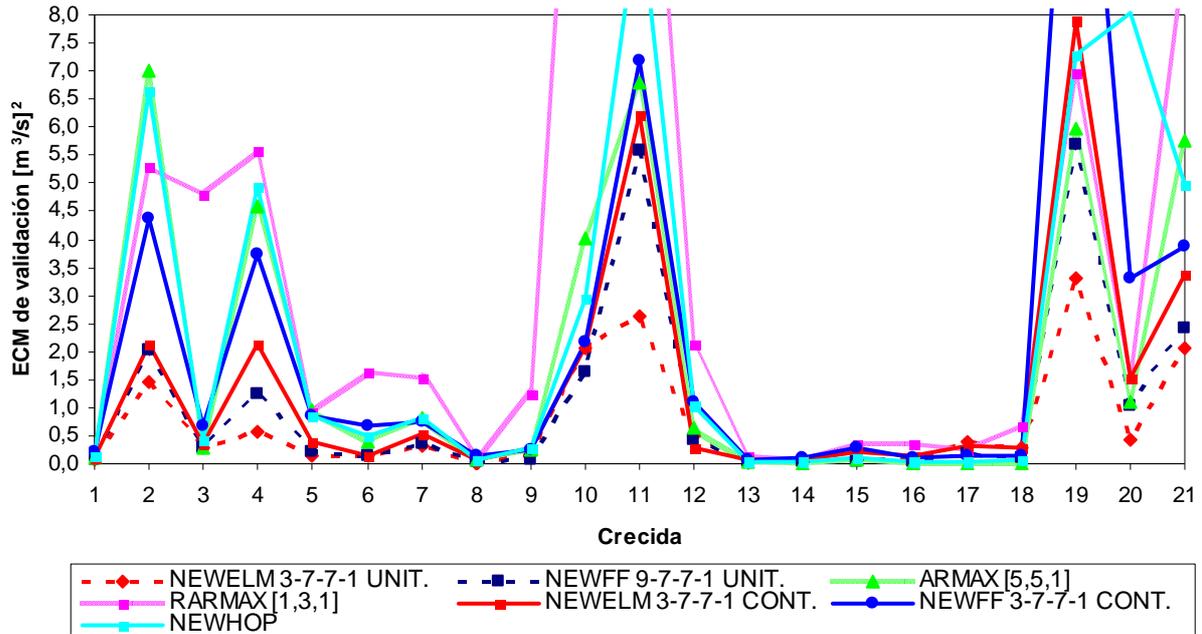
ECM PONDERADO [m ³ /s] ²	Entrenamiento	
	Unitario	Continuo
NEWELM 3-7-7-1	0,415	1,106
NEWFF 9-7-7-1	0,587	1,836
ARMAX	1,506	1,655
RARMAX	2,247	1,979
HOPFIELD	1,670	

En la tabla 4.31.b se presentan los errores de validación obtenidos para cada crecida, los cuales se pueden apreciar gráficamente en la figura 4.21. Se deduce que los modelos neuronales de pronóstico, no de identificación (Hopfield), presentan las mejores validaciones y por lo tanto mejores pronósticos, destacándose las claras ventajas también a este nivel de un modelo recurrente NEWELM 3-7-7-1 sobre cualquiera de los otros modelos planteados.

Tabla 4.31.b: Errores de validación por crecida según modelo de pronóstico.

Crecida	ECM de validación [m ³ /s] ²						
	Unitario		Continuo		ARMAX [5,5,1]	RARMAX[1,3,1]	HOPFIELD
	NEWELM 3-7-7-1	NEWFF 9-7-7-1	NEWELM 3-7-7-1	NEWFF 9-7-7-1			
1	0,075	0,088	0,108	0,204	0,099	0,128	0,130
2	1,469	2,043	2,120	4,387	7,004	5,291	6,663
3	0,271	0,278	0,370	0,683	0,290	4,801	0,415
4	0,573	1,252	2,134	3,723	4,572	5,593	4,950
5	0,150	0,207	0,406	0,839	0,947	0,929	0,841
6	0,140	0,142	0,130	0,682	0,390	1,652	0,500
7	0,325	0,355	0,534	0,764	0,833	1,527	0,816
8	0,016	0,036	0,080	0,146	0,066	0,145	0,081
9	0,172	0,086	0,294	0,257	0,239	1,235	0,267
10	2,049	1,645	2,106	2,179	4,033	16,039	2,941
11	2,626	5,577	6,214	7,197	6,786	14,908	9,970
12	0,485	0,439	0,286	1,117	0,649	2,149	1,041
13	0,063	0,034	0,057	0,086	0,019	0,128	0,030
14	0,022	0,020	0,085	0,109	0,016	0,082	0,031
15	0,053	0,051	0,218	0,298	0,063	0,370	0,091
16	0,073	0,023	0,136	0,099	0,017	0,360	0,047
17	0,399	0,164	0,305	0,150	0,003	0,234	0,053
18	0,277	0,116	0,279	0,145	0,007	0,662	0,082
19	3,316	5,694	7,891	14,500	5,983	6,980	7,303
20	0,413	1,032	1,532	3,306	1,093	1,530	8,036
21	2,045	2,415	3,366	3,864	5,748	9,041	4,977

Figura 4.21: Análisis gráfico de validación de crecidas, según modelo de pronóstico.



En la tabla 4.31.c, correspondiente al valor de los índices de error en el pronóstico de caudales máximos para modelos neuronales recurrentes de Elman, prealimentados y recurrentes de Hopfield, se destacan también, las ventajas de un modelo NEWELM 3-7-7-1. Cabe destacar que el índice TMEM ha sido calculado como un promedio del error obtenido en el pronóstico de máximos para cada crecida.

Tabla 4.31.c: Índice de error de pronóstico de máximos ponderado según modelo de pronóstico.

ÍNDICES DE ERROR [tanto por 1]	Entrenamiento					
	Unitario			Continuo		
	EMEM	EMRMOBS	TMEM	EMEM	EMRMOBS	TMEM
NEWELM 3-7-7-1	-0,023	-0,060	0,143	-0,086	-0,179	0,253
NEWFF 9-7-7-1	-0,087	-0,034	-0,182	-0,132	-0,213	0,284
NEWHOP	0,041	-0,100	1			

Dados los análisis anteriores, se deduce que el mejor modelo de pronóstico tanto en calidad como en valor es un modelo NEWELM 3-7-7-1 entrenado con los datos de la crecida número 4.

Las razones expuestas en el punto 4.3.1.5 en conjunto con el análisis de errores de validación gráfico presentado precedentemente, indican que realizar pronósticos a más pasos de adelanto con un modelo neuronal de identificación de Hopfield resulta factible e interesante desde el punto de vista de la diferencia de operación del modelo con respecto a los demás modelos neuronales. En el siguiente punto, dedicado a pronóstico a más hrs. de adelanto se detallan algunas de las maneras en que se podrían realizar pronósticos con dicha red neuronal. Además, en el capítulo 6 se recomiendan algunas alternativas para el uso de la red de Hopfield que

resultarían ser bastante útiles en el pronóstico de crecidas, tanto como modelo de pronóstico como también considerado como un proceso anterior e incorporado en la creación de un modelo de pronóstico con redes recurrentes de Elman.

En la figura 4.22 se observa el pronóstico realizado mediante el modelo neuronal recurrente de Elman en comparación a modelos de tipo paramétrico ARMAX y RARMAX, en un sector representativo del comportamiento de los pronósticos en el conjunto de validación, para entrenamientos de tipo continuo.

Se desprende que en general un modelo de tipo paramétrico presenta retardos en los pronósticos de caudal. Además, un modelo RARMAX la mayor parte del tiempo sobre-estima el caudal máximo, lo que no siempre sucede con un modelo ARMAX.

En modelos de RN de pronóstico, por ejemplo figura 4.23, existe una tendencia a subestimar el valor de los caudales máximos tanto en el tiempo de ocurrencia real como en el individual, en cambio la red de identificación de Hopfield sobre-estima los caudales máximos, pero en el momento de ocurrencia real tiende a subestimar.

Por otra parte, los desfases entre el valor pronosticado y observado, en cuanto al tiempo en que se producen son casi nulos y suficientemente menores para modelos de RNR de Elman en comparación a los que se obtienen para un modelo de redes neuronales prealimentado, y destacadamente sobre modelos de tipo paramétrico, Ver figura 4.22 y 4.23.

Figura 4.22: Análisis de Modelos en el conjunto de validación.

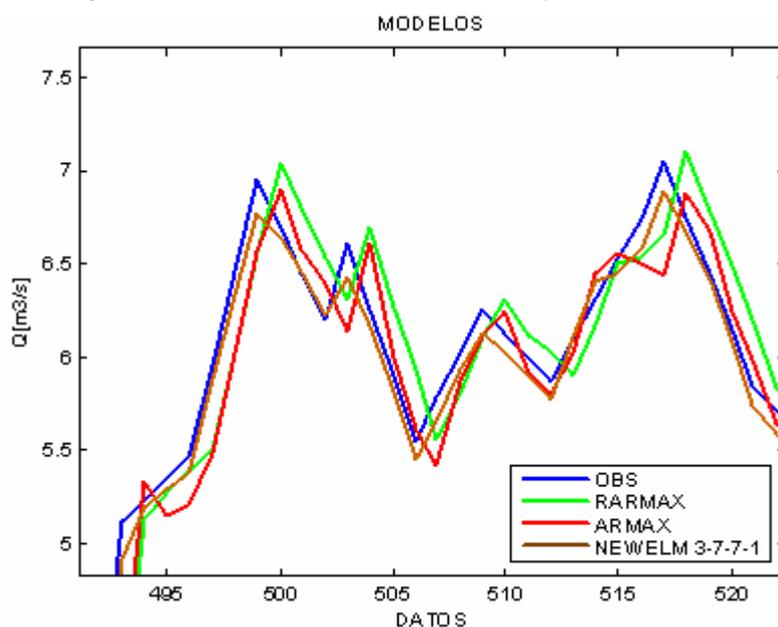
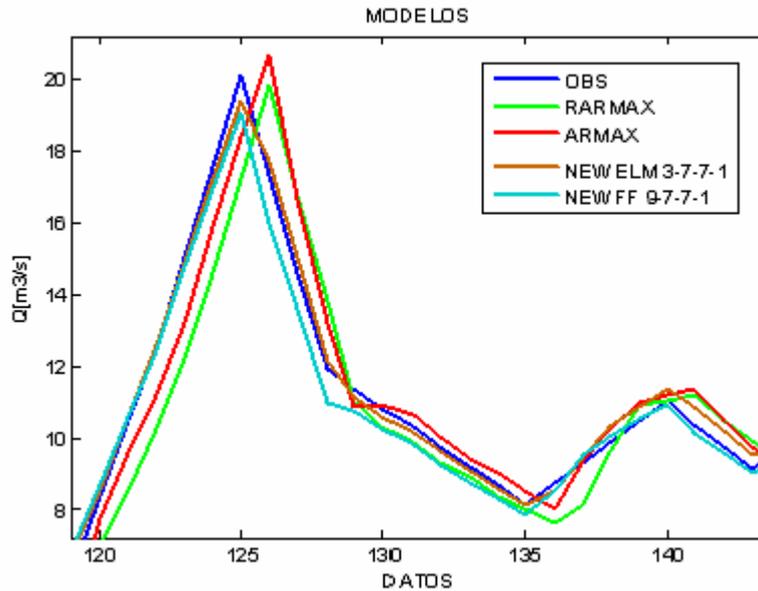


Figura 4.23: Análisis de Modelos en el conjunto de validación.



4.3.1.7 Pronósticos a varios pasos de adelanto

Se pretende realizar pronósticos a varios pasos de adelanto esencialmente con el modelo neuronal recurrente de Elman. Éste es comparado en términos de eficiencia con modelos prealimentados y modelos paramétricos, para finalizar con un pronóstico mediante la red neuronal de identificación recurrente de Hopfield y su comparación con el modelo de Elman.

Se realizan pronósticos de caudales a 4, 8, 12, 16 y 20 horas a partir de dos tipos de modelos, uno que toma como conjunto de datos de entrenamiento sólo la crecida número 4, dado que posee mayores amplitudes de frecuencia y define los mejores modelos a nivel de entrenamiento unitario y otro que considera como dicho conjunto, aquel empleado en el análisis tipo continuo.

Las variables de entrada son aquellas que definieron los mejores modelos de pronóstico, es decir el volumen de precipitación (t), el caudal (t) y el gradiente de este último con respecto a t para pronósticos a $t+x$ hrs.

Las arquitecturas neuronales óptimas para cada modelo de pronóstico, dependiendo de los pasos de adelanto se encuentran por medio de la misma metodología empleada anteriormente. Para modelos Newelm y Newff entrenados unitariamente y continuamente, la cantidad de neuronas óptima en las capas ocultas para cada paso de adelanto fue la misma que para un paso de adelanto, 7 neuronas.

En la tabla 4.32 se presenta los errores ponderados obtenidos de dichos pronósticos para modelos neuronales de Elman y prealimentados considerando un análisis continuo. Se desprende que si bien los errores ponderados para una red de tipo recurrente no son significativamente

menores a los de una red prealimentada, entrega consistentemente menores errores para todos los tiempos de adelanto y por lo tanto es un mejor modelo de pronóstico.

En la tabla 4.33 se presentan los errores de entrenamiento y ponderados obtenidos de realizar pronósticos de acuerdo al entrenamiento unitario con los datos de la crecida 4. Claramente los errores de entrenamiento son menores para una red de tipo recurrente, lo que es producto del beneficio que otorga gracias al retardo de la información en el fenómeno individual, lo mismo ocurre en los errores ponderados que reflejan los errores de validación obtenidos.

Comparando los errores ponderados obtenidos de análisis continuos y unitarios, se deduce que para obtener los mejores modelos de pronóstico no es necesario utilizar un conjunto que contenga un gran número de crecidas para el entrenamiento, sino más bien una crecida que contenga amplias componentes de distintas frecuencias, como es el caso de la crecida 4. Esto lleva a resultados más eficientes tanto en términos de pronóstico, como en cuanto a la facilidad de entrenamiento, dado que se necesita una menor cantidad de tiempo para entrenar la red neuronal, lo cual es muy útil en el caso de redes neuronales recurrentes.

Para la comparación de modelos neuronales con modelos de tipo paramétrico, se utiliza el mejor modelo de estos últimos, es decir, un modelo ARMAX de configuración [5, 5, 1] ([na, nb, nk]) entrenado con los datos de la crecida 4, cuyos valores de error ponderado para cada tiempo de pronóstico se presentan en la tabla 4.34. Se desprende que los errores por medio de estos modelos son mayores que los entregados por un modelo neuronal tanto recurrente como prealimentado.

Tabla 4.32: Errores Ponderados Modelos NEWELM y NEWFF. Análisis Continuo.

Modelo (Continuo)	Error Ponderado [m ³ /s] ²				
	t+4	t+8	t+12	t+16	t+20
Newelm 3-7-7-1	4,2	15,7	25,3	31,2	39,8
Newff 9-7-7-1	4,5	16,7	26,0	31,4	42,2

Tabla 4.33: Errores de entrenamiento y Ponderados Modelos NEWELM y NEWFF. Análisis Unitario.

Pronóstico	Error de Entrenamiento [m ³ /s] ²		Error Ponderado [m ³ /s] ²	
	Newelm 3-7-7-1	Newff 9-7-7-1	Newelm 3-7-7-1	Newff 9-7-7-1
t+4	12,0	13,0	3,9	4,7
t+8	47,6	60,4	12,6	13,3
t+12	70,4	91,4	21,0	23,1
t+16	72,1	104,8	26,2	31,6
t+20	98,0	103,9	32,9	37,0

Tabla 4.34: Errores Ponderados Modelo ARMAX [5, 5, 1]. Análisis Unitario.

modelo ARMAX (unitario)	Error Ponderado [m ³ /s] ²				
	t+4	t+8	t+12	t+16	t+20
	8,900	15,990	26,400	32,870	43,160

En la tabla 4.35 se presentan los caudales máximos observados y pronosticados para cada crecida luego de realizar un entrenamiento continuo de la red NEWELM 3-7-7-1, la correlación existente entre ambos valores se aprecia en la figura 4.24.

En la tabla 4.36 se presentan los caudales máximos observados y pronosticados para cada crecida para un modelo NEWELM 3-7-7-1 entrenado con los datos de la crecida nº 4, la correlación entre ambos valores se aprecia en a figura 4.25.

Comparando las correlaciones de máximos observados y pronosticados de la figura 4.24 y 4.25, un modelo NEWELM 3-7-7-1 con un entrenamiento unitario (crecida 4) entrega mejores pronósticos, ya que existe menos dispersión de los datos y la pendiente es más cercana.

En la tabla 4.37.a se presentan los índices de error en el pronóstico del caudal máximo, valores ponderados según el número de crecidas y pronósticos, para los distintos tipos de entrenamiento en modelos NEWELM. Se deduce que existen subestimaciones de caudal de hasta un 5 % y un retardo que en promedio es de 3 hrs. para obtener pronósticos de hasta 20 hrs. mediante un análisis unitario. Igualmente que en el caso de los errores ponderados, los índices más bajos se obtienen para un modelo NEWELM 3-7-7-1 entrenado con la crecida 4.

Tabla 4.35: Caudal máximo Observado y Pronosticado. Modelo NEWELM - Análisis Continuo.

Crecida	Q máx Observado [m ³ /s]	Q máx simulado [m ³ /s]				
		t+4	t+8	t+12	t+16	t+20
1	4,9	4,0	3,5	3,2	3,1	1,4
2	44,8	39,5	34,8	31,8	30,0	26,3
3	13,1	11,2	9,7	8,2	5,6	4,6
4	64,2	58,2	51,4	48,3	44,6	38,4
5	27,0	24,5	23,1	20,6	18,3	17,4
6	22,0	21,3	19,5	18,1	16,8	14,7
7	10,9	8,7	7,9	7,3	6,8	6,8
8	8,1	7,7	6,8	6,4	5,9	5,8
9	9,9	8,6	6,1	3,8	2,8	2,4
10	9,9	6,4	5,5			
11	27,5	22,3	19,1			
12	20,1	19,6	17,2	12,1		
13	8,9	9,1	8,2	7,8	7,2	6,9
14	4,9	4,7	4,1	4,0	3,7	3,7
15	7,1	6,6	6,1			
16	6,2	6,3	5,8	5,4		
17	4,8	5,1	4,6	4,4		
18	4,5	4,6				
19	19,8	17,0				
20	13,6	12,5				
21	15,6	12,6				

Tabla 4.36: Caudal máximo Observado y Pronosticado. Modelo NEWELM - Análisis unitario.

Crecida	Q máx Observado [m ³ /s]	Q máx simulado [m ³ /s]				
		t+4	t+8	t+12	t+16	t+20
1	4,9	4,3	4,2	3,6	4,1	1,3
2	44,8	41,9	39,8	40,4	36,3	36,0
3	13,1	11,9	11,1	6,7	6,5	6,1
4	64,2	62,9	59,3	54,5	44,7	52,2
5	27,0	26,3	24,4	26,8	20,3	25,5
6	22,0	22,1	20,6	25,6	20,0	15,7
7	10,9	9,1	8,7	9,0	7,2	8,7
8	8,1	7,8	7,1	7,6	6,3	8,1
9	9,9	9,1	7,1	2,9	3,0	3,4
10	9,9	6,9	6,8			
11	27,5	24,3	24,4			
12	20,1	20,5	19,3	9,5		
13	8,9	9,0	8,4	9,2	6,7	9,6
14	4,9	4,7	4,4	4,5	3,8	4,7
15	7,1	6,6	6,1			
16	6,2	6,4	6,0	6,4		
17	4,8	5,0	4,6	5,2		
18	4,5	4,6				
19	19,8	18,6				
20	13,6	13,2				
21	15,6	13,4				

Figura 4.24: Pronóstico de caudal máximo a t+x horas. Modelo NEWELM 3-7-7-1-Análisis continuo.

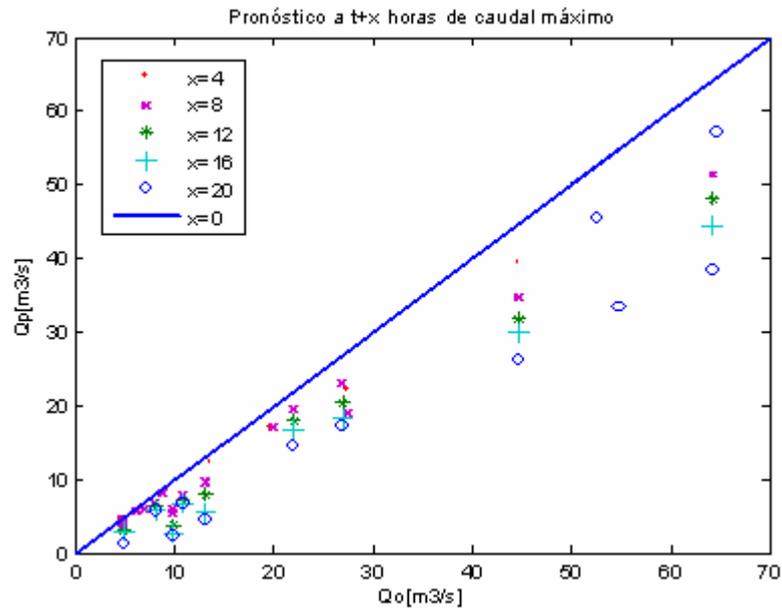


Figura 4.25: Pronóstico de caudal máximo a t+x horas. Modelo NEWELM 3-7-7-1-Análisis unitario.

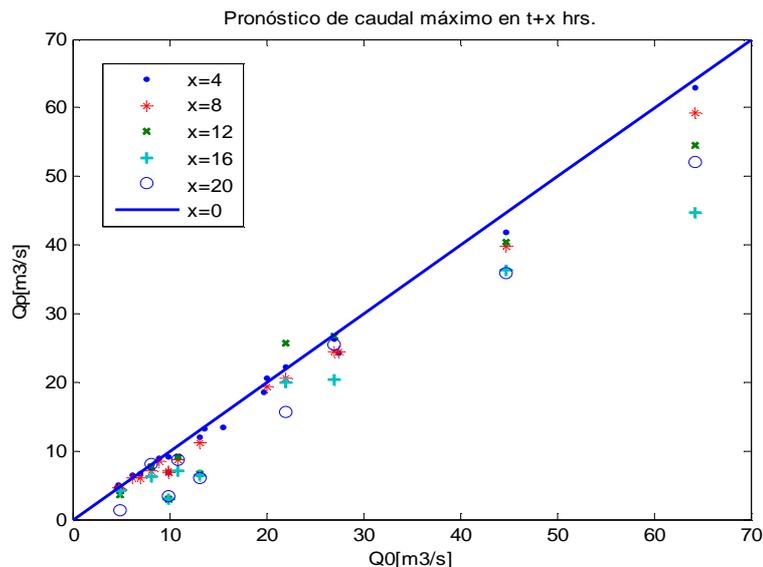


Tabla 4.37.a: Índices de Pronóstico Ponderados de caudal máximo. Modelos NEWELM 3-7-7-1.

Análisis Continuo		Análisis unitario	
EMEM [tanto por 1]	-0,105	EMEM [tanto por 1]	-0,049
EMRMOBS [tanto por 1]	0,260	EMRMOBS [tanto por 1]	-0,170
TMEM [hr.]	3,039	TMEM [hr.]	2,799

A continuación se realizan pronósticos utilizando la red neuronal de Hopfield. Los patrones que se deben almacenar son construidos al igual que para el modelo a un paso de adelanto con la información entregada por la crecida número 4, para ello las variables descritas en el punto 4.3.1.5, $P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y \dot{Q}_{t+1} , deben cambiar de tal manera de permitir la identificación del gradiente a t+x pasos de adelanto con $x \in \{4,8,12,16,20\}$. Por lo tanto, se elimina la variable \dot{Q}_{t-1} , y en vez de \dot{Q}_{t+1} es necesaria \dot{Q}_{t+x} , entonces, el patrón almacenado queda compuesto por las siguientes variables: $P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y \dot{Q}_{t+x} .

Nuevamente, la información de los patrones y vectores de entrada es codificada mediante la asignación de números binarios. Los vectores de entrada de cada crecida poseen las mismas variables que caracterizan a los patrones almacenados de tal manera de identificarlas, exceptuando la última variable que es la que se pretende rescatar (reemplazada por \dot{Q}_t). Posteriormente, en base a \dot{Q}_{t+x} y Q_t se deduce Q_{t+x} , recordando que modelo de pronóstico usado para la red de Hopfield posee dos etapas, un proceso de identificación y luego uno de estimación mediante: $(\dot{Q}_{t+x} + Q_t) * x = Q_{t+x}$.

En la tabla 4.37.b se presentan los errores de pronósticos de caudal máximo a 16 hrs. de adelanto, los cuales han sido cuantificados para cada crecida y posteriormente ponderados.

También, se presenta el error promedio cometido considerando todos los pronósticos realizados (a 4, 8, 12, 16 y 20hrs.), exceptuando el valor del ECM, valor que está indicado a las 20 hrs.

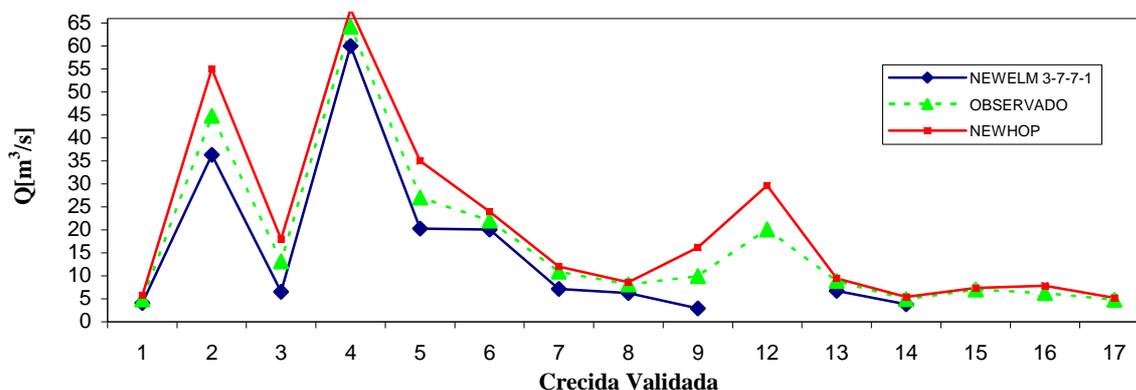
Se observa que el error de pronóstico se encuentra al nivel de un modelo ARMAX. En general, se producen sobre-estimaciones del caudal máximo, aunque para el tiempo de ocurrencia real de dicho valor se produce una subestimación, las razones de este comportamiento son las que se indican en el punto 4.3.1.5 y que tienen relación con la estimación del caudal máximo. La gran desventaja del método en relación a modelos neuronales recurrentes de Elman radica en el valor de los índices EMRMOBS y TMEM, lo que se debe a la segunda etapa del modelo de pronóstico (estimación), la cual como se ha indicado está basada sólo en la derivada, que si bien otorga una tendencia y por lo tanto información de pronóstico, el hacerla valedera a varios pasos de adelanto resulta difícil más que nada en términos de la hora en que se producen dichos valores.

En la figura 4.26 se observa el caudal máximo pronosticado para cada crecida, por cada una de las redes recurrentes estudiadas, la red de Elman (con entrenamiento unitario) y la red de Hopfield, los valores obtenidos para cada crecida son comparados con los observados. Se observa claramente la tendencia de la red de Elman a subestimar en comparación a la de Hopfield a sobre-estimar. La gran ventaja que presenta el modelo de Hopfield radica en que no necesita un gran set de datos de validación para pronosticar a muchos pasos de adelanto, no así la red de Elman o prealimentada, que sí requieren suficientes datos para obtener un pronóstico certero. Entonces, no es posible pronosticar las crecidas 10, 11, 12, 15, 16, 17, 19, 20 y 21 por medio de un modelo Elman.

Tabla 4.37.b: Índices de Pronóstico Ponderados de caudal máximo. Modelo NEWHOP.

ERROR a 16 hrs.		ERROR PROMEDIO.	
ECM[m ³ /s] ²	39,431	ECM[m ³ /s] ²	43,387
EMEM [tanto por 1]	0,108	EMEM [tanto por 1]	0,121
EMRMOBS [tanto por 1]	-0,449	EMRMOBS [tanto por 1]	-0,387
TMEM [hr.]	13,129	TMEM [hr.]	9,120

Figura 4.26: Caudal máximo a t+16 horas. Modelos Recurrentes.



4.3.2 Cuenca Mapocho en Los Almendros

Se cuenta con un total de 27 crecidas, 1626 datos cuyas características se presentan en la tabla 4.38. Dado el rápido tiempo de respuesta de la cuenca (aprox. 3 hrs.), se considera adecuado utilizar modelos de dos capas ocultas.

Tabla 4.38: Caracterización de Crecidas.

CRECIDA	INICIO	TERMINO	DURACIÓN [HR]	MEDIA [m ³ /s]	DESV EST. [m ³ /s]	Q Máx [m ³ /s]	QMin [m ³ /s]	Pp [mm]	Tª[°C]
1	24-8-79 9:00	28-8-79 2:00	89	7,1	4,1	27,2	1,8	66,2	10,3
2	30-8-79 19:00	31-8-79 22:00	27	30,6	17,1	74,7	10,9	29,4	8,4
3	10-4-80 2:00	10-4-80 23:00	21	46,4	34,5	120,0	7,8	34,1	12,5
4	2-5-81 14:00	4-5-81 14:00	48	6,8	4,0	18,7	3,3	20,5	12,5
5	16-7-82 1:00	17-7-82 3:00	26	51,3	19,0	90,0	20,6	23,2	10,7
6	23-8-82 13:00	29-8-82 11:00	142	20,6	4,3	32,7	15,8	23,4	11,2
7	4-7-84 2:00	5-7-84 13:00	35	10,2	5,9	25,4	3,0	79,8	7,5
8	27-5-86 5:00	28-5-86 22:00	41	8,8	6,3	30,5	2,3	91,3	9,2
9	9-6-86 20:00	10-6-86 16:00	20	13,9	2,5	19,4	8,9	16,7	9,9
10	13-6-86 3:00	14-6-86 1:00	22	16,5	5,0	23,0	10,6	24,9	11,7
11	15-6-86 6:00	19-6-86 14:00	104	107,5	98,6	325,0	13,9	101,9	11,2
12	2-6-87 1:00	5-6-87 9:00	80	10,7	6,4	30,9	3,2	41,9	11,3
13	11-7-87 23:00	19-7-87 9:00	178	50,4	44,1	230,0	14,1	266,1	7,9
14	10-8-87 17:00	17-8-87 15:00	166	31,0	17,7	99,3	10,5	148,3	7,8
15	22-8-89 13:00	25-8-89 1:00	60	32,5	41,3	178,9	4,3	77,3	9,2
16	27-5-91 2:00	29-5-91 23:00	69	27,5	9,6	69,0	13,2	26,3	14,2
17	29-4-92 2:00	30-4-92 10:00	32	7,0	3,2	14,6	3,7	27,4	8,9
18	25-5-92 18:00	27-5-92 6:00	36	9,4	4,1	20,8	2,8	81,6	7,2
19	17-4-93 4:00	18-4-93 4:00	24	5,7	1,1	8,6	4,4	12,9	12,8
20	19-4-93 3:00	19-4-93 23:00	20	13,2	6,1	28,4	6,5	48,6	12,2
21	2-6-93 16:00	5-6-93 4:00	60	14,4	4,7	24,7	7,2	23,1	11,7
22	4-6-97 0:00	5-6-97 3:00	27	8,2	3,5	21,3	3,5	37,3	10,4
23	19-6-97 20:00	24-6-97 23:00	123	25,1	15,1	82,2	11,5	139,0	9,0
24	13-6-00 18:00	14-6-00 22:00	28	35,4	24,7	108,0	17,5	132,9	9,3
25	29-6-00 11:00	4-7-00 0:00	109	34,0	19,7	105,0	13,3	55,4	9,0
26	25-5-02 20:00	26-5-02 18:00	22	25,3	11,9	49,2	11,7	68,3	10,1
27	3-6-02 5:00	5-6-02 17:00	60	20,2	4,9	33,4	11,8	219,5	6,3

Nuevamente, para esta cuenca y tipo de crecidas se realiza un análisis espectral de las crecidas, aplicando la transformada de Fourier a la serie de caudales que compone cada crecida con el objetivo de determinar cuales de ellas presentan mayores amplitudes de frecuencia, lo cual queda representado en la figura 4.27 y 4.28.

Se observa que las crecidas que mayor amplitud presentan son las crecidas número 11, 13, 15, 3, 23 y 25 y que las que presentan menor amplitud y variabilidad son las crecidas número 19, 9, 17, 22 y 4.

En la figura 4.28, se puede apreciar claramente que la crecida 11 tiene amplitudes destacadamente mayores que las demás en todo el espectro de frecuencias.

Cabe señalar que no se presenta todo el set de crecidas, especialmente aquellas con amplitud intermedia debido a la dificultad de distinguir las amplitudes de cada crecida.

Figura 4.27: Transformada de Fourier de la serie de caudales –tiempo.

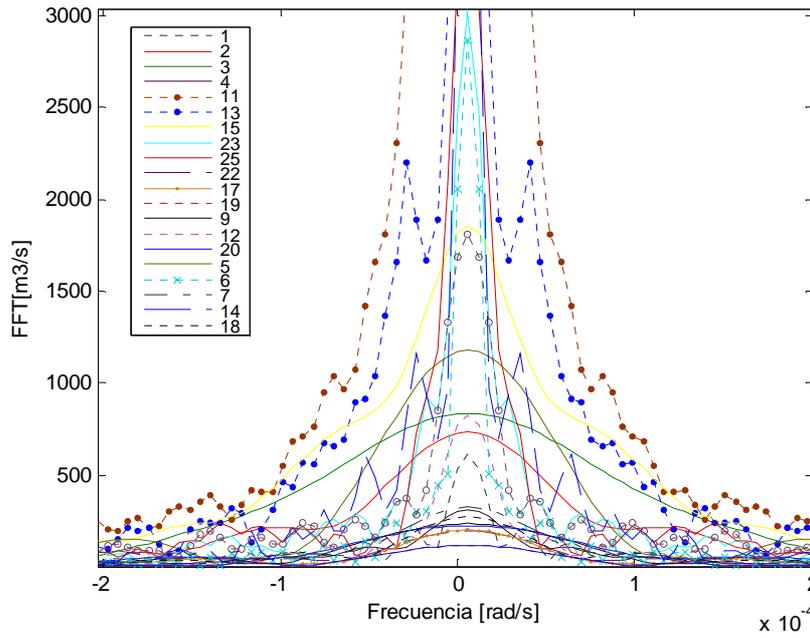
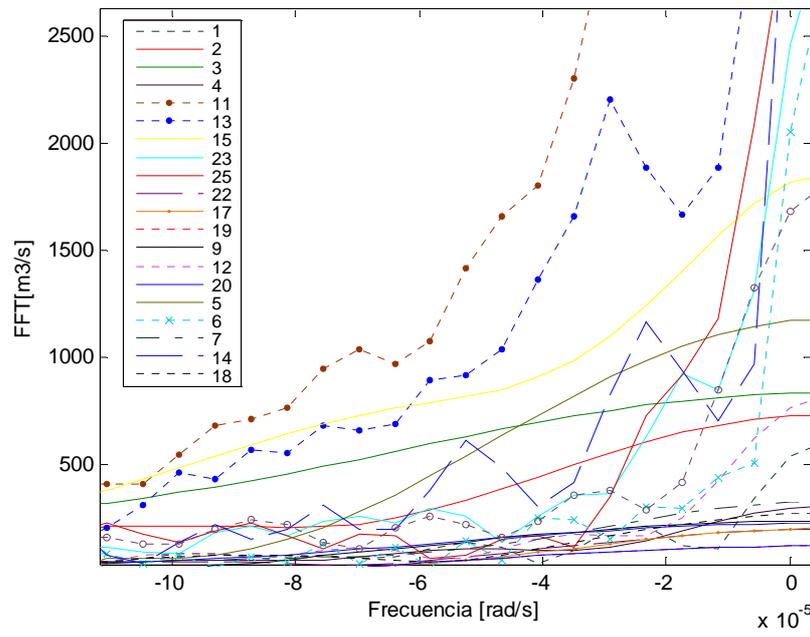


Figura 4.28: Transformada de Fourier de la serie de caudales –tiempo.



4.3.2.1 Pronósticos mediante Redes Neuronales de pronóstico - Entrenamiento Unitario

Primeramente y en este apartado se realizan pronósticos a un paso de adelanto.

De acuerdo al planteamiento general, cada crecida es usada como set de datos de entrenamiento, formando un modelo de pronóstico que es validado con las crecidas restantes. Se crean por lo tanto, 27 modelos independientes cuyos errores de entrenamiento y validación se aprecian en una matriz de errores, en que el error cuadrático medio de validación de la crecida de entrenamiento es el mismo error que se obtiene de su entrenamiento.

4.3.2.1.1 Pronóstico mediante Redes Neuronales Recurrentes de Elman

De acuerdo al planteamiento general, se consideran como parámetros de entrada las variables Q_t , $(P*A)_t$ y \dot{Q}_t . Su uso se justifica por medio de lo expuesto en el punto 4.3.1, que detalla un análisis equivalente en la cuenca del Estero Arrayán.

- Se realizan análisis para las siguientes 3 configuraciones de entrada
 - P_t*A, Q_t (Configuración 0).
 - P_t*A, Q_t, \dot{Q}_t (configuración 1).
 - $P_{t+1}*A, Q_t, \dot{Q}_t$ (Configuración 2).

Dado que el tiempo de respuesta de la cuenca ante eventos de precipitación es “**rápido**”, de aprox. 3hrs. (capítulo 3), se consideran modelos de dos capas ocultas.

Nuevamente, se prueba un distinto número de neuronas en cada una de las capas ocultas, escogiéndose la mejor arquitectura y el modelo en base a la (as) crecida (as) óptima (s) de entrenamiento para realizar pronósticos.

A. Configuración de entrada 0

Se modelan diferentes arquitecturas neuronales, con 3 neuronas en la capa de entrada, el número de neuronas de cada capa oculta es variable, de 4 a 9 y una neurona de salida (pronóstico).

En la figura 4.29 se presenta el comportamiento de los errores de validación al variar el número de neuronas de cada capa oculta para un modelo creado considerando como conjunto de datos de

entrenamiento la crecida 11. Se considera dicha crecida dado que es la que posee mayor variabilidad frecuencial y por lo tanto puede resultar la más conveniente desde el punto de vista de ser usada como set de entrenamiento, lo cual es corroborado conforme se desarrollan los modelos. Se observa que la curva mínima de errores se obtiene para una red con 5 neuronas en cada capa oculta y 1200 épocas de entrenamiento.

Luego, se considera como la arquitectura del modelo de pronóstico elegido 3-5-5-1, para el cual únicamente se varía el conjunto de datos de entrenamiento, dependiendo de la crecida que se utilice para ello.

En la tabla 4.39.a se presentan los errores de entrenamiento y ponderados obtenidos para 6 modelos NEWELM 3-5-5-1 que arrojan los errores más bajos tanto de entrenamiento, validación y prueba. Estos modelos son aquellos creados a partir del entrenamiento con las crecidas 3, 11, 13, 15, 23 y 25. Se presentan dichas crecidas por motivos de simplicidad al mostrar la información, ya que se trata de una matriz de 27* 28 en las cuales sólo destaca lo que se presenta en la tabla procedente con base en lo encontrado para el análisis de Fourier.

También, se presentan los errores obtenidos al entrenar el modelo con la crecida 24, a modo de mostrar el aumento considerable que se produce en los errores ponderados para aquellas crecidas que no presentan variabilidad frecuencial ni amplitudes considerables de dichas frecuencias. Esta consideración es extensiva a todas las matrices de este tipo que se analizan en este apartado.

Se observa que el error de entrenamiento más alto es el que se obtiene al utilizar la crecida 11 como set de datos de entrenamiento y a su vez el error ponderado más bajo es obtenido mediante este mismo entrenamiento, consecuentemente los errores de validación de cada crecida son considerablemente menores que el de entrenamiento (figura 4.39.b), lo cual otorga una amplia capacidad de generalización al modelo en el sentido de que es capaz de pronosticar tanto modelos con grandes amplitudes de frecuencia como modelos que sean menos variables.

Figura 4.29: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.

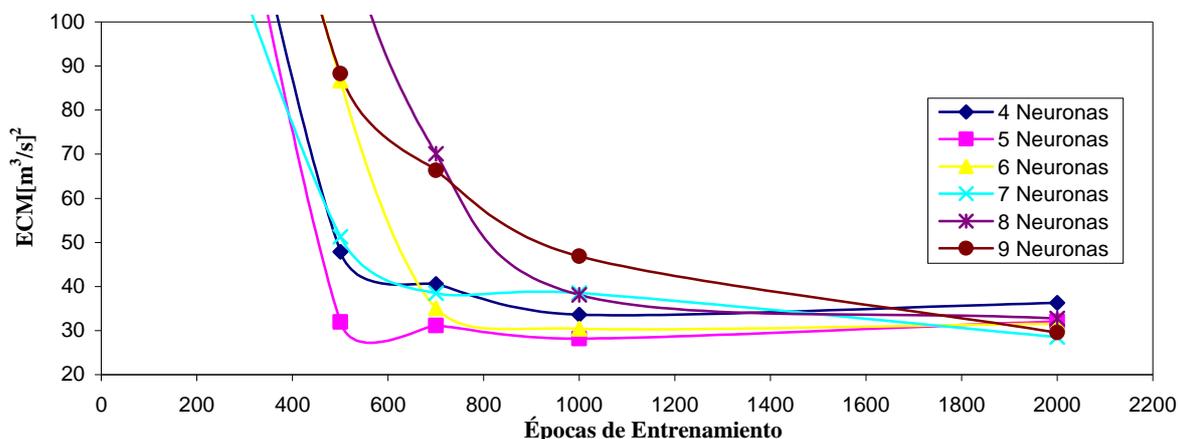


Tabla 4.39.a: Error de entrenamiento y ponderado. Modelo NEWELM, Conf. 0.

Crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	43,371	124,036
11	101	90,751	33,962
13	175	87,991	68,938
15	23	72,100	72,806
23	120	18,061	51,465
24	25	60,170	81,503
25	106	18,773	67,516

Tabla 4.39.b: Errores de validación por crecida, Conf. 0. Análisis unitario.

Crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	2,724	31,726	43,371	4,435	69,284	12,505	13,376	13,001	7,163	15,648	1111,997	9,340	251,416
11	4,129	56,338	301,313	1,457	38,918	0,376	3,809	16,174	2,387	1,455	90,751	2,776	84,801
13	6,335	93,180	383,318	4,127	57,885	4,801	5,611	16,362	5,320	7,248	439,512	3,489	87,991
15	4,268	46,340	153,933	2,423	72,091	15,396	4,983	14,844	5,427	8,858	538,602	4,942	147,987
23	4,845	74,935	270,543	1,794	70,285	1,320	4,964	17,240	3,805	3,679	261,801	3,594	102,873
24	5,064	81,699	142,639	3,844	121,737	26,938	9,121	18,326	11,412	19,029	367,877	9,243	240,153
25	5,400	76,998	297,992	3,502	58,336	7,386	5,368	15,363	5,705	8,985	418,063	3,746	116,270

Tabla 4.39.c: Errores de validación por crecida, Conf.0. Análisis unitario.

Crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	36,186	111,693	36,866	3,327	8,854	0,750	18,420	13,312	11,362	55,234	132,632	26,873	32,882	7,373
11	9,196	131,266	31,165	1,345	8,013	0,329	23,741	3,212	11,079	16,360	138,497	15,596	25,836	3,186
13	15,462	272,481	41,314	2,891	8,686	0,474	21,674	5,095	10,497	30,460	215,493	22,521	53,854	5,073
15	34,219	72,100	47,546	1,352	7,883	0,655	19,837	9,498	11,556	28,252	98,782	28,554	24,807	7,535
23	12,868	187,668	32,520	1,923	8,638	0,486	24,420	4,028	9,721	18,061	179,012	20,973	29,841	3,962
24	54,709	112,106	74,147	3,140	10,375	1,020	25,766	15,449	11,361	43,516	60,170	38,414	35,652	13,708
25	18,380	236,190	39,611	2,582	8,113	0,478	20,544	6,005	9,868	29,303	180,349	18,773	43,035	5,108

B. Configuración de entrada 1

Mediante la configuración de entrada en cuestión se pretende verificar si para el fenómeno de pronóstico de crecidas en la cuenca definida por la estación Mapocho en Los Almendros, la variable \dot{Q}_t es relevante o no .

En la tabla 4.40.a se presentan los errores de entrenamiento y ponderados obtenidos para los modelos NEWELM 3-5-5-1. Nuevamente, se observa que las crecidas 11 y 13 presentan altos errores de entrenamiento en relación al error ponderado. Considerando que el error ponderado está midiendo más que nada la capacidad de generalización de la red dada la disparidad del número de datos en los conjuntos de entrenamiento y validación, nuevamente se obtienen errores de validación bastante pequeños en algunas crecidas, lo cual se puede apreciar en la figura 4.40.b y 4.40.c.

Tabla 4.40.a: Error de entrenamiento y Ponderado. Modelo NEWELM, Conf. 1.

Crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	11,187	58,324
11	101	67,406	28,425
13	175	54,022	49,193
15	23	29,816	42,817
23	120	12,493	28,830
24	25	38,545	115,349
25	106	9,981	49,379

Tabla 4.40.b: Error de validación por crecida, Conf. 1. Modelos NEWELM, Análisis unitario.

Crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	2,950	22,673	11,187	3,376	86,056	10,325	6,742	10,734	3,334	9,919	233,851	5,175	191,131
11	3,952	43,459	118,402	2,314	41,544	4,950	2,821	15,025	3,340	2,965	67,406	2,286	87,808
13	4,662	50,349	126,042	1,838	56,034	5,177	4,028	12,743	4,680	7,283	382,432	2,321	54,022
15	4,155	41,094	83,304	2,221	65,568	9,859	3,590	15,331	5,793	7,800	301,235	2,716	83,042
23	4,557	42,082	104,495	0,649	36,051	2,941	3,833	16,401	4,652	2,890	93,905	3,057	89,273
24	3,257	42,871	84,524	2,509	112,620	22,159	5,307	12,429	9,094	15,994	1228,425	4,002	144,581
25	4,190	41,036	135,191	0,691	51,175	4,343	3,413	14,720	3,814	4,924	392,573	2,499	84,416

Tabla 4.40.c: Error de validación por crecida, Conf. 1. Modelos NEWELM, Análisis unitario.

Crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	38,389	105,290	33,096	1,570	6,157	0,476	13,056	7,689	9,426	46,198	107,845	34,572	34,069	3,593
11	10,214	64,008	31,101	0,624	7,201	0,501	21,701	4,752	11,172	15,027	112,781	18,643	36,706	5,020
13	13,341	146,069	28,106	1,356	6,751	0,558	15,463	5,638	9,674	23,438	109,052	13,714	57,568	6,266
15	15,355	29,816	33,671	0,837	8,368	0,673	22,262	6,726	9,843	17,112	88,047	19,093	35,712	6,999
23	9,731	58,871	28,293	0,973	8,520	0,555	25,289	4,072	9,515	13,493	97,494	12,118	39,486	6,444
24	35,371	58,721	35,218	1,503	7,311	0,842	18,038	11,885	7,949	30,085	38,545	22,394	25,134	11,009
25	11,121	85,304	27,210	0,980	7,234	0,469	18,918	4,443	9,024	17,689	100,063	9,981	35,540	4,952

C. Configuración de entrada 2

Dados los altos errores de pronóstico obtenidos en relación a la cuenca del Estero Arrayán en La Montosa, se estudia una forma de disminuir dichos errores, para ello se analiza la eficiencia de considerar la variable volumen de precipitación en el instante de pronóstico conocida. Para ello y para que pudiera considerarse como modelo de pronóstico, dicha variable debería ser pronosticada previamente y en caso de que no, un pronóstico de caudal mediante esta configuración de entrada sólo tendría eficiencia cuando se presentaran fallas en los equipos de medición de caudal y se quisiera pronosticar o deducir esta última variable, en base a la precipitación en el instante.

En las tablas 4.41.a, 4.41.b y 4.41.c se presentan los errores obtenidos para los distintos modelos NEWELM 3-5-5-1, formados a partir de dicha configuración de entrada.

En general, se observa que el error de entrenamiento y ponderado aumenta en relación a un modelo que no toma la variable en cuestión en el instante de pronóstico (configuración 1), razón por la cual añadirla no aporta información para el modelo de pronóstico a una hora de adelanto.

Tabla 4.41.a: Error de entrenamiento y Ponderado. Modelos NEWELM, Conf. 2.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	30,837	64,744
11	101	74,394	31,176
13	175	83,632	56,451
15	23	31,890	44,329
23	120	13,991	39,383
24	25	25,066	118,960
25	106	10,363	57,940

Tabla4.41. b: Error de validación por crecida, Conf. 2. Modelos NEWELM, Análisis unitario.

crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	3,418	32,267	30,837	3,218	80,025	17,244	6,139	12,387	6,241	11,566	350,232	5,255	189,149
11	4,412	43,747	140,350	3,907	48,117	7,812	3,110	16,990	4,019	3,984	74,394	2,836	89,494
13	4,890	52,398	186,042	1,628	34,973	3,788	4,210	14,882	3,760	5,458	443,969	2,683	83,632
15	3,918	37,756	70,645	1,793	59,263	12,599	3,754	14,926	7,393	9,178	297,618	2,901	96,755
23	4,470	30,489	108,351	1,657	42,305	4,795	2,832	17,741	2,620	2,499	113,339	3,006	85,190
24	6,517	118,275	142,549	5,912	257,793	63,406	13,599	18,464	25,250	35,383	356,555	14,658	402,838
25	5,174	48,656	143,280	0,795	41,374	6,806	4,186	15,916	6,013	5,784	464,891	3,210	99,911

Tabla4.41.c: Error de validación por crecida, Conf. 2. Modelos NEWELM, Análisis unitario.

crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	37,274	91,334	36,845	1,627	6,211	0,690	16,234	9,613	9,075	36,649	86,229	25,911	38,664	6,610
11	10,323	79,953	33,190	0,707	8,573	0,540	25,904	4,882	10,901	14,349	124,317	24,818	34,752	6,306
13	12,538	138,852	31,648	1,640	7,850	0,309	20,647	4,376	9,338	23,036	122,495	11,894	40,425	4,030
15	19,062	31,890	32,242	0,906	8,120	0,773	22,885	7,514	9,373	18,982	70,633	16,073	42,048	9,458
23	8,954	66,473	29,363	1,025	8,139	0,415	23,435	4,013	9,925	13,991	105,781	150,288	28,886	3,784
24	106,630	138,545	115,178	5,458	13,915	2,214	32,929	29,356	11,001	70,218	25,066	73,791	51,476	29,563
25	13,688	100,861	33,416	1,393	8,377	0,559	22,904	4,881	9,406	20,928	105,106	10,363	41,677	6,875

C. Análisis y Comparación de Modelos Recurrentes de Elman

Mediante una simple inspección del valor del ECM, se desprende que un modelo que NO involucra la variable gradiente de caudal no es un buen modelo de pronóstico de tipo neuronal recurrente frente a uno que si involucra dicha variable como parámetro de entrada a la red.

En la tabla 4.42 se presentan los coeficientes de determinación obtenidos para los modelos NEWELM con configuraciones de entrada de tipo 1 y 2 a partir de la crecida que es utilizada para el entrenamiento. Entonces, se desprende además, que un modelo que NO involucra la precipitación en el instante de pronóstico es un mejor modelo de pronóstico a 1 hora de adelanto.

Entonces, analizando aquel modelo formado a partir de la configuración de entrada tipo 1, la crecida utilizada en el entrenamiento que entrega menores errores de pronóstico (validación), según se desprendió anteriormente, es la número 11.

En la figura 4.30 se presenta el comportamiento de los errores de validación en modelos NEWELM 3-5-5-1 con configuraciones de entrada tipo 1 y 2, creados a partir de las diferentes crecidas de entrenamiento. Además, sólo se observa para algunas crecidas validadas, escogidas como las más representativas del fenómeno general, debido a que la presencia de todas impide apreciar el comportamiento de cada modelo a nivel de validación de crecidas.

Cabe señalar, que las líneas segmentadas de la figura representan errores de validación para modelos tipo 2 (referido a la configuración de entrada) y líneas continuas representan modelos tipo 1 por crecida de entrenamiento. Además, en la leyenda adjunta Crecida x. y (con y 1 ó 2) quiere decir, validación de la crecida x en un modelo NEWELM 3-5-5-1 de configuración de entrada y.

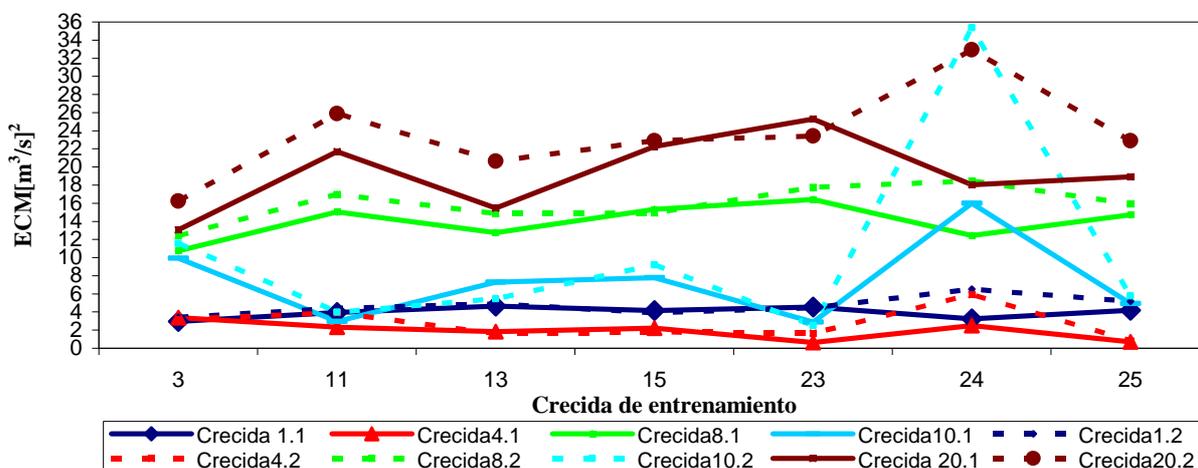
Al igual que en el punto 4.3.1.1.1.C, las curvas en sí no presentan ningún significado y sólo tienen por objetivo mostrar de manera más simple que en general el modelo tipo 2 presenta mayores errores de validación para todas las crecidas de entrenamiento.

Se corrobora la ventaja de modelos entrenados con los datos de la crecida número 11.

Tabla 4.42: Coeficiente de determinación entrenamiento y ponderado- Conf. 1 y 2.

Crecida	r^2 entrenamiento		r^2 validación	
	NEWELM 3-5-5-1 tipo 1	NEWELM 3-5-5-1 tipo 2	NEWELM 3-5-5-1 tipo 1	NEWELM 3-5-5-1 tipo 2
3	0,986	0,948	0,885	0,619
11	0,995	0,993	0,854	0,856
13	0,951	0,953	0,780	0,766
15	0,981	0,980	0,764	0,750
23	0,927	0,934	0,825	0,840
24	0,901	0,954	0,466	-0,383
25	0,972	0,971	0,788	0,723

Figura 4.30: Errores de Validación para cada modelo NEWELM.



Un análisis completo de los errores, cuyos valores se presentan en las matrices de error precedentes, muestra que un 86 % de los modelos (crecidas) presenta menores errores ponderados con un modelo tipo 1 y un 14 % mantiene sus errores o es mejor con un modelo tipo 2.

Para un modelo NEWELM 3-5-5-1 tipo 1 entrenado con los datos de la crecida 11, se obtiene un error ponderado equivalente aprox. al 9% del caudal máximo de dicha crecida.

Luego, para fines de pronóstico se puede realizar un rápido análisis de las salidas (errores) de entrenamiento, validación y prueba (pronóstico) del programa creado en función de los valores de entrada, de tal manera de encontrar la crecida que sirva como mejor modelo de pronóstico en términos de ser ocupada para entrenar el modelo neuronal y generar un modelo de pronóstico.

De antemano, es recomendable la crecida número 11, de las aquí señaladas.

4.3.2.1.2 Pronóstico mediante Redes Neuronales Prealimentadas

Con el objetivo de comparar la eficiencia de pronósticos de modelos de RNR se plantean modelos neuronales prealimentados.

Analizando las características de la cuenca, punto 3.2.2, se observa al igual que para la cuenca del Estero Arrayán un tiempo de concentración de aprox. 3 hrs. Entonces, la variable caudal que interesa debe estar retardada hasta en 3 hrs., en cuanto al volumen de precipitación se hará lo mismo.

Se evalúan dos tipos de modelos, cada uno con la siguiente arquitectura:

- 8 neuronas de entrada $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$ (Conf. 0).
- 9 neuronas de entrada. $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, \dot{Q}_t$ (Conf.1).
- 10 neuronas de entrada. $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, \dot{Q}_t, P_{t+1} * A_{t+1}$ (Conf.2).

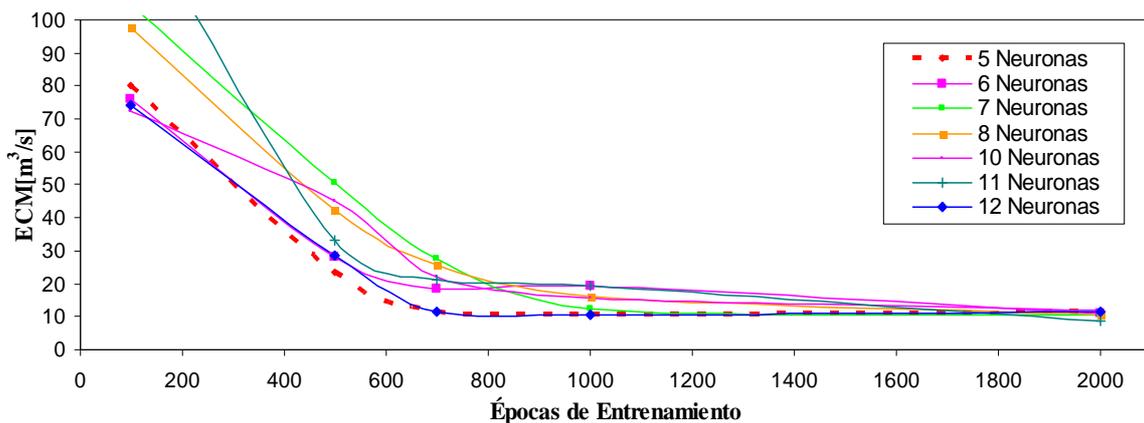
Una vez más, dada la naturaleza del problema, en este tipo de redes también se emplean dos capas ocultas, en las cuales se prueba un distinto número de neuronas para diferentes épocas de entrenamiento.

A. Configuración de entrada 0

En la figura 4.43 se aprecia la envolvente de errores de validación para modelos tipo NEWFF. Se encuentra que el número de neuronas en las capas ocultas es de 5 (envolvente mínima) y que la cantidad óptima de épocas de entrenamiento es de aprox.1500.

Luego, se elige un modelo NEWFF de arquitectura 8-5-5-1.

Figura 4.43: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.



La matriz de errores se presenta en las tablas 4.43.a, 4.43.b y 4.43.c. En términos generales los errores de entrenamiento y ponderados son del mismo orden de magnitud, pero mayores que los obtenidos para un modelo de Elman equivalente y si se observan los errores de validación, consecuentemente también son bastante mayores en algunas crecidas.

Al igual que en el caso de usar redes neuronales recurrentes de Elman, se aprecia las ventajas de un modelo entrenado con los datos de la crecida 11, en términos de que entrega los errores más bajos de pronóstico, no así de entrenamiento, donde entrega los más altos. El error ponderado de este modelo es de aprox. $42[m^3/s]^2$, correspondiente a un considerablemente mayor al obtenido para un modelo de Elman equivalente aprox. $34 [m^3/s]^2$.

Tabla 4.43.a: Error de entrenamiento y Ponderado. Modelos NEWFF, Conf. 0.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	44,895	170,445
11	101	93,651	42,901
13	175	90,891	76,427
15	23	73,399	87,255
23	120	20,095	56,927
24	25	59,975	122,410
25	106	23,801	95,317

Tabla 4.43.b: Error de validación por crecida, Conf. 0. Modelos NEWFF, Análisis unitario.

crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	10,040	165,226	44,895	9,354	123,040	15,569	18,640	23,817	7,899	12,423	287,600	20,341	620,954
11	5,097	74,371	245,415	1,726	86,518	0,496	4,257	18,672	1,925	0,631	93,651	3,699	113,372
13	7,118	112,588	299,430	4,345	90,398	4,819	6,654	16,129	3,700	4,486	497,510	4,323	90,891
15	5,361	90,100	301,864	3,985	156,155	32,435	7,970	17,997	7,683	10,293	246,042	10,056	297,229
23	5,035	69,479	254,129	1,651	82,495	2,031	4,867	17,586	3,536	3,254	357,217	3,899	94,150
24	5,733	63,652	183,258	1,207	203,814	7,987	5,406	22,788	8,361	19,803	541,850	8,784	446,678
25	6,026	105,535	293,791	4,443	105,913	11,735	4,727	14,337	4,121	7,061	599,520	4,344	184,733

Tabla 4.43.c: Error de validación por crecida, Conf. 0. Modelos NEWFF, Análisis unitario.

crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	145,305	524,702	122,649	9,447	16,444	1,535	41,614	12,267	16,887	142,851	452,355	172,427	101,030	21,266
11	8,770	158,394	44,457	1,078	8,676	0,550	28,312	3,717	15,242	21,843	265,383	28,252	28,257	4,728
13	15,169	299,434	47,186	2,677	9,540	0,379	22,638	5,391	11,787	34,153	333,973	28,209	33,625	4,356
15	70,542	73,399	80,694	1,743	10,478	0,878	28,879	17,119	13,375	51,645	161,401	64,816	17,778	13,120
23	10,029	163,824	41,022	1,612	9,298	0,450	27,287	4,216	12,790	20,095	230,348	23,035	20,219	3,791
24	41,288	378,815	51,118	2,765	12,997	1,701	31,311	7,562	14,026	31,456	59,975	46,913	13,370	16,944
25	24,718	307,692	47,771	1,516	6,972	0,456	20,068	7,774	11,019	42,510	314,778	23,801	21,625	5,596

B. Configuración de entrada 1

En la tabla 4.44.a se presentan los errores de entrenamiento y ponderados obtenidos para modelos NEWFF 9-5-5-1 tipo 1. Se observa la disminución considerable en los errores de entrenamiento y validación promedio (a través del ponderado) con respecto al modelo de configuración de entrada 0. Nuevamente, se aprecian las ventajas de la crecida 11 como set de datos de entrenamiento. Las tablas 4.43.b y 4.43.c, muestra los errores de validación, lo que refleja mejores pronósticos en casi todas las crecidas con respecto al modelo de configuración 0.

Tabla 4.44.a: Error de entrenamiento y Ponderado. Modelos NEWFF, Conf. 1.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	28,785	58,296
11	101	70,439	32,823
13	175	82,934	56,432
15	23	34,433	51,739
23	120	14,244	35,585
24	25	37,741	116,386
25	106	12,140	77,555

Tabla 4.44.b: Error de validación por crecida, Conf.1. Modelos NEWFF, Análisis unitario.

crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	3,770	24,449	28,785	1,529	64,232	3,697	6,014	15,026	2,825	5,923	412,735	4,973	128,704
11	3,853	35,333	105,474	2,307	37,470	4,119	3,397	14,646	1,738	3,822	76,439	2,530	109,820
13	5,383	59,458	216,626	1,403	45,653	4,173	4,978	14,431	2,757	4,196	432,311	2,942	85,934
15	4,149	39,063	85,102	1,824	49,252	8,887	3,561	14,173	3,015	4,892	385,203	2,684	105,998
23	4,183	37,074	106,112	0,512	36,186	2,262	3,983	15,533	2,572	2,644	210,002	2,825	83,593
24	2,440	23,821	60,967	1,726	110,800	17,308	6,413	12,554	9,735	18,837	1197,159	5,195	177,204
25	5,459	71,826	254,259	1,824	59,017	8,826	3,758	13,478	3,286	7,360	631,867	2,558	129,516

Tabla 4.44.c: Error de validación por crecida, Conf.1. Modelos NEWFF, Análisis unitario.

crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	23,575	75,167	38,153	2,177	10,412	0,543	23,055	4,894	9,574	27,949	93,420	35,059	26,227	4,937
11	12,828	65,055	32,290	1,059	8,214	0,449	23,452	5,686	12,152	17,348	135,464	30,981	31,657	4,800
13	13,049	129,701	29,399	1,629	8,156	0,275	18,069	4,218	9,073	23,271	139,514	14,397	22,654	3,345
15	21,958	37,433	30,383	0,846	6,665	0,432	17,166	6,944	9,518	20,649	103,595	24,324	21,967	4,640
23	9,886	64,006	26,132	1,094	8,199	0,351	21,923	3,475	9,938	14,244	105,662	13,511	19,934	3,492
24	36,612	59,385	30,081	2,533	7,585	1,468	20,778	10,815	10,193	27,205	37,741	27,335	27,809	17,135
25	18,726	152,404	31,221	1,073	6,121	0,388	16,740	5,784	8,899	28,845	159,661	12,140	21,340	4,571

C. Configuración de entrada 2

En la tabla 4.45.a se presentan los errores de entrenamiento y ponderados. Se puede apreciar que no existe una tendencia a mejorar o no los pronósticos, comportamiento reflejado en los errores tanto de entrenamiento, ponderados e individualmente por validación de crecidas (tablas 4.45.b y 4.45.c). Para el modelo formado a partir de la crecida de interés (11), se observa incluso que este empeora, por lo tanto incluir dicha variable no aporta información útil para el modelo de pronóstico y en consecuencia se considera irrelevante. Este último punto se refuerza de acuerdo al análisis planteado en el punto 4.3.1.3

Tabla 4.45.a: Error de entrenamiento y Ponderado. Modelos NEWFF, Conf. 2.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Ponderado [m ³ /s] ²
3	18	35,985	80,874
11	101	76,307	33,157
13	175	83,838	57,633
15	23	36,353	61,065
23	120	14,266	35,734
24	25	26,806	137,039
25	106	10,987	70,983

Tabla 4.45.b: Error de validación por crecida, Conf.2. Modelos NEWFF, Análisis unitario.

crecida	Error de validación [m ³ /s] ²												
	Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11	Valid 12	Valid 13
	86	24	18	45	23	139	32	38	17	19	101	77	175
3	3,340	22,365	35,985	2,741	72,926	10,561	6,396	13,587	3,992	8,588	672,657	5,201	155,874
11	3,796	35,795	110,808	2,300	36,485	4,125	3,402	14,173	1,853	4,099	74,307	2,459	111,090
13	5,415	57,441	206,257	1,483	39,493	4,370	4,977	14,635	3,064	4,752	456,678	2,918	83,838
15	4,259	39,449	79,582	1,949	52,393	11,625	3,676	14,059	3,820	6,265	492,306	2,953	116,906
23	4,179	38,508	106,302	0,494	38,357	2,020	4,042	15,622	2,453	2,555	208,004	2,886	85,735
24	2,205	28,411	40,929	1,210	139,039	19,309	7,142	13,119	11,297	23,944	1483,507	5,607	190,807
25	5,166	62,455	207,811	1,536	47,969	7,978	3,710	13,886	3,580	7,176	587,017	2,611	116,045

Tabla 4.45.c: Error de validación por crecida, Conf. .2. Modelos NEWFF, Análisis unitario

crecida	Error de validación [m ³ /s] ²													
	Valid 14	Valid 15	Valid 16	Valid 17	Valid 18	Valid 19	Valid 20	Valid 21	Valid 22	Valid 23	Valid 24	Valid 25	Valid 26	Valid 27
	163	57	66	29	29	21	17	34	24	120	25	106	19	57
3	30,155	91,765	36,005	1,979	8,333	0,534	20,174	8,585	9,271	36,352	117,277	30,595	25,226	4,005
11	13,925	68,555	33,582	1,072	8,079	0,439	23,653	5,906	12,283	17,795	134,486	30,320	31,891	5,028
13	13,499	130,939	29,909	1,800	8,128	0,270	19,159	4,368	9,204	24,142	135,946	13,155	23,279	3,294
15	27,089	36,353	34,105	0,925	6,729	0,463	17,771	8,091	9,473	23,099	97,551	25,870	20,430	5,401
23	9,956	62,258	25,803	1,085	8,187	0,370	21,756	3,447	9,895	14,266	107,529	14,033	20,335	3,535
24	42,408	50,305	31,825	2,968	8,405	1,565	23,254	12,360	10,670	28,083	26,806	24,163	22,527	17,522
25	17,861	135,162	31,253	1,231	6,447	0,363	18,322	5,444	8,921	26,691	142,771	10,987	20,760	4,354

4.3.2.1.3 Análisis y comparación de Modelos Neuronales

Al igual que para la cuenca del Estero Arrayán, se esquematiza el comportamiento de los errores ponderados y de los índices que cuantifican la eficiencia de un modelo de pronóstico en el máximo.

En la figura 4.31 se presentan los errores ponderados de los 4 modelos analizados. Se observa, según se desprendió unitariamente que la crecida 24 es una crecida que no serviría para realizar algún tipo de pronóstico, puesto que en todos los modelos planteados presenta errores significativamente grandes, al contrario la crecida 11 presenta los menores errores ponderados en cada uno de los modelos analizados.

Se deduce que un modelo de Elman del tipo 3-5-5-1, entrenado con la crecida 11 ó 23 con las variables definidas para un modelo tipo 1 (configuración de entrada 1) es el mejor modelo de pronóstico desde el punto de vista de obtención de los mínimos errores cuadráticos medios.

En las tablas 4.46.a, 4.46.b y 4.46.c se presentan los valores del coeficiente de determinación obtenido para cada uno de los modelos NEWELM y NEWFF tipo 0, 1 y 2. Se desprende que desde este índice de eficiencia, el modelo NEWELM 3-5-5-1 tipo 1 sigue siendo el mejor modelo de pronóstico.

Figura 4.31: Errores Ponderados para cada modelo NEWELM y NEWFF.

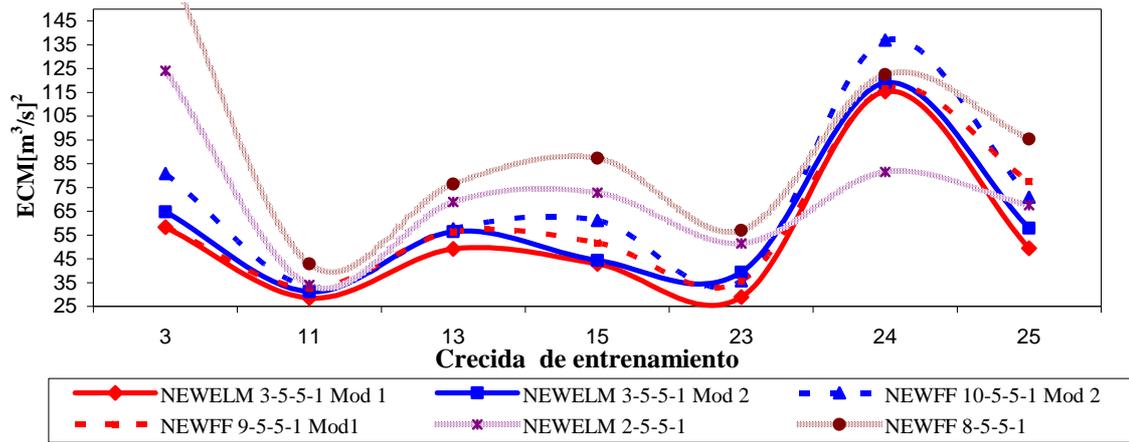


Tabla 4.46.a: Coeficiente de determinación. Modelos NEWELM y NEWFF.

crecida	Modelo 0 NEWELM 3-5-5-1		Modelo 0 NEWFF 9-5-5-1	
	r^2 entrenamiento	r^2 ponderado	r^2 entrenamiento	r^2 ponderado
3	0,935	0,477	0,935	-0,739
11	0,988	0,860	0,984	0,859
13	0,942	0,694	0,931	0,679
15	0,952	0,640	0,950	0,523
23	0,907	0,791	0,908	0,789
24	0,877	0,322	0,960	-0,020
25	0,945	0,643	0,930	0,534

Tabla 4.46.b: Coeficiente de determinación. Modelos NEWELM y NEWFF.

crecida	Modelo 1 NEWELM 3-5-5-1		Modelo 1 NEWFF 9-5-5-1	
	r^2 entrenamiento	r^2 ponderado	r^2 entrenamiento	r^2 ponderado
3	0,986	0,885	0,971	0,718
11	0,995	0,854	0,992	0,845
13	0,951	0,780	0,950	0,770
15	0,981	0,764	0,976	0,780
23	0,927	0,825	0,927	0,833
24	0,901	0,466	0,926	0,427
25	0,972	0,788	0,965	0,693

Tabla 4.46.c: Coeficiente de determinación. Modelos NEWELM y NEWFF.

crecida	Modelo 2 NEWELM 3-5-5-1		Modelo 2 NEWFF 9-5-5-1	
	r^2 entrenamiento	r^2 ponderado	r^2 entrenamiento	r^2 ponderado
3	0,948	0,619	0,964	0,680
11	0,993	0,856	0,992	0,845
13	0,953	0,766	0,952	0,759
15	0,980	0,750	0,977	0,745
23	0,934	0,840	0,927	0,835
24	0,954	-0,383	0,948	0,235
25	0,971	0,723	0,969	0,705

En las figuras 4.32 y 4.33 se presentan los índices de eficiencia de pronóstico de caudal máximo en los datos de entrenamiento para cada uno de los modelos NEWELM y NEWFF. De la misma forma, en las figuras 4.35 y 4.36 se presentan dichos índices medidos en el pronóstico de la crecida 13, a modo de ejemplo, para cada uno de los modelos planteados y de acuerdo a la crecida utilizada en el entrenamiento.

No se presenta la crecida 24 como crecida de entrenamiento debido a los altos errores de validación obtenidos, razón por la cual ser estudiada a este nivel no tiene relevancia frente al resto.

Se observa que para un modelo NEWELM tipo 1 los errores de pronosticar el caudal máximo de una determinada crecida son menores en relación a cualquier otro modelo de pronóstico de redes neuronales, tanto en el valor al tiempo de ocurrencia observado, como en cuanto a los máximos pronosticados.

En particular, para un modelo NEWELM 3-5-5-1 tipo 1 entrenado con la crecida 11, el error de pronóstico de caudal máximo es inferior al 3% de sobrestimación, aunque en general los modelos formados a partir del entrenamiento por medio de otras crecidas producen una subestimación del caudal máximo.

Con respecto al pronóstico de máximos durante validación y ejemplificando dicho comportamiento en la crecida 13, se observa que la diferencia entre el máximo observado y pronosticado es inferior al 1%, con un desfase nulo, utilizando el modelo anteriormente mencionado. Valores mucho más grandes en el caso de un modelo NEWFF 8-5-5-1 ó 9-5-5-1 ó 10-5-5-1.

En la tabla 4.47 se aprecian los desfases existentes entre la ocurrencia del máximo observado y pronosticado para cada modelo, también, en el entrenamiento y en la validación de la crecida 13. Cabe señalar, que no se ha presentado gráficamente por razones de simplicidad, ya que los valores son ± 1 ó 0 y sólo se estudian las 6 crecidas de entrenamiento mencionadas.

Nuevamente, desde estos índices de eficiencia se aprecia las ventajas de considerar la variable \dot{Q} como una variable de entrada relevante para el pronóstico, no así el volumen de precipitación en el instante de pronóstico, que no produce mejoras.

Figura 4.32: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.

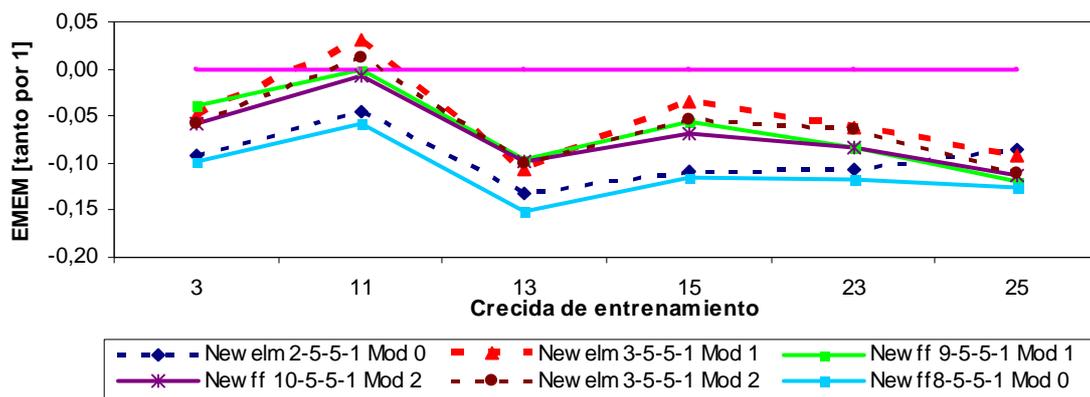


Figura 4.33: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.

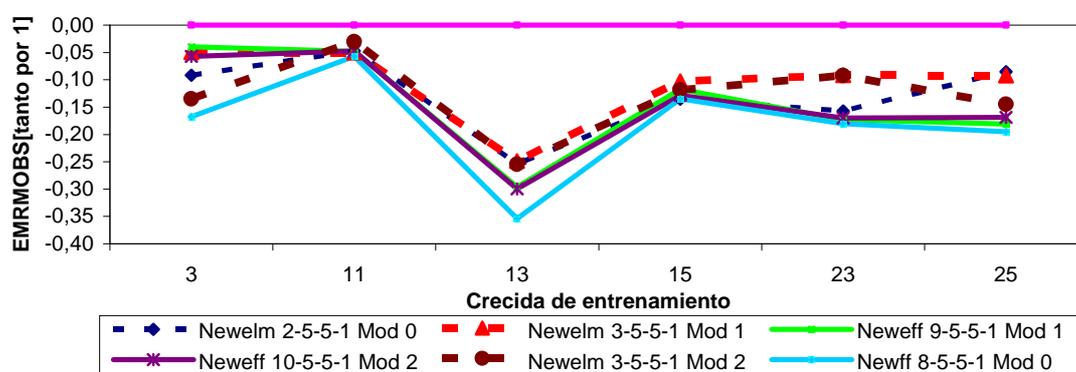


Figura 4.34: Índice EMEM validación Crecida 13. Modelos NEWELM y NEWFF.

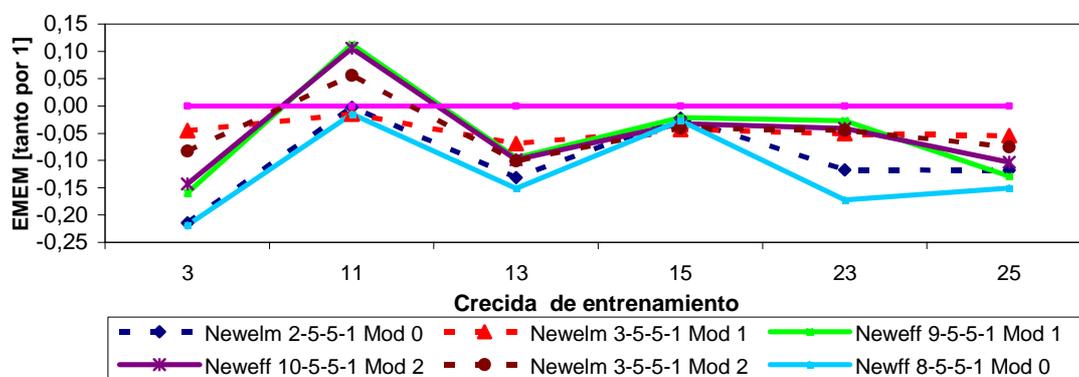


Figura 4.35: Índice EMRMOBS validación Crecida 13. Modelos NEWELM y NEWFF.

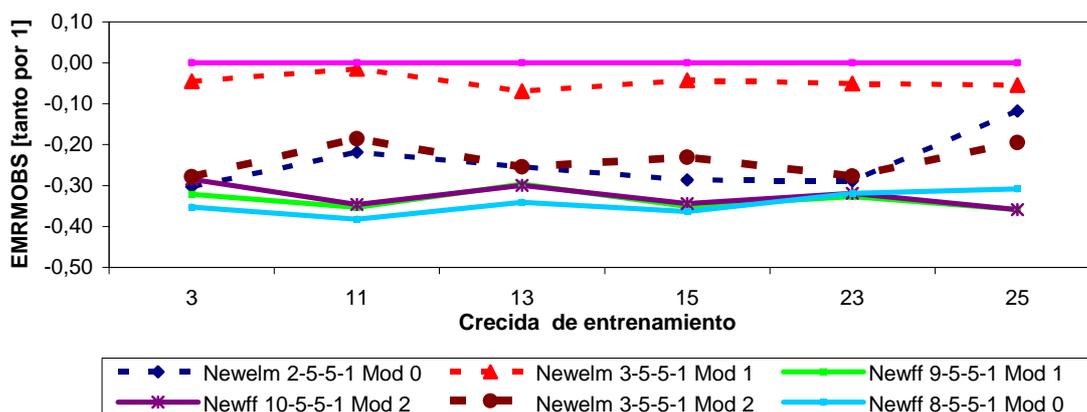


Tabla 4.47: Índice TMEM entrenamiento Crecida 13. Modelos NEWELM y NEWFF.

Crecida	TMEM Entrenamiento [hr]						TMEM Validación Crecida 13 [hr]					
	Modelo NEWELM			Modelo NEWFF			Modelo NEWELM			Modelo NEWFF		
	Modelo 0	Modelo 1	Modelo 2	Modelo 0	Modelo 1	Modelo 2	Modelo 0	Modelo 1	Modelo 2	Modelo 0	Modelo 1	Modelo 2
3	0	0	0	-1	0	0	-1	-1	1	1	1	-1
11	1	0	0	1	1	1	1	0	0	1	0	1
13	1	0	1	1	1	1	1	0	1	1	1	1
15	1	0	0	1	0	1	1	1	1	1	1	1
23	-1	1	1	2	1	1	1	0	0	1	-1	0
25	1	0	0	1	1	1	0	0	1	1	1	1

4.3.2.2 Pronóstico mediante Redes Neuronales -Entrenamiento Continuo

4.3.2.2.1 Pronóstico mediante Redes Neuronales Recurrentes de Elman

De acuerdo al planteamiento general, se forman tres conjuntos, uno de entrenamiento, otro de validación y un último de prueba, los cuales contienen 897 (crecidas 2, 3, 4, 5, 11, 13, 14, 15, 16, 23, 24, 25), 307(crecidas 7, 8, 10, 12, 20, 21, 27) y 422 (crecidas 6, 9, 14, 17, 18, 19, 22) datos respectivamente.

El conjunto de datos de entrenamiento es formado considerando las crecidas que tienen más componentes de frecuencia y más amplitud de ellas, debido a que si bien a partir de esa decisión los errores de entrenamiento aumentan, los errores de validación y prueba decrecen considerablemente y por lo tanto el error ponderado de realizar pronósticos es más bajo que en el caso de que no se considere dicho análisis.

Luego, un modelo de pronóstico creado a partir de esta selección de crecidas tiene por objetivo lograr una mayor capacidad de generalización o de la misma manera que el modelo sea capaz de pronosticar tanto crecidas con mucha variabilidad y altas frecuencias como crecidas que sean más simples.

Las características de cada uno de los conjuntos de entrenamiento se presentan en la tabla 4.48.

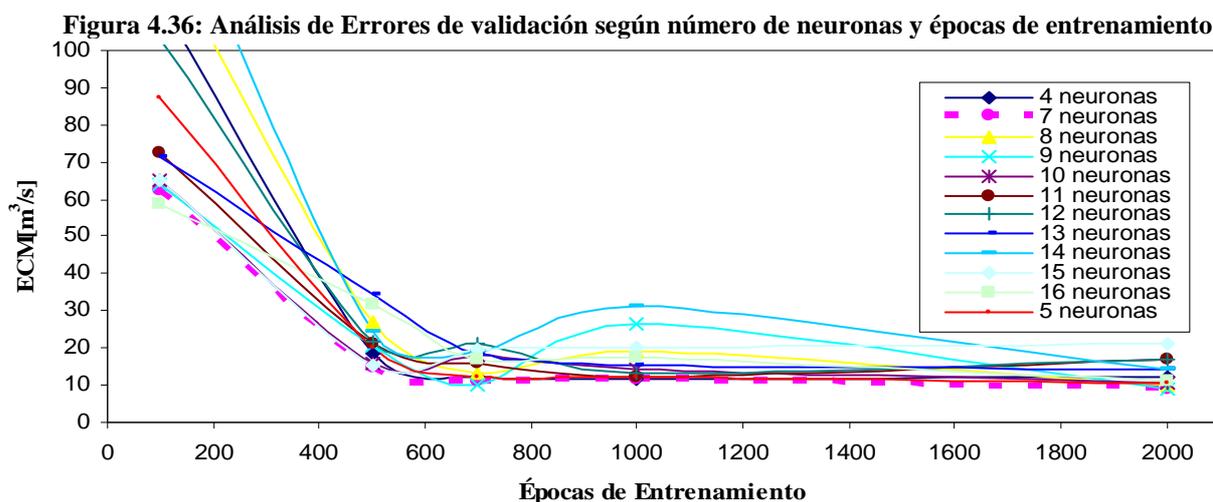
Tabla 4.48: Caracterización conjuntos de datos.

CRECIDA	MEDIA [m ³ /s]	DESV ESTANDAR [m ³ /s]	Q Máx [m ³ /s]	QMin [m ³ /s]
ENTRENAMIENTO	41,4	48,6	325,0	4,3
VALIDACIÓN	13,4	7,0	33,4	2,3
PRUEBA	8,7	7,4	32,7	1,8

Nuevamente, se utilizan modelos neuronales de 2 capas ocultas. Se realiza configuraciones con 3 neuronas en la capa de entrada definidas por las mismas variables que a nivel de entrenamiento unitario resultaron ser más eficientes: $P_t * A$, Q_t y \dot{Q}_t (configuración 1).

A. Configuración de entrada 1

Se prueban distintos números de neuronas (5 a 15) en cada una de las capas ocultas. La figura 4.36 muestra el análisis de los errores de validación en función del número de épocas de entrenamiento y neuronas de las capas ocultas, se deduce que el modelo óptimo considera 7 neuronas en cada capa oculta y aproximadamente 1800 épocas de entrenamiento.



En la tabla 4.49 se presentan los valores de los errores obtenidos en cada uno de los set de datos considerados para distintas arquitecturas de RNR, donde se aprecia claramente que para un modelo NEWELM los menores errores cuadráticos medios de pronóstico se obtienen para un modelo con 7 neuronas en cada capa oculta.

En cuanto a los valores de los errores ponderados se observa que los menores errores varían aprox. entre 33 y 38 $[m^3/s]^2$, con un mínimo de 33,01 $[m^3/s]^2$, valor marcado principalmente por el alto error de entrenamiento y el bajo error de validación y prueba.

Cabe destacar que NO debe entenderse que simplemente se considera las crecidas más “difíciles” de pronosticar en el entrenamiento y por eso se obtienen errores de validación y prueba pequeños, sino que nuevamente, mediante un entrenamiento de ese tipo, el modelo es capaz de lograr una buena generalización de tal manera de pronosticar todo tipo de crecidas, además de las incluidas en este análisis, donde pueden presentarse otras con amplitud de frecuencia mayor.

Tabla 4.49: Errores de entrenamiento y Ponderado Modelos NEWELM, Conf. 1. Análisis Continuo.

Configuraciones	ECM [m ³ /s] ²			Ponderado [m ³ /s] ²	r ² entrenamiento	r ² validación	r ² prueba
	entrenamiento	Validación	Prueba				
	897	307	422				
3-4-4-1	80,546	12,132	6,767	48,481	0,961	0,849	0,963
3-5-5-1	65,315	10,491	6,979	39,824	0,961	0,843	0,961
3-7-7-1	53,611	9,137	6,588	33,010	0,971	0,884	0,966
3-8-8-1	65,558	10,081	5,072	39,385	0,973	0,889	0,975
3-9-9-1	59,684	9,049	6,287	36,266	0,973	0,898	0,969
3-10-10-1	62,751	10,571	7,894	38,662	0,972	0,881	0,963
3-11-11-1	101,796	16,775	15,686	63,395	0,954	0,797	0,921
3-12-12-1	98,931	17,113	17,744	62,413	0,956	0,794	0,919
3-13-13-1	110,954	14,151	10,777	66,677	0,951	0,848	0,950
3-14-14-1	105,643	14,481	18,148	65,723	0,954	0,856	0,913
3-15-15-1	90,858	21,232	15,377	58,122	0,960	0,808	0,922
3-16-16-1	66,978	11,316	8,261	41,230	0,971	0,874	0,960
3-7-9-1	96,524	12,631	10,553	58,372	0,956	0,857	0,945
3-7-5-1	71,765	11,508	6,070	43,338	0,966	0,862	0,968
3-5-10-1	54,184	9,528	7,137	33,542	0,974	0,891	0,964

4.3.2.2.2 Pronóstico mediante Redes Neuronales Prealimentadas

Igualmente, se elige como mejor configuración de entrada aquella que entrega mejores pronósticos a nivel de entrenamiento unitario. Luego, las entradas a la red quedan definidas por las siguientes 9 variables: $P_t * A_t, P_{t-1} * A_{t-1}, P_{t-2} * A_{t-2}, P_{t-3} * A_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}, \dot{Q}_t$

La figura 4.37 muestra el análisis de los errores de validación en función del número de épocas de entrenamiento y neuronas de las capas ocultas. Se observa que la envolvente mínima de errores considera 5 neuronas en cada capa oculta para un óptimo de 1200 épocas de entrenamiento, al igual que en el caso unitario.

En la tabla 4.50 se muestran los errores obtenidos en cada uno de los set de datos para distintas arquitecturas de RNR, donde se aprecia que el número óptimo de neuronas en cada capa oculta es precisamente de 5 neuronas. Aunque se puede observar errores ponderados cercanos a los obtenidos para este último modelo, en aquellos con arquitectura 9-7-7-1, 9-10-10-1 y 9-12-12-1.

El mínimo ECM ponderado es de 43,983 [m³/s]². Considerablemente superior al encontrado para el modelo NEWELM equivalente (en el sentido de que considera las mismas variables pero sin los retardos correspondientes) y a su vez a los errores obtenidos para el modelo unitario tanto NEWELM como NEWFF.

Figura 4.37: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.

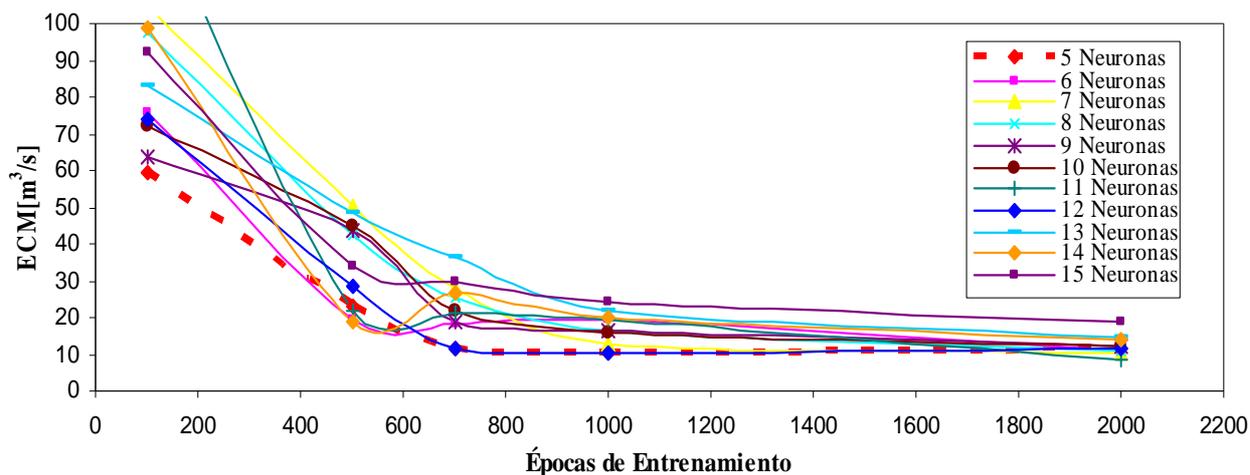


Tabla 4.50: Errores de entrenamiento y Ponderado Modelos NEWFF, Conf.1. Análisis Continuo.

Configuraciones	ECM [m^3/s] ²			Ponderado [m^3/s] ²	r ² entrenamiento	r ² validación	r ² prueba
	entrenamiento	Validación	Prueba				
	897	307	422				
9-5-5-1	72,046	11,335	8,082	43,983	0,957	0,856	0,955
9-6-6-1	79,111	10,920	7,571	47,669	0,962	0,866	0,959
9-7-7-1	73,839	10,596	7,217	44,608	0,965	0,872	0,962
9-8-8-1	78,239	10,790	7,386	47,115	0,963	0,870	0,962
9-9-9-1	83,124	12,145	10,391	50,846	0,962	0,852	0,947
9-10-10-1	72,291	11,841	9,398	44,555	0,967	0,863	0,954
9-11-11-1	75,482	8,763	5,997	44,851	0,975	0,896	0,971
9-12-12-1	73,140	11,749	6,795	44,330	0,968	0,865	0,966
9-13-13-1	72,146	14,266	9,944	45,074	0,968	0,818	0,947
9-14-14-1	79,984	13,645	10,396	49,399	0,964	0,816	0,946
9-15-15-1	84,176	18,492	23,898	56,131	0,963	0,782	0,890
9-7-5-1	78,634	10,866	7,493	47,376	0,962	0,867	0,960
9-5-7-1	81,951	11,089	7,749	49,314	0,960	0,862	0,958
9-5-10-1	83,671	11,232	8,880	50,584	0,960	0,862	0,953
9-5-12-1	75,710	11,465	7,796	45,954	0,9643	0,8584	0,9588

En la tabla 4.51 se presentan los índices de error producidos en el pronóstico de caudales máximos para cada conjunto de datos y según tipo de modelo empleado.

Se aprecia que un modelo NEWELM presenta mejores pronósticos de caudal máximo tanto en valor como en desfase, en relación al último modelo estudiado (NEWFF 9-5-5-1), al igual que en el caso de modelos con entrenamiento unitario.

Para un modelo NEWELM 3-7-7-1 tipo 1, el error entre el caudal observado y pronosticado promedio, en el momento en que ocurre el máximo, es de aprox. un 13%, diferencia que se reduce a un 7% en el caso de máximos observados y pronosticados en su momento de ocurrencia individual. Estos valores son bastante mayores cuando se utiliza un modelo NEWFF 9-5-5-1.

Se observa que la incorporación de la variable \dot{Q}_t no tiene el mismo efecto que en un análisis tipo unitario, ya que si bien el desfase entre caudales observados y pronosticados es nulo durante el entrenamiento en un modelo de Elman, no lo es en el conjunto de validación y prueba. La

razón de ello puede ser la inestabilidad que crea en el pronóstico la serie “continua” o pegada de crecidas.

Tabla 4.51: Errores de Pronóstico de caudal máximo a un paso de adelanto.

Datos	ENTRENAMIENTO-VALIDACIÓN-PRUEBA -2000 épocas de entrenamiento					
	Modelo 3-7-7-1 NEWELM			Modelo 9-5-5-1 NEWFF		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [hr.]
Entrenamiento	-0,071	-0,071	0	-0,129	-0,132	1
Validación	-0,069	-0,122	1	-0,162	-0,249	1
Prueba	-0,055	-0,199	1	-0,094	-0,264	1

4.3.2.3 Comparación de Modelos Neuronales de Pronóstico por crecidas

A continuación en las figuras 4.38 y 4.39 se presentan una serie de crecidas validadas con los modelos neuronales NEWELM y NEWFF creados a partir de configuraciones que involucran \hat{Q}_t , específicamente con la configuración de entrada 1 y para los distintos tipos de entrenamiento estudiados. Se puede apreciar gráficamente las ventajas de realizar pronósticos mediante redes neuronales recurrentes de Elman entrenadas con los datos de la crecida 11 sobre modelos de redes prealimentadas entrenadas tanto continuamente como unitariamente.

Figura 4.38: Validación de crecidas según red neuronal y tipo de entrenamiento.

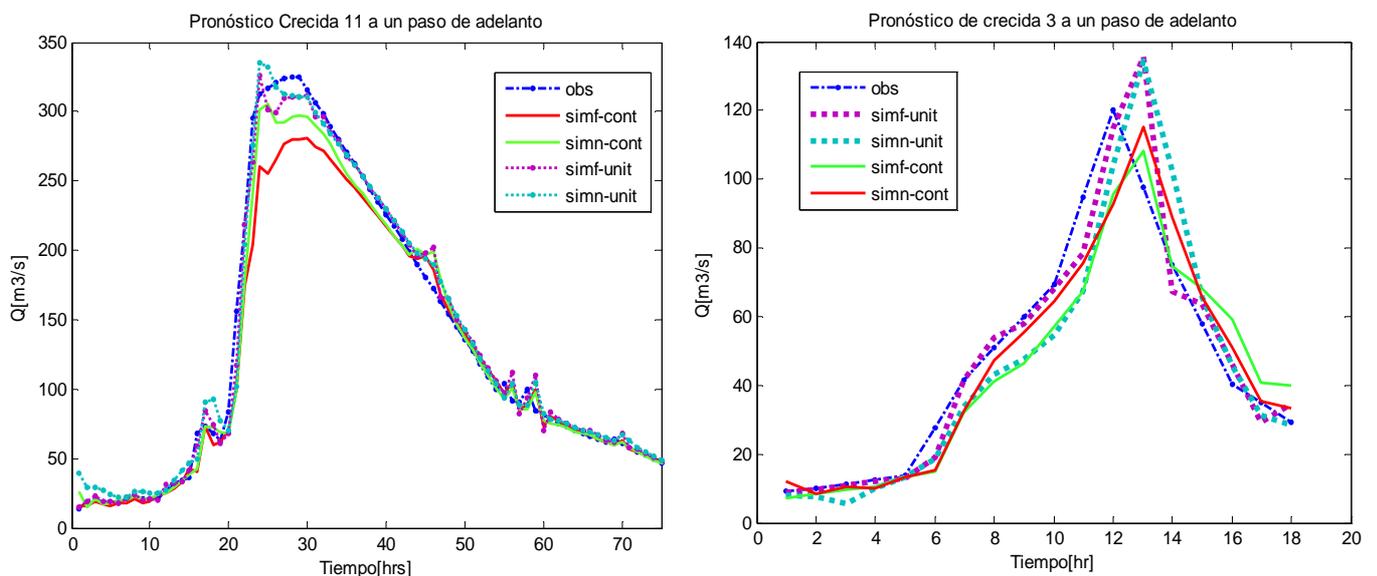
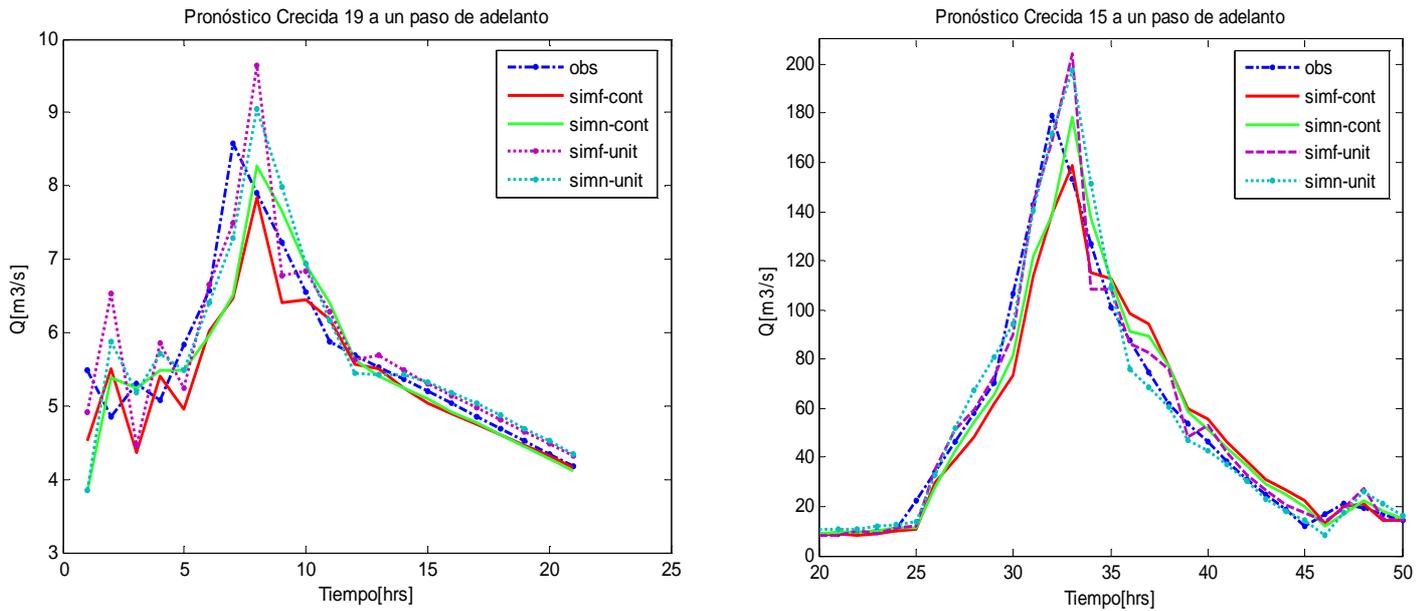


Figura 4.39: Validación de crecidas según red neuronal y tipo de entrenamiento.



4.3.2.4 Pronóstico mediante Modelos Paramétricos

Los tres conjuntos usados para el pronóstico mediante modelos neuronales se utilizan para realizar un análisis de pronóstico mediante modelos paramétricos tipo ARMAX y RARMAX a una hora de adelanto.

Las variables de entrada son las mismas que se consideran para los modelos anteriores, el volumen de precipitación, el gradiente de caudal y el caudal en tiempos anteriores para producir la salida en $t+1$ c/r a entradas en t .

Los retardos de caudal (n_a), volumen de precipitación (n_b) y gradiente de caudal se varían de 1 a 6 hrs. y el desfase inicial de precipitación desde 1 a 5 hrs. (n_k). Luego, se forman 180 modelos, de tal manera de encontrar la mejor configuración para cada uno de los modelos en base al ECM ponderado, en el sentido de encontrar cual es el límite de influencia de cada una de las variables en cuestión.

Al realizar el análisis de los errores se observa, al igual que para la cuenca del Estero Arrayán que un modelo que involucra la variable gradiente de caudal como parámetro de entrada no es más eficiente que uno que no lo hace, por lo tanto en un modelo de tipo paramétrico dicha variable no aporta información adicional al modelo de pronóstico y resulta innecesaria. Los

errores entregados para estos modelos no se presentan debido a su similitud con los modelos que NO involucran \dot{Q}_i .

En la tabla 4.52 se observan los mejores modelos tipo ARMAX y RARMAX (su configuración), obtenidos de acuerdo a una comparación del error cuadrático medio ponderado, valores que se observan en la tabla 4.53.

Los valores del ECM y el coeficiente de determinación asociado a cada conjunto de datos para ambos tipos de modelos se pueden observar en la tabla 4.54 y 4.55.

Tabla 4.52: Mejores Modelos ARMAX y RARMAX.

Tabla 4.52.a: Modelos ARMAX.

Modelo	na	nb	nk
31	3	5	1
42	7	6	1
84	7	5	2
38	3	6	1
33	5	5	1
26	5	4	1
39	4	6	1

Tabla 4.52.b: Modelos RARMAX.

Modelo	na	nb	nk
100	2	1	3
107	2	2	3
2	2	1	1
247	2	1	6
99	1	1	3
150	3	1	4
198	2	1	5

Tabla 4.53: Error Cuadrático Medio Ponderado Modelo ARMAX y RARMAX.

Tabla 4.53.a: Modelos ARMAX.

Modelo	ECM ponderado [m ³ /s] ²	R ² ponderado
31	34,769	0,949
42	34,817	0,948
84	34,837	0,950
38	34,898	0,950
33	34,960	0,950
26	34,979	0,949
39	34,980	0,947

Tabla 4.53.b: Modelos RARMAX.

Modelo	ECM ponderado [m ³ /s] ²	R ² ponderado
100	46,657	0,928
107	50,038	0,924
2	50,064	0,927
247	51,253	0,928
99	51,684	0,934
150	52,790	0,912
198	53,659	0,927

Tabla 4.54: Error cuadrático medio Modelos ARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento [m ³ /s] ²	ECM Validación [m ³ /s] ²	ECM Prueba [m ³ /s] ²
31	0,977	0,848	0,965	54,696	13,946	7,562
42	0,977	0,841	0,965	54,519	14,645	7,614
84	0,977	0,849	0,966	54,948	13,896	7,322
38	0,977	0,851	0,964	54,974	13,671	7,668
33	0,977	0,851	0,965	55,138	13,689	7,546
26	0,977	0,844	0,965	54,944	14,310	7,578
39	0,977	0,837	0,964	54,649	15,008	7,701

Tabla 4.55: Error cuadrático medio Modelos RARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento [m ³ /s] ²	ECM Validación [m ³ /s] ²	ECM Prueba [m ³ /s] ²
100	0,969	0,839	0,956	72,926	20,975	9,503
107	0,967	0,845	0,958	78,487	22,512	9,594
2	0,966	0,847	0,961	79,390	21,102	8,799
247	0,965	0,838	0,956	81,605	20,284	9,267
99	0,965	0,859	0,965	83,599	17,862	8,451
150	0,966	0,846	0,958	81,393	27,767	10,194
198	0,964	0,848	0,959	86,030	20,047	9,303

Se desprende que un modelo ARMAX presenta mejores pronósticos que un modelo RARMAX, la diferencia se debe principalmente al aumento de los errores tanto de entrenamiento, validación y prueba que se cometen en el segundo modelo con respecto al primero.

Eligiendo la mejor configuración de cada modelo ARMAX y RARMAX (31 y 100) y realizando una comparación del error cuadrático medio obtenido para cada uno con un modelo de Elman 3-5-5-1 entrenado con los datos de la crecida 11, se desprende que el error de este último es un 6 % menor que el obtenido para un modelo ARMAX y un 30% menor que el de un modelo RARMAX a nivel de entrenamientos continuos.

Al igual que para modelos neuronales, se realiza también un análisis unitario, donde cada modelo tiene la mejor configuración obtenida para ARMAX y RARMAX “continuo” y es entrenado de acuerdo a los datos de cada crecida, formándose 27 modelos. En la tabla 4.56 se presentan los errores ponderados obtenidos para cada modelo, siendo los mejores aquellos señalados en amarillo.

Tabla 4.56: Errores Ponderados Modelo ARMAX y RARMAX.

Nª Crecida	RARMAX	ARMAX
	Error Ponderado [m3/s] ²	
1	70,111	38,888
2	70,364	88,898
3	77,715	254,224
4	70,024	43,966
5	70,012	47,622
6	70,128	44,692
7	70,286	56,662
8	70,958	111,857
9	72,058	40,354
10	70,146	63,028
11	73,170	34,234
12	70,554	39,706
13	71,760	35,220
14	75,734	36,976
15	71,155	45,476
16	70,564	54,901
17	70,444	45,446
18	70,357	38,233
19	70,141	490,742
20	70,519	84,482
21	70,208	38,439
22	70,271	270,493
23	70,224	35,874
24	70,825	47,899
25	70,875	40,529
26	70,050	159,695
27	70,100	42,667

Se desprende que modelos RARMAX entrenados continuamente entregan mejores resultados que aquellos formados a través de entrenamientos de tipo unitario. Las razones que fundamentan aquello son las mismas que se exponen en el punto 4.3.1.4, respecto del ajuste de datos punto a

punto. En general modelos ARMAX entrenados continuamente entregan menores errores, pero para los datos de la crecida 11 el valor del ECM es un poco menor.

Las crecidas 11, 13 y 23 como datos de entrenamiento forman los mejores modelos ARMAX para realizar pronósticos y las crecidas 5, 4 y 26 definen los mejores modelos tipo RARMAX. Entonces, al igual que los resultados encontrados para la cuenca del Estero Arrayán, las crecidas con mayor amplitud de frecuencias como datos de entrenamiento forman los mejores modelos ARMAX y a la inversa, las crecidas con menor amplitud de frecuencias forman los mejores modelos RARMAX.

Con respecto al valor de los errores ponderados presentados en la tabla 4.56 se observa que si bien el modelo RARMAX en general entrega valores más altos que un modelo ARMAX, entrega también, un desempeño mucho más equilibrado para cada uno de los modelos (crecidas) en comparación a un modelo ARMAX, y en aquellos destacados en naranja incluso presenta errores más pequeños. La razón, al igual que en el caso de la cuenca del Estero Arrayán radica en que en un modelo RARMAX el entrenamiento fija **sólo** las condiciones iniciales de los parámetros característicos y no los parámetros en sí.

En términos porcentuales un modelo de redes neuronales recurrentes NEWELM 3-7-7-1 entrenado con los datos de la crecida 11 entrega un error aprox. 18% menor que un modelo ARMAX con un entrenamiento del mismo tipo, porcentaje aún mayor en el caso de ser comparado con el modelo RARMAX (60%).

4.3.2.5 Pronóstico mediante Red Neuronal Recurrente de Hopfield - Identificación

Los puntos expuestos acerca de este tipo de red neuronal en el apartado 4.2.5 sobre su utilidad, beneficios y discapacidades tienen validez en cada cuenca que se desee estudiar y por lo tanto en la cuenca definida por la estación Mapocho en Los Almendros.

Cada patrón es formado de tal manera de contener la información relevante para pronosticar 1 hora después y con el objetivo de representar a un determinado “punto” (tiempo, caudal (t)).

Se consideran 101 patrones constituidos por la información de la crecida 11, crecida que posee la mayor capacidad de generalización y que forma el mismo conjunto que se utiliza para realizar un análisis considerando un entrenamiento unitario con resultados eficientes en una red de Elman.

Las variables contenidas en cada uno de los patrones son: $P_t * A_t$, Q_t , \dot{Q}_{t-1} , \dot{Q}_t y \dot{Q}_{t+1} . Cabe recordar que la red es de interconexión total entre neuronas y que cada componente del patrón

define una neurona con diferentes estados que caracterizan las trayectorias posibles a dicho patrón.

Los vectores de entrada poseen las mismas variables que los patrones almacenados, con la excepción de la última variable (\dot{Q}_{t+1}), la ausencia de ésta es reemplazada por la variable antecesora y con el objetivo de ser rescatada en el proceso de identificación del vector.

Para garantizar el correcto funcionamiento de la red, cada uno de los patrones y vectores de entrada es codificado binariamente, para ello se considera el mismo proceso indicado en el caso de pronósticos mediante ésta red en la cuenca del Estero Arrayán en La Montosa. Luego, lo que se conoce como vector patrón de la red caracterizado por 5 variables, queda definido por un vector que posee 30 componentes en total, por lo tanto la red está formada por 30 neuronas y una totalidad de 435 ($30 \cdot 29 / 2$) interconexiones entre ellas.

Una vez creada la red, en base a la realización de pruebas en que se varía el número de iteraciones necesarias para garantizar la convergencia hacia el patrón almacenado, se decide que una cantidad de 8 iteraciones es suficiente para alcanzar la convergencia a un determinado patrón dependiendo de la entrada.

El funcionamiento y teoría del modelo es la misma que se detalla en la cuenca del Estero Arrayán en La Montosa, expuesta en el punto 4.3.1.5. Sólo a modo de situar al lector en las bases del modelo, se recuerda que éste posee dos etapas, una de identificación de \dot{Q}_{t+1} y otra de estimación de Q_{t+1} (valor pronosticado).

En la tabla 4.57 se presentan los errores estándar y cuadráticos medios para cada crecida validada, así como también los índices de eficiencia de pronóstico de caudal máximo. Se observa que los errores son del mismo orden que los obtenidos para modelos neuronales de pronóstico tanto prealimentados como recurrentes de Elman, aunque mayores que los obtenidos para modelos unitarios, pero, en términos generales bien similares a los obtenidos para un modelo prealimentado entrenado continuamente.

Se observa que en general existe sobre-estimación del caudal máximo, pero, en el momento de ocurrencia real de dicho valor se presenta más bien una subestimación con respecto al valor observado. El hecho de que existan mayormente sobre-estimaciones de los valores máximos de caudal se debe a que el modelo de pronóstico trabaja en base a la identificación del gradiente conocida la tendencia precedente. Marcadamente cuando se está en la zona de inflexión, donde es importante la diferencia entre el gradiente antecesor al tiempo de pronóstico y el valor nulo de

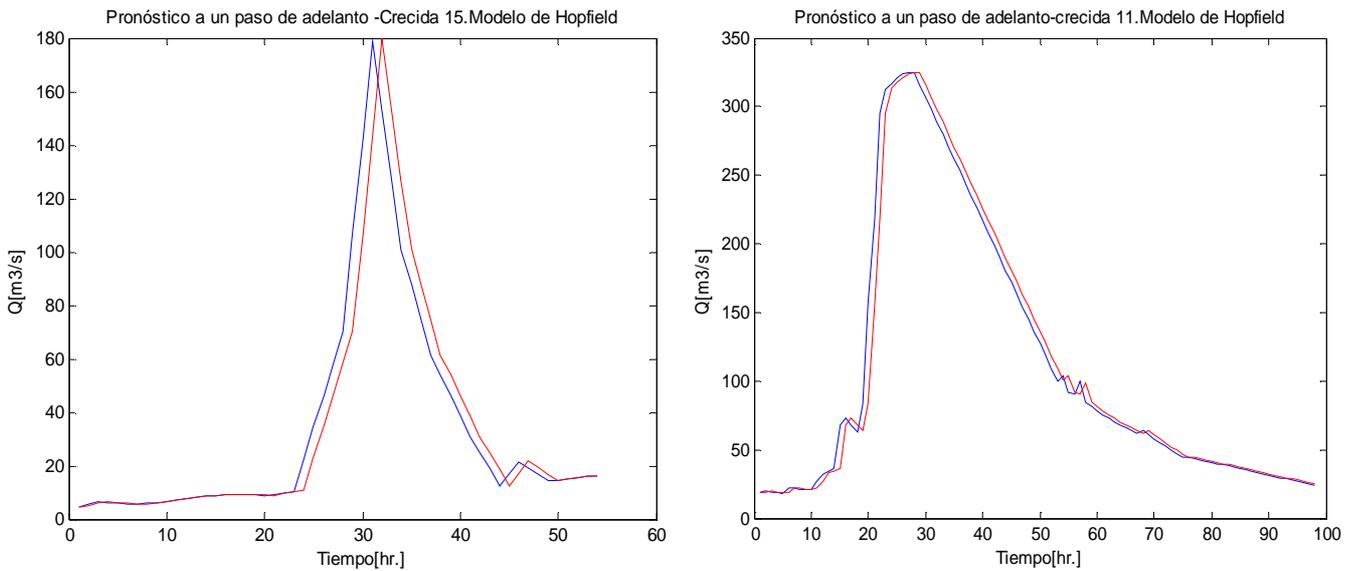
este que se produce en el peak, hasta que encuentra un valor negativo y bruscamente pasa a la zona de receso.

Se aprecia además, que existe un desfase entre máximos de 1 hora, lo cual hace que el modelo de Hopfield no sea competitivo a este nivel con la eficiencia de modelos de redes de pronóstico (Elman y prealimentadas) con entrenamientos de tipo unitario. Sin embargo, las diferencias entre los valores máximos observados y pronosticados a su tiempo de ocurrencia son en promedio de un 1% y al tiempo real de ocurrencia de un 23%. El primer valor se encuentra al nivel de una red de Elman entrenada unitariamente, en consecuencia el pronóstico en valor de los caudales máximos es muy bueno, pero, el segundo valor es grande en comparación a cualquier modelo neuronal de pronóstico estudiado, por lo tanto, la calidad del caudal pronosticado es mala en relación a la obtenida con los demás modelos analizados. En la figura 4.40 se presenta, a modo de ejemplo, el caudal observado y simulado para la crecida 11 y 15.

Tabla 4.57: Errores estándar de validación. Modelo de Hopfield.

Crecida	Duración	HOPFIELD				
		E. Estandar [m ³ /s]	ECM [m ³ /s] ²	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]
1	86	0,002	4,078	0,020	-0,537	1
2	24	0,865	50,928	0,007	-0,147	1
3	18	0,060	210,454	0,006	-0,123	1
4	45	0,135	0,903	0,031	-0,082	1
5	23	-0,865	55,652	0,006	-0,153	1
6	139	0,108	0,128	0,016	-0,035	1
7	32	0,111	4,494	0,024	-0,344	1
8	38	-0,057	18,587	0,016	-0,773	1
9	17	0,050	2,764	0,024	-0,670	1
10	19	-0,040	2,284	0,016	-0,770	1
11	101	0,016	197,182	0,000	-0,001	1
12	77	0,025	3,365	0,017	-0,431	1
13	175	0,120	101,849	0,003	-0,300	1
14	163	0,053	11,040	0,005	-0,214	1
15	57	-0,043	137,948	0,003	-0,199	1
16	66	0,275	31,454	0,007	-0,330	1
17	29	0,185	1,373	0,034	-0,166	1
18	29	0,059	9,105	0,026	-0,664	1
19	21	-0,040	0,454	0,003	-0,330	1
20	17	0,110	31,146	0,023	-0,023	1
21	34	0,207	3,341	0,015	-0,023	1
22	24	0,201	10,745	0,029	-0,422	1
23	120	0,133	15,947	0,004	-0,052	1
24	25	0,272	167,757	0,004	-0,163	1
25	106	0,103	18,416	0,004	-0,160	1
26	19	-0,140	8,538	0,002	-0,120	1
27	57	0,513	1,856	0,014	-0,164	1
Ponderado		0,099	42,869	0,010	-0,234	1

Figura 4.40: Pronósticos a un paso de adelanto. Modelo de Hopfield.



4.3.2.6 Resumen y Análisis de modelos de pronóstico a un paso de adelanto

Esta sección tiene por objetivo, situar al lector en los desarrollos de pronósticos realizados hasta el momento, a fin de continuar con pronósticos a varios pasos de adelanto en la cuenca del Mapocho en Los Almendros.

Se comparan modelos de redes neuronales recurrentes de Elman y de redes prealimentadas a nivel de entrenamiento unitario y continuo, como también el modelo neuronal de Hopfield y modelos paramétricos ARMAX y RARMAX [na,nb,nk]. Para ello se consideran los mejores modelos de pronóstico de cada tipo analizado. En el ámbito de los modelos neuronales con entrenamientos por crecidas se considera como set de datos de entrenamiento la crecida 11, dado que se ha demostrado su mayor capacidad de generalización.

Los parámetros de comparación utilizados son el ECM ponderado, el ECM de validación por crecida y los índices de error en el pronóstico de máximos. Los cuales se presentan en las tablas 4.58.a ,4.58.b y 4.58.c respectivamente.

En la tabla 4.58.a se observan las ventajas a nivel de errores ponderados de realizar entrenamientos de tipo unitario en modelos neuronales. Para este tipo de entrenamiento y además, para el de tipo continuo, el menor error se obtiene para un modelo NEWELM, de arquitectura 3-5-5-1 ó 3-7-7-1 respectivamente.

En general, al igual que en la cuenca del Estero Arrayán, los modelos paramétricos presentan errores superiores a los de redes neuronales y sin una clara tendencia “común” a entregar

menores errores con entrenamientos de tipo continuo o unitario. Se deducen mejores pronósticos a nivel continuo para modelos RARMAX y a nivel unitario para modelos ARMAX, en éste último dicha aseveración no es netamente concluyente, dado que la diferencia de error ponderado para ambos tipos de entrenamiento es pequeña.

Se observa además, que el modelo de Hopfield entrega una eficiencia similar en términos de ECM a un modelo prealimentado entrenado continuamente.

Tabla 4.58.a: Errores ponderados según modelo de pronóstico.

ECM PONDERADO [m ³ /s] ²	Entrenamiento	
	Unitario	Continuo
NEWELM	28,425	33,010
NEWFF	32,435	43,983
ARMAX	34,234	34,769
RARMAX	70,012	46,657
HOPFIELD	42,869	

En la tabla 4.58.b se presentan los errores de validación obtenidos para cada crecida, los cuales se pueden apreciar gráficamente en la figura 4.41 y 4.42. Se deduce que los modelos neuronales de pronóstico, no de identificación (Hopfield) presentan las mejores validaciones y por lo tanto mejores pronósticos, destacándose las claras ventajas, también a este nivel, de un modelo recurrente NEWELM 3-5-5-1 entrenado unitariamente.

Tabla 4.58.b: Errores de validación por crecida según modelo de pronóstico.

Crecida	ECM de validación [m ³ /s] ²						
	Unitario		Continuo		ARMAX [3,5,1]	RARMAX[2,1,3]	HOPFIELD
	NEWELM 3-5-5-1	NEWFF 9-5-5-1	NEWELM 3-7-7-1	NEWFF 9-5-5-1			
1	3,952	3,853	4,242	5,388	24,375	9,873	4,078
2	43,459	35,333	47,231	54,458	43,830	858,613	50,928
3	118,402	105,474	116,637	153,115	207,011	425,430	258,454
4	2,314	2,307	0,709	0,672	3,560	2,419	0,903
5	41,544	37,470	44,542	44,665	41,944	99,926	55,652
6	4,950	4,119	3,313	3,612	0,373	2,063	0,128
7	2,821	3,397	2,533	5,302	13,965	8,060	4,494
8	15,025	14,646	15,961	16,809	22,818	16,990	18,587
9	3,340	1,738	3,133	3,653	0,111	9,701	2,764
10	2,965	3,822	2,181	4,213	5,578	15,872	2,284
11	67,406	70,439	166,272	366,880	83,396	186,286	197,182
12	2,286	2,530	2,422	3,629	3,734	4,945	3,365
13	87,808	109,820	78,030	102,881	103,400	109,238	101,849
14	10,214	12,828	8,105	14,649	9,259	8,846	11,040
15	64,008	65,055	81,112	95,857	92,589	222,468	137,948
16	31,101	32,290	28,620	29,344	30,859	59,680	31,454
17	0,624	1,059	0,705	1,646	3,070	3,200	1,373
18	7,201	8,214	8,059	9,248	2,861	72,023	9,105
19	0,501	0,449	0,422	0,417	3,824	4,711	0,454
20	21,701	23,452	22,397	24,254	1,387	188,097	31,146
21	4,752	5,686	4,473	4,505	2,226	22,272	3,341
22	11,172	12,152	10,301	9,547	3,230	79,357	10,745
23	15,027	17,348	13,765	19,573	14,981	20,729	15,947
24	112,781	135,464	100,373	122,985	156,095	256,512	167,757
25	18,643	30,981	14,810	16,242	17,164	23,472	18,416
26	36,706	31,657	32,998	25,147	7,162	219,331	8,538
27	5,020	4,800	4,850	4,909	4,438	17,050	1,856

En la tabla 4.58.c, correspondiente al valor de los índices de error de pronóstico de máximos para modelos recurrentes de Elman, prealimentados y de Hopfield, se destacan también las ventajas de un modelo NEWELM 3-7-7-1. Cabe destacar que el índice TMEM ha sido calculado como un promedio del error obtenido en el pronóstico de máximos para cada crecida.

Los tres modelos neuronales en promedio presentan sobrestimaciones de los caudales máximos a nivel de entrenamientos unitarios, en mayor medida el modelo NEWELM y el modelo NEWHOP, cuyos valores son equivalentes. En el tiempo de ocurrencia del máximo real, las sobrestimaciones aumentan en este primer modelo y pasan a subestimaciones considerables en un modelo de Hopfield, lo mismo ocurre en el modelo prealimentado pero a un nivel bastante menor. A nivel continuo en cambio, se ocasiona subestimaciones para el modelo de Elman y el prealimentado.

No se puede concluir una tendencia de cada modelo a sobrestimar o subestimar en promedio el valor de los caudales máximos de cada crecida, pero en general, un modelo neuronal recurrente de Elman o uno prealimentado subestima el valor del caudal máximo de cada crecida, no así un modelo de Hopfield que otorga sobrestimaciones, producto de la forma en que se realizan sus estimaciones (punto 4.3.2.5).

La gran desventaja del modelo de Hopfield salta a la vista y es el tiempo de retraso en los caudales máximos pronosticados.

Tabla 4.58.c: Índice de error de pronóstico de máximos ponderado, según modelo de pronóstico.

ÍNDICES DE ERROR	Entrenamiento					
	Unitario			Continuo		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]
NEWELM	0,010	0,091	0,374	-0,067	-0,114	0,423
NEWFF	0,062	-0,168	0,551	-0,126	-0,188	0,601
NEWHOP	0,010	-0,234	1			

Figura 4.41: Análisis gráfico de validación de crecidas, según modelo de pronóstico.

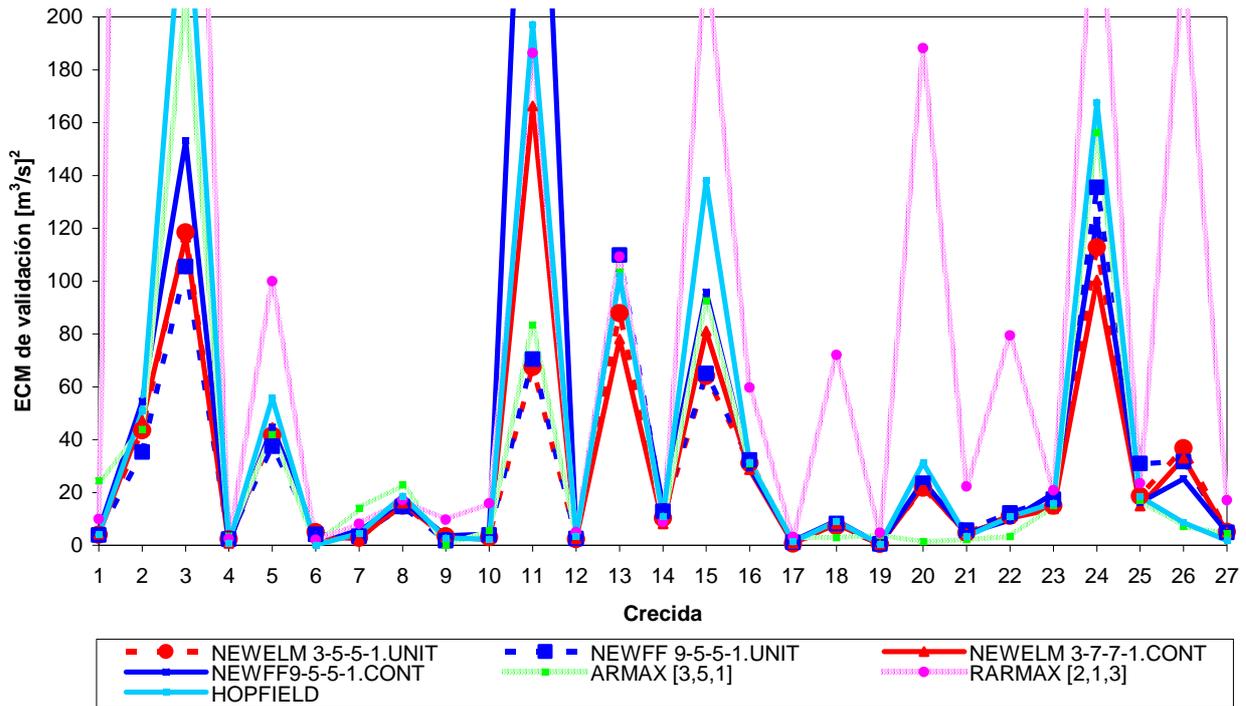
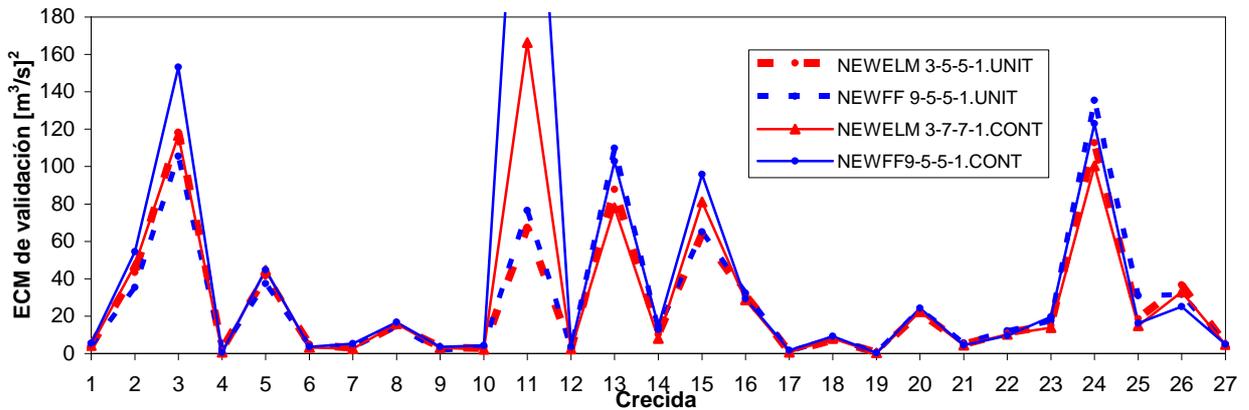


Figura 4.42: Análisis gráfico de validación de crecidas, según modelo de pronóstico.



Dados los análisis anteriores, se deduce que el mejor modelo de pronóstico tanto en calidad como en valor es un modelo NEWELM 3-5-5-1 entrenado con los datos de la crecida número 11. A este nivel se obtiene un mejor pronóstico de dicha crecida usando este último modelo y no un modelo NEWFF, dicho comportamiento hace predecir que para crecidas con mayor amplitud de frecuencias y características de fenómenos extremos se obtendrán mejores pronósticos que a nivel de entrenamiento continuo con cualquier tipo de red.

A nivel de detalles y de acuerdo a los errores de pronóstico, con un modelo de Elman entrenado unitariamente (crecida 11), los errores son menores mayormente en las crecidas intermedias, un 60% presenta mejores validaciones con este modelo respecto del equivalente NEWFF. Esto

puede no apreciarse gráficamente debido a que en muchos casos dichas mejoras no superan 1 orden de magnitud.

Las razones expuestas en el punto 4.3.2.5, en conjunto con el análisis de errores de validación gráfico presentado precedentemente, indican que realizar pronósticos a más pasos de adelanto con un modelo neuronal de identificación de Hopfield puede resultar eficiente y comparable con modelos neuronales netamente de pronóstico, pero con la consideración de que si bien se encuentran buenos pronósticos de caudal en algunas crecidas y en otros regulares, existen algunas de ellas que arrojan errores significativamente mayores a los encontrados para una red neuronal de pronóstico. En el capítulo 6 se recomiendan algunas alternativas para el uso de la red de Hopfield bastante útiles en el pronóstico de crecidas, tanto como modelo de pronóstico y como un proceso anterior e incorporado en la creación de un modelo de pronóstico con redes recurrentes de Elman.

De la misma manera observando los errores de validación obtenidos para un modelo RARMAX, se distingue el gran error de pronosticar algunas crecidas con respecto a modelos ARMAX y de redes neuronales. Por lo tanto, no se recomienda su uso a menos de que se contenga un gran set de datos de crecidas, de tal manera que el modelo pueda procesar la mayor cantidad de información durante el entrenamiento, fijando las condiciones iniciales de ajuste de los pesos para posteriores pronósticos.

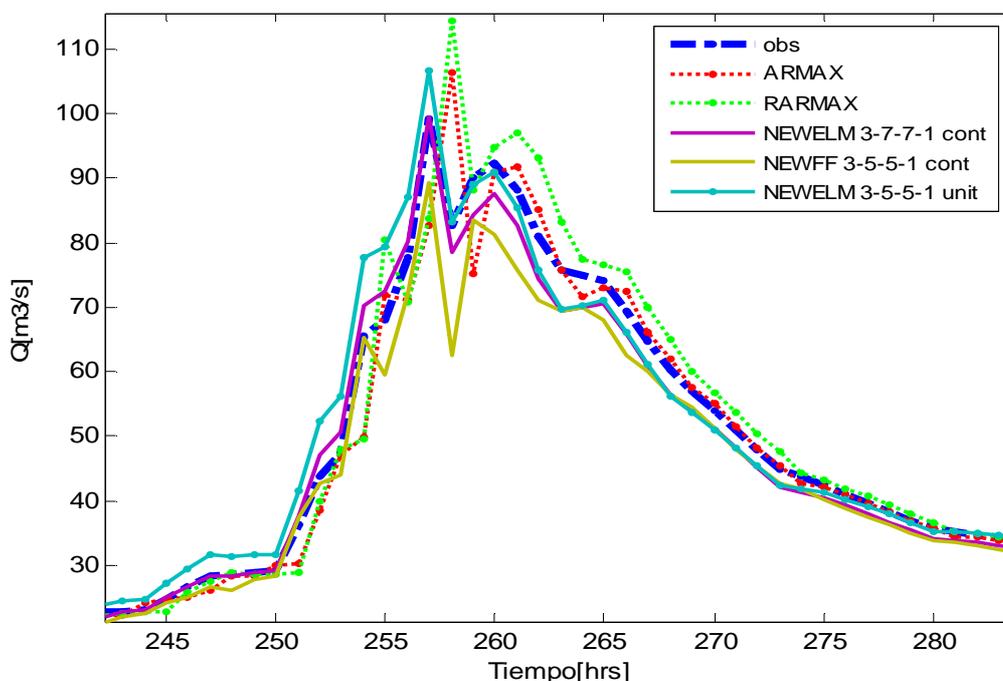
En la figura 4.43 se observa los pronósticos realizados mediante el modelo neuronal recurrente de Elman en comparación a modelos prealimentados y paramétricos, en un sector representativo del comportamiento de los pronósticos en el conjunto de validación.

Se observan los retardos en los pronósticos de caudal producidos por un modelo de tipo paramétrico, a lo que se suma la sobre-estimación de los caudales producida por el modelo RARMAX, lo que no siempre sucede con un modelo ARMAX.

En modelos de redes neuronales, existe una tendencia promedio a subestimar el valor de los caudales máximos, más notoria en modelos NEWFF. En particular, el modelo NEWELM 3-5-5-1 entrenado con los datos de la crecida 11 se encuentra sobrestimando los valores observados al inicio de la crecida y luego subestimando, sin desfase en los tiempos de ocurrencia, en cambio el modelo NEWELM 3-7-7-1 entrenado continuamente pronostica casi con exactitud el caudal máximo y al tiempo en que se produce realmente, aunque pasado el valor máximo tiende a subestimar los caudales en mayor cantidad que el modelo unitario

En términos generales los desfases entre el valor observado y pronosticado son casi nulos y menores para modelos de RNR de Elman en comparación a los que se obtienen para un modelo de redes neuronales prealimentado, y destacadamente para modelos de tipo paramétrico como los indicados.

Figura 4.43: Análisis de Modelos en el conjunto de validación.



4.3.2.7 Pronósticos a varios pasos de adelanto

Se realizan pronósticos de caudales a 4, 8, 12, 16 y 20 horas utilizando modelos neuronales recurrentes de Elman y prealimentados, cuyas arquitecturas son elegidas dependiendo de los pasos de adelanto a los que se quiere pronosticar y de acuerdo a la misma metodología utilizada para cada modelo individualmente.

Al finalizar el apartado se realizan pronósticos con un modelo neuronal de Hopfield, su eficiencia es comparada con los demás modelos neuronales y modelos paramétricos.

En la tabla 4.59 se presentan los modelos neuronales de pronóstico escogidos.

Tabla 4.59: Arquitecturas Neuronales óptimas.

Análisis	t+4		t+8		t+12		t+16		t+20	
	Unitario	Continuo	Unitario	Continuo	Unitario	Continuo	Unitario	Continuo	Unitario	Continuo
Newelm	3-5-5-1	3-5-5-1	3-5-5-1	3-5-5-1	3-5-5-1	3-9-10-1	3-5-5-1	3-10-9-1	3-5-5-1	3-10-9-1
Newff	9-5-5-1	9-7-7-1	9-5-5-1	9-7-7-1	9-5-5-1	9-9-9-1	9-5-5-1	9-10-10-1	9-5-5-1	9-10-10-1

Las variables de entrada son aquellas que definieron los mejores modelos de pronóstico a una hora de adelanto, es decir $P_t * A_t, Q_t$ y \dot{Q} con respecto a t para pronósticos a $t+x$ horas, además de los correspondientes retardos involucrados para un modelo neuronal prealimentado.

En la tabla 4.60 se presentan los errores ponderados obtenidos de dichos pronósticos para modelos neuronales de Elman y prealimentados considerando un análisis continuo. Se observan las ventajas del primer modelo con respecto al segundo.

En la tabla 4.61 se presentan los errores de entrenamiento y ponderados obtenidos de realizar pronósticos de acuerdo al análisis unitario por medio de la crecida 11. Claramente los errores de entrenamiento son menores para una red de tipo recurrente, lo mismo ocurre en los errores ponderados que reflejan los errores de validación obtenidos.

En las figuras 4.44 y 4.45 se aprecia la correlación existente entre los caudales máximos observados y pronosticados para modelos NEWELM entrenados continua y unitariamente. Se aprecian mejores pronósticos con un modelo unitario.

Con el fin de comparar la eficiencia de modelos neuronales con modelos de tipo paramétrico, se utiliza el mejor modelo de estos últimos a nivel de pronósticos a una hora, es decir, un modelo ARMAX [3,5,1](recordar [na,nb,nk]). Los valores del ECM obtenido para distintos tiempos de antelación se presentan en la tabla 4.62. Se observa que dichos valores son mayores que los entregados por un modelo neuronal recurrente de Elman, pero destacándose que son del mismo orden de magnitud. Además, las diferencias entre ambos modelos (ARMAX y neuronales) disminuyen a medida que se realizan pronósticos a más horas de adelanto.

Se infiere que un modelo neuronal recurrente de Elman entrenado con los datos de la crecida 11 arroja pronósticos más certeros a diferentes horas de adelanto que modelos del mismo tipo entrenados continuamente y que modelos prealimentados formados con ambos tipos de entrenamiento, así como también significativamente sobre un modelo ARMAX (de acuerdo al valor del ECM).

En la tabla 4.63 se presentan los caudales máximos observados y pronosticados para cada crecida considerando un modelo NEWELM 3-5-5-1 entrenado con los datos de la crecida 11. Si bien existen marcadas diferencias sobre todo para los pronósticos a 16 y 20 horas, se ha conseguido mejorar considerablemente los pronósticos de crecidas en relación a modelos neuronales prealimentados, figuras 4.44 y 4.45.

En las tablas 4.64 y 4.65 se presenta el valor de los índices de eficiencia de pronóstico de caudal máximo para modelos NEWELM con ambos tipos de entrenamiento. Se observa que existen

subestimaciones de caudal de hasta un 6 % y un retardo que en promedio es de 7 hrs. para obtener pronósticos hasta 20 hrs. de adelanto mediante un análisis unitario. Al igual que en el caso de los errores ponderados, los índices más bajos se obtienen para un modelo NEWELM 3-5-5-1 entrenado con la crecida 11.

Tabla 4.60: Errores Ponderados Modelos Newelm y Newff. Análisis Continuo.

Modelo (Continuo)	ECM Ponderado $[m^3/s]^2$				
	t+4	t+8	t+12	t+16	t+20
Newelm 3-7-7-1	249,2	867,4	1010,5	1230,1	1653,2
Newff 9-5-5-1	323,2	930,6	1250,1	1468,3	1730,2

Tabla 4.61: Errores de entrenamiento y Ponderados Modelos NEWELM y NEWFF. Análisis unitario.

Pronóstico	ECM Entrenamiento $[m^3/s]^2$		ECM Ponderado $[m^3/s]^2$	
	Newelm	Newff	Newelm	Newff
t+4	926,1	1133,1	235,6	257,0
t+8	997,3	1378,2	857,3	932,9
t+12	956,8	1861,2	998,6	1226,4
t+16	926,1	3390,3	1217,3	1336,7
t+20	1256,6	3071,8	1568,9	1727,2

Tabla 4.62: Errores Ponderado Modelo ARMAX.

modelo ARMAX (unitario)	Error Ponderado $[m^3/s]^2$				
	t+4	t+8	t+12	t+16	t+20
	352,123	945,500	1227,751	1459,520	1631,640

Figura 4.44: Pronóstico de caudal máximo a t+x horas. Análisis Continuo NEWELM.

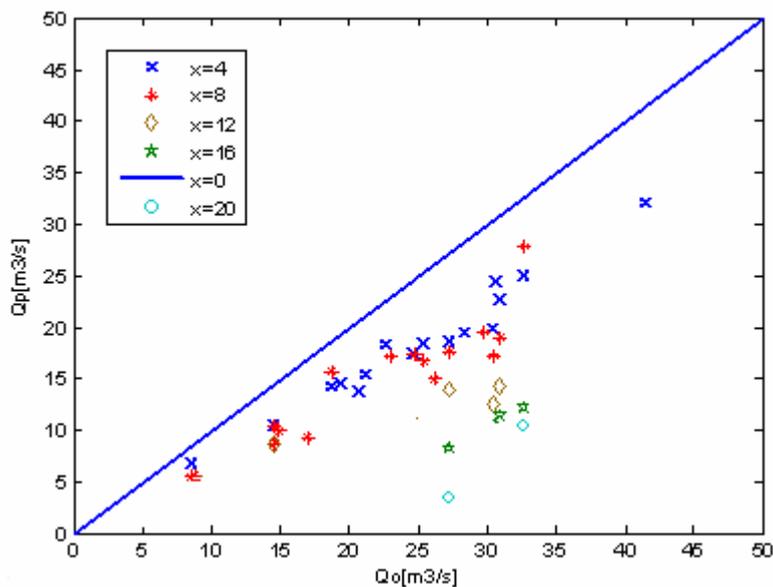


Figura 4.45: Pronóstico de caudal máximo a t+x horas. Análisis Unitario NEWELM 3-5-5-1.

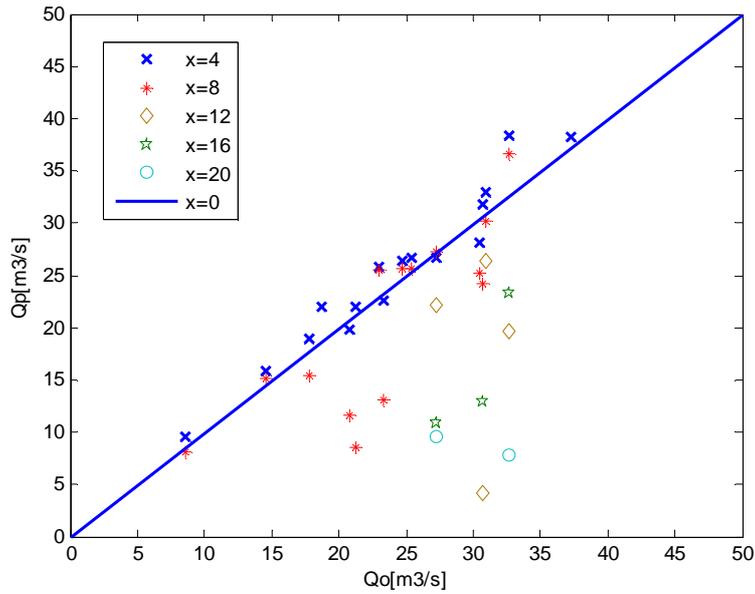


Tabla 4.63: Caudal máximo Observado y Pronosticado para cada crecida. Análisis unitario.

Crecida	Q máx Observado [m³/s]	Q máx simulado [m³/s]				
		t+4	t+8	t+12	t+16	t+20
1	27,2	26,6	27,2	22,1	10,9	9,6
2	74,7	79,7				
3	120,0	138,1				
4	18,7	22,0				
5	90,0	103,3				
6	32,7	38,4	36,6	19,7	23,3	7,8
7	25,4	26,7	25,6			
8	30,5	28,1	25,2			
9	17,8	18,9	15,5			
10	23,0	25,8	25,5			
11	325,0	362,3	354,0			
12	30,9	33,0	30,2	26,3		
13	230,0	248,6	231,6	303,1	289,6	225,2
14	99,3	109,9	109,1	111,0		
15	178,9	206,5	204,3			
16	69,0	70,1	70,4	32,5	33,6	
17	14,6	15,8	15,2			
18	20,8	19,9	11,6			
19	8,6	9,6	8,2			
20	23,4	22,5	13,1			
21	24,7	26,4	25,6			
22	21,3	22,0	8,6			
23	82,2	82,4	86,1	57,3	45,7	34,7
24	108,0	122,7	115,2			
25	105,0	123,5	112,9	55,2	68,2	45,3
26	37,3	38,2				
27	30,8	31,8	24,2	4,1	12,9	

Tabla 4.64: Índices de Pronóstico de caudal máximo. Análisis Continuo NEWELM.

ÍNDICE	
EMEM [tanto por 1]	-0,210
EMRMOBS [tanto por 1]	-0,350
TMEM [hr.]	7,123

Tabla 4.65: Índices de Pronóstico de caudal máximo. Análisis Unitario NEWELM 3-5-5-1.

ÍNDICE	
EMEM [tanto por 1]	-0,058
EMRMOBS [tanto por 1]	-0,239
TMEM [hr.]	6,815

A continuación se realizan pronósticos utilizando la red neuronal de Hopfield. Se utiliza la misma metodología expuesta en la cuenca del Estero Arrayán en La Montosa, en el punto 4.3.1.7. El modelo de pronóstico usado posee dos etapas, un proceso de identificación de \dot{Q}_{t+x} y uno de estimación $((\dot{Q}_{t+x} + Q_t) * x = Q_{t+x})$.

En la tabla 4.66 se presentan los errores de pronósticos de caudal máximo a 8 hrs. de adelanto y el error ponderado cometido considerando todos los pronósticos realizados (a 4, 8, 12, 16 y 20hrs. de adelanto), exceptuando el valor del ECM que se indica a para el pronóstico a 20 hrs.

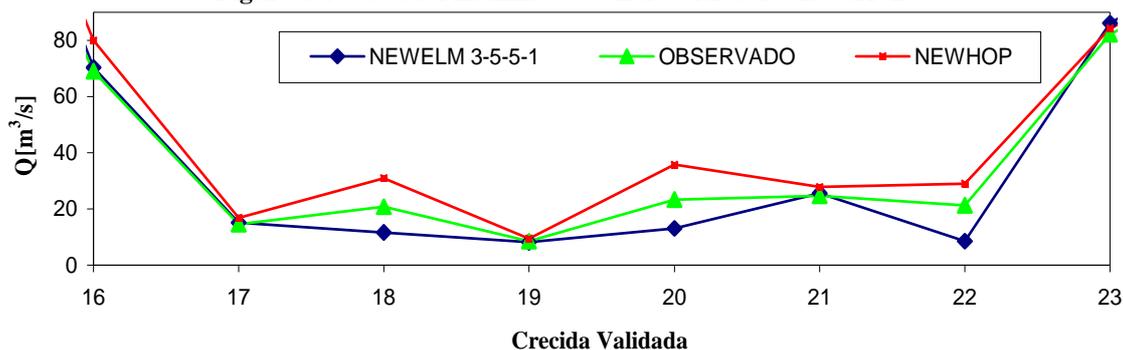
Se observa que el error de pronóstico se encuentra al nivel de un modelo neuronal prealimentado entrenado continuamente. En general, se producen sobre-estimaciones del caudal máximo, aunque para el tiempo de ocurrencia real de dicho valor se produce una subestimación, las razones de este comportamiento son las que se indican en el punto 4.3.1.5 y que tienen relación con la estimación del caudal máximo. La gran desventaja del método en relación a modelos neuronales recurrentes de Elman radica en el valor de los índices EMRMOBS y TMEM, lo que se debe a la manera de estimación del valor pronosticado.

En la figura 4.46 se observa el caudal máximo pronosticado para algunas crecidas (16 a 23), por cada una de las redes recurrentes estudiadas, la red de Elman (con entrenamiento unitario) y la red de Hopfield, los valores obtenidos para cada crecida son comparados con los observados. Se observa claramente la tendencia de la red de Elman a subestimar en comparación a la de Hopfield a sobre-estimar. La gran ventaja que presenta el modelo de Hopfield radica en que no necesita un gran set de datos de validación para pronosticar a muchos pasos de adelanto, no así la red de Elman o prealimentada, que sí requieren suficientes datos para obtener un pronóstico certero.

Tabla 4.66: Índices de Pronóstico Ponderados de caudal máximo. Modelo NEWHOP.

ERROR a 8 hrs.		ERROR PROMEDIO.	
ECM[m ³ /s] ²	934,345	ECM[m ³ /s] ²	1721,340
EMEM [tanto por 1]	0,029	EMEM [tanto por 1]	0,061
EMRMOBS [tanto por 1]	-0,431	EMRMOBS [tanto por 1]	-0,765
TMEM [hr.]	2	TMEM [hr.]	15,210

Figura 4.46: Caudal máximo a t+8 horas. Modelos Recurrentes.



4.4 Aplicación a Crecidas de deshielo

4.4.1 Cuenca del Estero Arrayán en La Montosa

Se cuenta con un total de 11 crecidas (10.912 datos) cuyas características se presentan en la tabla 4.67. Al observar los desfases entre la variable temperatura y caudal se deduce que existe un desfase de aprox. 7 horas entre la ocurrencia de los valores máximos, de tal manera que el fenómeno no responde rápidamente ante aumentos de temperatura y por lo tanto, un modelo neuronal de pronósticos de deshielo en esta cuenca resulta adecuado con una sola capa oculta.

Tabla 4.67: Caracterización de Crecidas.

CRECIDA	INICIO	TERMINO	DURACIÓN [HR]	MEDIA [m3/s]	DESV ESTANDAR [m3/s]	Q Máx [m3/s]	Q Mín [m3/s]	Pp ACUMULADA [mm]	T°[°C]
1	13-12-1984 16:00	04-01-1985 0:00	513	4,8	0,5	6,0	3,7	0,0	20,0
2	23-10-1986 22:00	30-12-1986 3:00	1614	4,6	1,6	9,9	2,1	43,1	18,9
3	24-09-1987 19:00	29-01-1988 9:00	3039	7,3	2,9	17,4	3,5	84,8	19,0
4	05-11-1989 22:00	28-11-1989 3:00	534	3,8	0,5	5,1	2,9	0,0	19,1
5	21-10-1991 20:00	26-12-1991 1:00	1566	4,6	1,1	7,4	2,8	10,5	17,6
6	12-10-1992 9:00	01-12-1992 21:00	1213	3,9	0,9	7,3	2,6	11,9	17,1
7	09-11-1994 11:00	25-11-1994 5:00	379	3,2	0,3	4,1	2,7	0,0	20,0
8	18-09-1997 21:00	23-09-1997 0:00	100	3,2	0,3	4,2	2,8	15,0	14,9
9	13-10-1997 21:00	04-11-1997 13:00	521	3,8	0,5	5,4	2,5	43,0	14,3
10	02-12-1997 14:00	25-01-1998 3:00	1286	6,3	1,1	8,8	3,6	46,6	21,4
11	12-11-1999 4:00	18-11-1999 6:00	147	3,4	0,4	4,1	2,5	0,0	19,4

Se realiza un análisis espectral de las crecidas por medio de la transformada de Fourier, lo cual queda representado en la figura 4.47 y 4.48. Se observa que las crecidas que mayor amplitud de frecuencia presentan son las crecidas número 2, 3, 10, 5, 6 y que las que menor amplitud y menor variabilidad frecuencial presentan son las crecidas número 8, 11, 7 y 4.

Figura 4.47: Transformada de Fourier de la serie de caudales –tiempo.

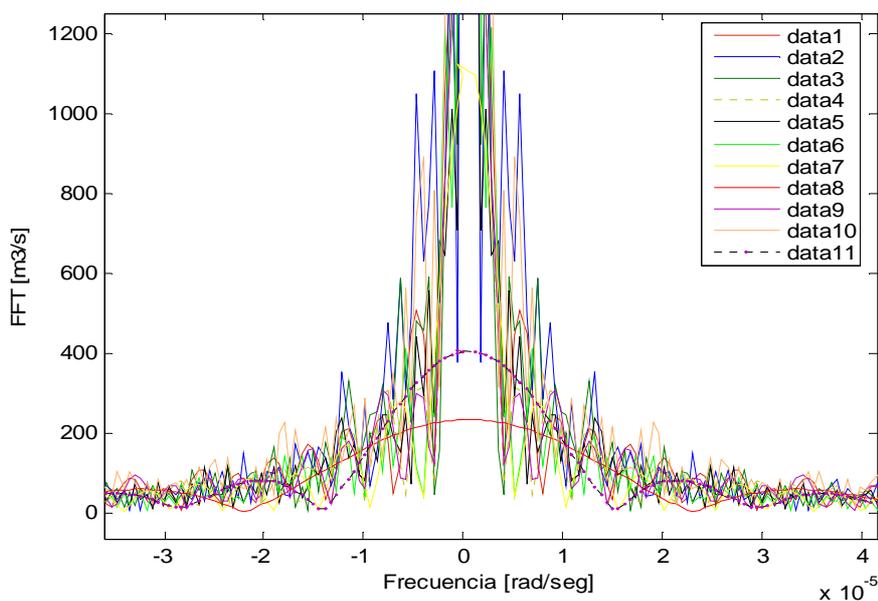
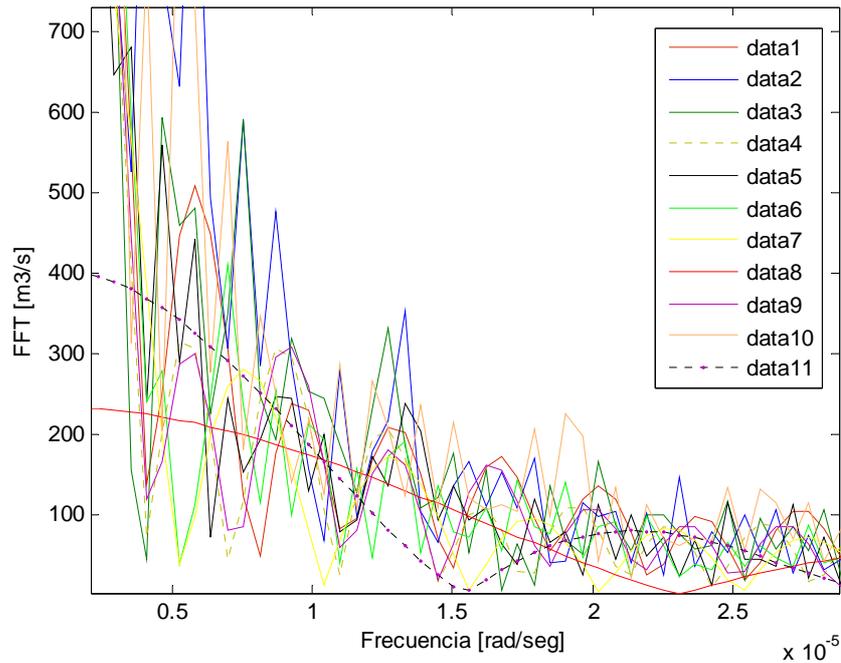


Figura 4.48: Transformada de Fourier de la serie de caudales –tiempo.



4.4.1.1 Pronósticos mediante Redes Neuronales - Entrenamiento Unitario

Primeramente y en este apartado se realizan pronósticos a un paso de adelanto.

De acuerdo al planteamiento general, cada crecida es usada como set de datos de entrenamiento, formando un modelo neuronal que es validado con las crecidas restantes, por lo tanto, se obtienen 11 modelos independientes. Esto crea una matriz de errores en que el error cuadrático medio de validación de la crecida de entrenamiento es el mismo error que se obtiene de su entrenamiento.

4.4.1.1.1 Pronósticos mediante Redes Neuronales Recurrentes de Elman

Las configuraciones planteadas involucran como parámetros de entrada Q_t , Q_{t-23} o Q_{t-24} , T_{t+1-D} o T_{t-D} , y \dot{Q}_t , donde D es el desfase existente entre la ocurrencia de la temperatura máxima y el caudal máximo. El uso de estas variables se justifica de acuerdo a los siguientes puntos:

- Pronóstico de caudales => Comportamiento de la variable Q_t .
- Comportamiento cíclico => Comportamiento de la variable Q_{t-23} o Q_{t-24} .
- Temperatura, causa la salida del derretimiento en el punto de salida de la cuenca (T_{t+1-D} o T_{t-D}).

- Gradiente de caudales, contribuye a dar representatividad al efecto producido por las curvas de ascenso, máximo y descenso de caudal en el sentido de derivada positiva (ascenso), nula (peak), negativa (descenso).

Se estudian las siguientes configuraciones de entrada:

- Q_t, Q_{t-23}, T_{t+1-D} .(Configuración 0).
- $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$ (configuración 1).
- $Q_t, Q_{t-24}, T_{t-D}, \dot{Q}_t$ (configuración 2).

A. Configuración de entrada 0

Se modelan diferentes arquitecturas neuronales, con 3 neuronas en la capa de entrada dadas las variables indicadas, el número de neuronas de la capa oculta es variable, de 9 a 15 y 1 neurona de salida (pronóstico).

En la figura 4.49 se presenta el comportamiento del error ponderado promedio general, error entendido como el promedio de los errores ponderados de validación de todos los modelos (formados a partir de la crecida de entrenamiento) en función del número de neuronas en la capa ocultas y de las épocas de entrenamiento. Se desprende que 10 neuronas en la capa oculta constituye la envolvente inferior de los errores, se escoge 2000 épocas como número óptimo de épocas de entrenamiento.

En la figura 4.50 se presenta el análisis de los errores de validación al considerar la crecida 2 como set de datos de entrenamiento. Se observa al igual que para un modelo general que el óptimo es considerar 10 neuronas en cada capa oculta y 2000 épocas.

Por lo tanto, se elige un modelo neuronal que considera 3 neuronas en la capa de entrada, 10 neuronas en la capa oculta y una neurona de salida. El número de neuronas de la capa oculta es extensivo a las demás configuraciones, según se explicó en el planteamiento general.

En la tabla 4.68 se presenta la matriz de errores de entrenamiento y validación para cada uno de los modelos (a partir de la crecida de entrenamiento). Se observa que tanto los errores de entrenamiento como de validación son bastante pequeños y en términos generales del orden de $10^{-2} [m^3/s]^2$.

En la tabla 4.69 se observan los errores ponderados que presentan consecuentemente los mismos órdenes de magnitud.

Analizando los errores, se desprende que el análisis de Fourier presenta validez en crecidas de deshielo, donde la señal es más o menos periódica cada 24 hrs. y por lo tanto el comportamiento de Fourier debería ser aún más evidente. Esto queda de manifiesto al observar los errores, ya que las crecidas con mayor amplitud de frecuencia, destacadas en amarillo, son las que presentan mejor capacidad de generalización y recíprocamente para las peores, destacadas en verde.

Figura 4.49: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.

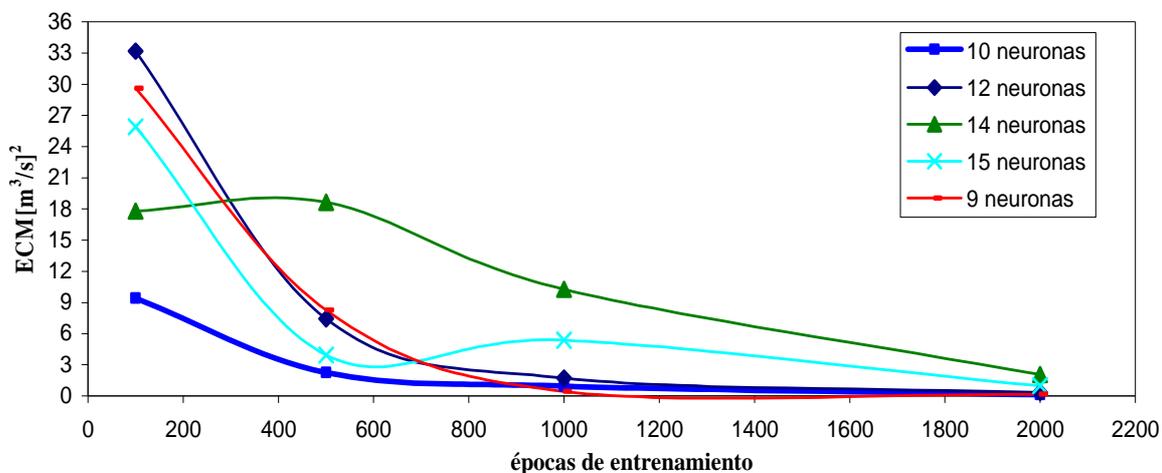


Figura 4.50: ECM validación según número de neuronas en cada capa oculta y épocas de entrenamiento. Crecida 2 de entrenamiento.

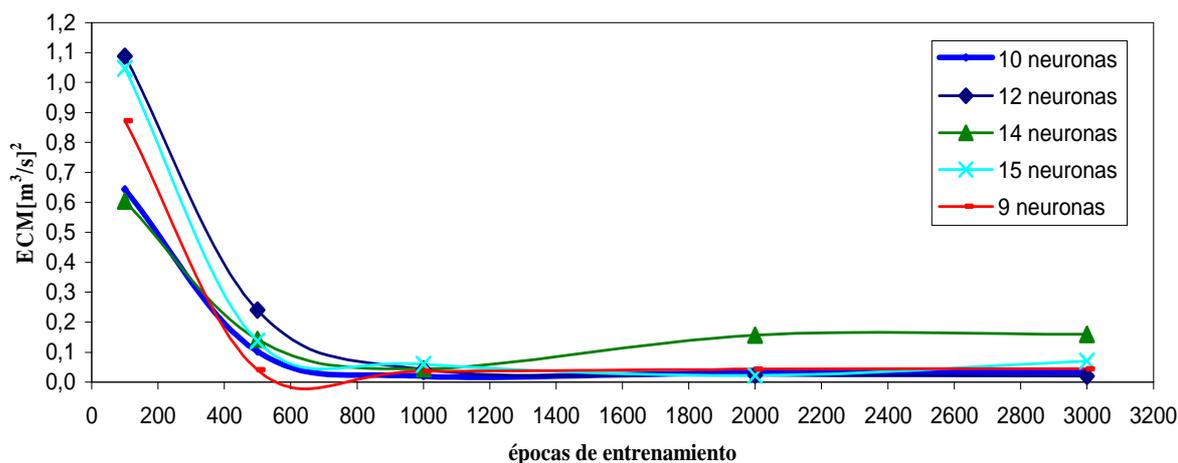


Tabla 4.68: Errores de Entrenamiento y Validación. NEWELM 3-10-1.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Validación [m ³ /s] ²										
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11
1	487	0,007	0,007	0,087	0,311	0,007	0,028	0,037	0,004	0,007	0,012	0,017	0,007
2	1588	0,030	0,024	0,030	0,055	0,016	0,021	0,014	0,011	0,023	0,015	0,040	0,023
3	3013	0,088	0,043	0,076	0,088	0,027	0,036	0,030	0,019	0,060	0,034	0,059	0,041
4	508	0,011	0,029	0,106	0,317	0,011	0,028	0,030	0,008	0,022	0,018	0,055	0,016
5	1540	0,014	0,019	0,039	0,077	0,010	0,014	0,014	0,006	0,020	0,013	0,026	0,011
6	1187	0,023	0,037	0,049	0,084	0,022	0,032	0,023	0,017	0,042	0,027	0,061	0,036
7	353	0,005	0,016	0,130	0,483	0,009	0,029	0,038	0,005	0,011	0,009	0,041	0,009
8	74	0,008	0,019	0,301	1,202	0,013	0,067	0,093	0,007	0,008	0,013	0,060	0,016
9	495	0,021	0,031	0,044	0,110	0,020	0,033	0,024	0,014	0,017	0,021	0,058	0,027
10	1260	0,011	0,010	0,027	0,060	0,008	0,010	0,010	0,005	0,010	0,007	0,011	0,008
11	121	0,008	0,043	0,215	0,697	0,020	0,055	0,072	0,011	0,026	0,022	0,082	0,008

Tabla 4.69: Errores Ponderados. NEWELM 3-10-1.

crecida	Nº de datos	Error Ponderado [m ³ /s] ²
1	487	0,113
2	1588	0,033
3	3013	0,058
4	508	0,123
5	1540	0,037
6	1187	0,051
7	353	0,172
8	74	0,416
9	495	0,056
10	1260	0,026
11	121	0,260

B. Configuración de entrada 1

Además de las variables indicadas para la configuración 0, se agrega como variable de entrada \dot{Q}_i , considerando la misma arquitectura con la diferencia de que se tienen 4 neuronas de entrada.

En la tabla 4.70 se presenta la matriz de errores con cada una de las crecidas de entrenamiento que definen los modelos y las validaciones respectivas de cada crecida en cada modelo. Se observa que tanto los errores de entrenamiento como de validación son más pequeños que para la configuración anterior, en términos generales del orden de 10^{-2} [m³/s]². Los errores ponderados, que se observan en la tabla 4.71, presentan disminuciones y los mismos órdenes de magnitud con respecto a la primera configuración.

En general todas las crecidas son bien pronosticadas a un paso de adelanto, en particular se observa que las crecidas nº 7, 9 y 11 presentan errores de pronóstico mínimos con cada uno de los modelos. Se observa que los menores errores de validación se encuentran en un rango de 10^{-3} [m³/s]² a 10^{-2} [m³/s]².

El análisis de Fourier se confirma nuevamente, debido a que al igual que para la configuración precedente, en el análisis de tipo unitario las mejores crecidas desde el punto de vista de servir como datos de entrenamiento son las que poseen mayores amplitudes de frecuencias.

Tabla 4.70: Errores de Entrenamiento y Validación. NEWELM 4-10-1.

crecida	Nº de datos	Error de Entrenamiento [m ³ /s] ²	Error Validación [m ³ /s] ²										
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11
1	487	0,008	0,008	0,025	0,065	0,011	0,013	0,009	0,003	0,011	0,007	0,019	0,008
2	1588	0,017	0,024	0,017	0,034	0,013	0,013	0,011	0,010	0,019	0,012	0,035	0,019
3	3013	0,082	0,060	0,055	0,082	0,033	0,040	0,028	0,026	0,050	0,037	0,083	0,048
4	508	0,011	0,032	0,053	0,151	0,011	0,021	0,025	0,014	0,016	0,022	0,041	0,020
5	1540	0,013	0,024	0,037	0,076	0,012	0,013	0,020	0,010	0,022	0,015	0,027	0,010
6	1187	0,010	0,011	0,029	0,066	0,011	0,014	0,010	0,004	0,016	0,008	0,024	0,010
7	353	0,004	0,012	0,115	0,384	0,013	0,039	0,045	0,004	0,009	0,016	0,042	0,005
8	74	0,003	0,011	0,103	0,438	0,010	0,025	0,029	0,004	0,003	0,006	0,028	0,009
9	495	0,006	0,011	0,024	0,066	0,010	0,012	0,009	0,004	0,014	0,006	0,024	0,010
10	1260	0,018	0,028	0,038	0,083	0,011	0,012	0,023	0,013	0,027	0,018	0,018	0,015
11	121	0,004	0,019	0,268	0,989	0,017	0,073	0,098	0,006	0,012	0,024	0,069	0,004

Tabla 4.71: Errores Ponderados. NEWELM 4-10-10-1.

crecida	Nº de datos	Error Ponderado [m ³ /s] ²
1	487	0,029
2	1588	0,022
3	3013	0,058
4	508	0,065
5	1540	0,037
6	1187	0,031
7	353	0,144
8	74	0,151
9	495	0,030
10	1260	0,039
11	121	0,353

C. Configuración de entrada 2

Lo que se pretende mediante esta configuración es encontrar el límite de las variables, en el sentido de que dado que la red neuronal otorga relevancia a los retardos de las variables de entrada, puede que, dichos retardos no tengan una importancia “instantánea” en cuanto a la periodicidad (24 hrs.) de la variable de salida (en t+1), sino que de la variable de entrada caudal (en t).

En las tablas 4.72 y 4.73 se observa la matriz de errores para esta configuración. Con respecto a los valores presentados se concluye que en general son del mismo orden y variables, existen algunos modelos (crecidas) que mejoran su error ponderado y otros que lo empeoran con respecto a la configuración anterior. En general, la tendencia muestra que la configuración

número 1, que involucra un desfase menos en las variables temperatura y caudal cíclico (Q_{t+24}) respecto de Q_t presenta menores errores en un 55 % de las crecidas contra un 45 % que empeora o permanece estable para esta configuración.

Tabla 4.72: Errores de Entrenamiento y Validación. NEWELM 4-10-1.

crecida	Nº de datos	Error de Entrenamiento $[m^3/s]^2$	Error Validación $[m^3/s]^2$										
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11
1	487	0,008	0,008	0,024	0,065	0,011	0,013	0,008	0,003	0,010	0,006	0,019	0,009
2	1588	0,017	0,028	0,017	0,041	0,015	0,016	0,011	0,011	0,021	0,013	0,035	0,019
3	3013	0,043	0,040	0,023	0,043	0,019	0,022	0,015	0,015	0,020	0,018	0,059	0,028
4	508	0,011	0,043	0,072	0,215	0,011	0,022	0,037	0,018	0,022	0,023	0,044	0,022
5	1540	0,011	0,017	0,030	0,073	0,008	0,011	0,012	0,006	0,013	0,011	0,021	0,007
6	1187	0,012	0,012	0,034	0,074	0,012	0,016	0,012	0,004	0,016	0,010	0,026	0,010
7	353	0,005	0,018	0,288	0,453	0,019	0,077	0,100	0,005	0,017	0,023	0,071	0,010
8	74	0,002	0,006	0,099	0,422	0,008	0,023	0,028	0,002	0,002	0,006	0,023	0,006
9	495	0,004	0,008	0,022	0,069	0,008	0,010	0,007	0,003	0,008	0,004	0,019	0,007
10	1260	0,018	0,043	0,030	0,067	0,010	0,012	0,025	0,018	0,021	0,022	0,018	0,012
11	121	0,005	0,029	0,342	1,255	0,021	0,090	0,122	0,009	0,016	0,030	0,091	0,005

Tabla 4.73: Errores Ponderados. NEWELM 4-10-1.

crecida	Nº de datos	Error Ponderado $[m^3/s]^2$
1	487	0,028
2	1588	0,025
3	3013	0,032
4	508	0,089
5	1540	0,033
6	1187	0,035
7	353	0,205
8	74	0,145
9	495	0,028
10	1260	0,034
11	121	0,449

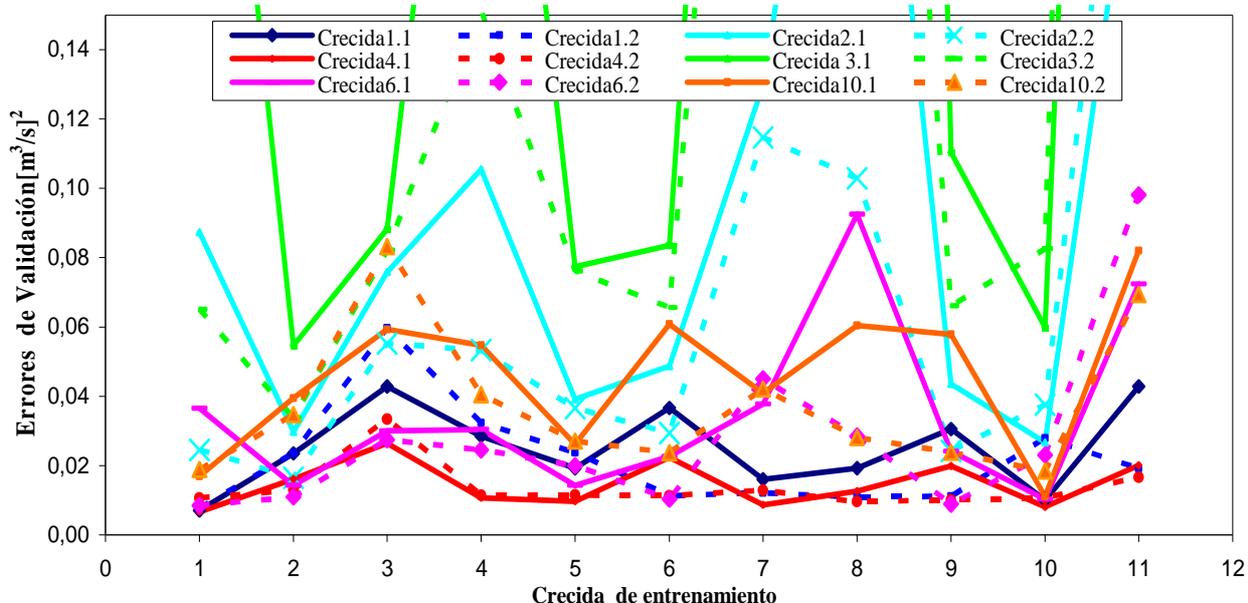
D. Análisis y Comparación de Modelos Recurrentes de Elman

Un 73 % de los modelos NEWELM que consideran la variable gradiente de caudal mejoran sus pronósticos con respecto a un modelo que no lo hace (porcentaje entendido sobre el nº de crecidas que definen los modelos), pero, en general los errores son bastante pequeños, del orden de $10^{-2} [m^3/s]^2$ y para algunos modelos aumenta a $10^{-1} [m^3/s]^2$.

En la figura 4.51 se presenta esquemáticamente el comportamiento de los errores de validación de algunas crecidas, de acuerdo a la crecida utilizada en el entrenamiento de la red, para la configuración 0 y 1. La curva en sí no posee ningún significado más que facilitar el entendimiento del comportamiento de los errores de validación obtenidos según la crecida con la

cual el modelo es entrenado. Líneas punteadas indican un modelo que incorpora el gradiente (Configuración de entrada 1) y línea continua un modelo que no lo hace (Conf. 0); curvas del mismo color indican la misma crecida validada en uno u otro modelo. Se observa que en general se obtienen menores errores para un modelo que incorpora el gradiente.

Figura 4.51: Errores modelos NEWELM 1 con y sin gradiente.



En la tabla 4.74 se presentan los coeficientes de determinación promedio de validación obtenidos para cada uno de los modelos planteados. Se observa que todos los modelos presentan coeficientes sobre 0,8 y en términos generales los mejores se obtienen para el modelo 4-10-1 con la configuración de entrada número 1.

Tabla 4.74: Coeficientes de determinación de modelos NEWELM.

Crecida	r ² validación ponderado		
	Newelm 4-10-1 Conf. Entrada 1	Newelm 3-10-1 Conf. Entrada 0	Newelm 4-10-1 Conf. Entrada 2
1	0,971	0,970	0,983
2	0,972	0,964	0,971
3	0,922	0,924	0,954
4	0,957	0,941	0,948
5	0,972	0,973	0,979
6	0,978	0,941	0,976
7	0,947	0,943	0,868
8	0,952	0,858	0,958
9	0,980	0,951	0,985
10	0,972	0,984	0,971
11	0,882	0,900	0,847

4.4.1.1.2 Pronósticos mediante Redes Neuronales Prealimentados

Se plantean modelos neuronales prealimentados con el objetivo de comparar su eficiencia al realizar pronósticos con respecto a modelos neuronales recurrentes.

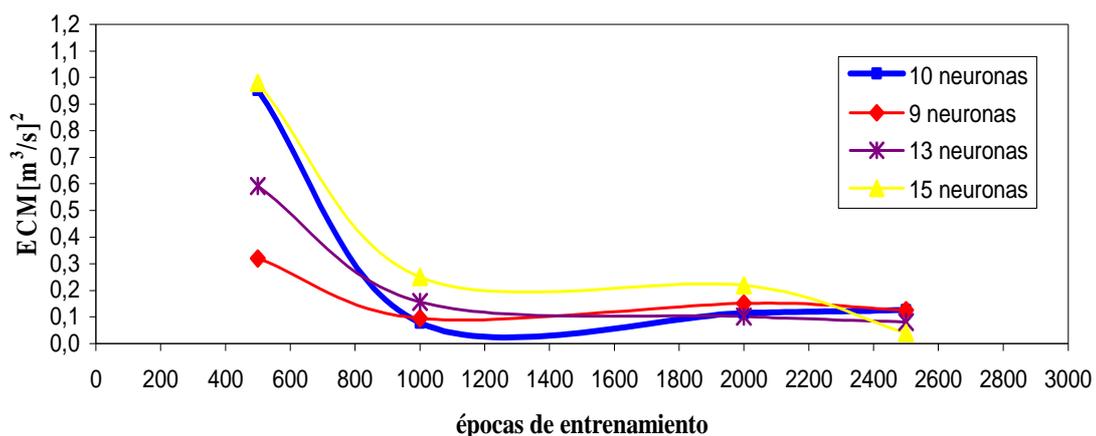
Analizando las variables de entrada representativas del fenómeno de deshielo y los retardos que se deben considerar de dichas variables. La variable caudal se considera retardada hasta en 2 hrs. a partir del instante t y del instante $t-23$, en cuanto a la temperatura, se consideraran 3 retardos a partir del instante $t + 1 - D$, con D el desfase entre los máximos de temperatura y caudal observados (7 horas aprox.) y se consideran modelos con y sin la variable gradiente de caudal.

Luego, se evalúan dos tipos de modelos que poseen la siguiente configuración de entrada:

- 9 neuronas de entrada $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}$ (Conf. 0).
- 10 neuronas de entrada. $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}, \dot{Q}_t$ (Conf. 1).

Se varía el número de neuronas y épocas de entrenamiento de tal manera de encontrar el óptimo, que no necesariamente es el mismo que para una RRS, encontrándose que el número de épocas y de neuronas en las capas ocultas es de 1300 y 10 respectivamente (Figura 4.52), al igual que para una RRS.

Figura 4.52: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.



A. Configuración de entrada 0

Se obtiene la matriz de errores para la configuración indicada, se deduce que el orden de los errores tanto de entrenamiento como de validación son de $10^{-2} [\text{m}^3/\text{s}]^2$ y en algunos pocos casos $10^{-1} [\text{m}^3/\text{s}]^2$. (Tabla 4.75 y 4.76).

Se observa que los peores errores de validación se obtienen al igual en el caso de RNR para aquellos modelos formados a partir de las crecidas con menor amplitud de frecuencia y los mejores modelos siguen siendo los que se obtienen al entrenar con las crecidas con mayores amplitudes de frecuencia.

Con respecto a los errores ponderados, se observa que para los modelos formados a partir del entrenamiento con los datos de la mayoría de las crecidas exceptuando la 1, 6, 8 un modelo de RNR de Elman es mejor.

Tabla 4.75: Matriz de errores modelo NEWFF 9-10-1.

crecida	Nº de datos	Error de Entrenamiento $[\text{m}^3/\text{s}]^2$	Error Validación $[\text{m}^3/\text{s}]^2$										
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11
			487	1588	3013	508	1540	1187	353	74	495	1260	121
1	487	0,015	0,015	0,044	0,096	0,011	0,018	0,018	0,007	0,014	0,011	0,025	0,014
2	1588	0,039	0,027	0,039	0,062	0,018	0,026	0,019	0,013	0,020	0,019	0,045	0,028
3	3013	0,088	0,046	0,082	0,088	0,030	0,044	0,035	0,022	0,048	0,038	0,067	0,046
4	508	0,014	0,036	0,196	0,653	0,014	0,048	0,057	0,010	0,025	0,022	0,080	0,022
5	1540	0,025	0,032	0,069	0,137	0,016	0,025	0,027	0,009	0,025	0,018	0,042	0,013
6	1187	0,020	0,024	0,051	0,091	0,014	0,023	0,020	0,010	0,022	0,018	0,038	0,022
7	353	0,010	0,034	0,267	0,962	0,015	0,062	0,078	0,010	0,022	0,020	0,086	0,022
8	74	0,012	0,032	0,153	0,547	0,022	0,045	0,048	0,014	0,012	0,016	0,057	0,034
9	495	0,013	0,021	0,082	0,251	0,012	0,026	0,027	0,008	0,015	0,013	0,044	0,017
10	1260	0,026	0,023	0,063	0,108	0,016	0,024	0,024	0,010	0,028	0,020	0,026	0,021
11	121	0,010	0,058	0,341	1,156	0,028	0,084	0,112	0,014	0,034	0,029	0,124	0,010

Tabla 4.76: Errores Ponderados modelo NEWFF 9-10-1.

crecida	Nº de datos	Error Ponderado $[\text{m}^3/\text{s}]^2$
1	487	0,044
2	1588	0,038
3	3013	0,062
4	508	0,241
5	1540	0,065
6	1187	0,047
7	353	0,345
8	74	0,201
9	495	0,098
10	1260	0,053
11	121	0,424

B. Configuración de entrada 1

En la tabla 4.77 se presenta la matriz de errores, al igual que para la primera configuración el orden de los errores es de $10^{-2} [\text{m}^3/\text{s}]^2$ y en algunos pocos casos de $10^{-1} [\text{m}^3/\text{s}]^2$.

Se observa una considerable disminución de los errores de entrenamiento y validación, dentro del rango y orden en que se encuentran y por ende de los errores ponderados en los 11 modelos formados (tabla 4.78), con respecto a la primera configuración. Queda de manifiesto también al observar el coeficiente de determinación para cada modelo (Tabla 4.79).

Tabla 4.77: Matriz de errores modelo NEWFF 10-10-1.

crecida	Nº de datos	Error de Entrenamiento $[\text{m}^3/\text{s}]^2$	Error Validación $[\text{m}^3/\text{s}]^2$										
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	Valid 11
1	487	0,009	0,009	0,036	0,102	0,015	0,019	0,011	0,004	0,007	0,007	0,025	0,012
2	1588	0,017	0,016	0,017	0,047	0,010	0,011	0,009	0,006	0,008	0,009	0,023	0,009
3	3013	0,078	0,024	0,053	0,078	0,017	0,025	0,017	0,009	0,023	0,021	0,041	0,020
4	508	0,009	0,037	0,117	0,380	0,009	0,028	0,047	0,013	0,016	0,020	0,046	0,011
5	1540	0,010	0,019	0,028	0,060	0,008	0,010	0,014	0,006	0,010	0,012	0,020	0,006
6	1187	0,010	0,018	0,029	0,056	0,017	0,021	0,010	0,008	0,013	0,014	0,036	0,018
7	353	0,004	0,015	0,146	0,501	0,014	0,040	0,047	0,004	0,011	0,015	0,045	0,008
8	74	0,006	0,020	0,037	0,163	0,017	0,019	0,007	0,007	0,006	0,007	0,039	0,017
9	495	0,007	0,012	0,025	0,074	0,013	0,016	0,008	0,005	0,006	0,007	0,031	0,013
10	1260	0,024	0,026	0,032	0,068	0,012	0,014	0,017	0,008	0,014	0,015	0,024	0,005
11	121	0,005	0,028	0,205	0,697	0,019	0,061	0,080	0,008	0,014	0,023	0,072	0,005

Tabla 4.78: Errores Ponderados modelo NEWFF 10-10-1.

crecida	Nº de datos	Error Ponderado $[\text{m}^3/\text{s}]^2$
1	487	0,043
2	1588	0,023
3	3013	0,044
4	508	0,144
5	1540	0,029
6	1187	0,032
7	353	0,183
8	74	0,062
9	495	0,033
10	1260	0,034
11	121	0,258

Tabla 4.79: Coeficiente de determinación modelos NEWFF.

Crecida	r^2 validación ponderado	
	Newff 9-10-1	Newff 10-10-1
1	0,958	0,976
2	0,969	0,981
3	0,935	0,958
4	0,933	0,940
5	0,969	0,979
6	0,962	0,969
7	0,929	0,935
8	0,734	0,965
9	0,963	0,977
10	0,965	0,975
11	0,948	0,906

4.4.1.1.3 Comparación y Análisis de Modelos Neuronales.

Se esquematiza el comportamiento de los errores de validación y ponderados, así como también de los índices que cuantifican la eficiencia del modelo de pronóstico en el caudal máximo.

De ahora en adelante se entiende por Newelm (Newff) Mod 1., el modelo neuronal recurrente (prealimentado) analizado con las variables de entrada dadas por la configuración 0, y por c/d (s/d) a si además involucra (no) el gradiente.

En la figura 4.53 se observan los errores ponderados de los 4 modelos analizados, la serie en sí unida y representada como una curva no tiene ningún significado físico, sólo es una esquematización de los errores obtenidos.

Se observa, según se desprendió unitariamente que las crecidas 11, 8 y 4 son crecidas que no servirían para realizar algún tipo de pronóstico, puesto que en todos los modelos planteados presentan errores superiores al resto, al contrario las crecidas 2, 3, 5 y 6 validan los pronósticos de crecidas de manera óptima.

En la figura 4.54 se presentan los errores ponderados para cada modelo de pronóstico sin los modelos formados por el entrenamiento de las crecidas 11, 8 y 4. Donde se aprecia con mayor claridad que un modelo de Elman del tipo 4-10-1, con las variables definidas anteriormente es el mejor modelo de pronóstico y sobre- todo cuando es entrenado con los datos de la crecida número 2.

Figura 4.53: Curvas de error ponderado modelos NEWELM y NEWFF.

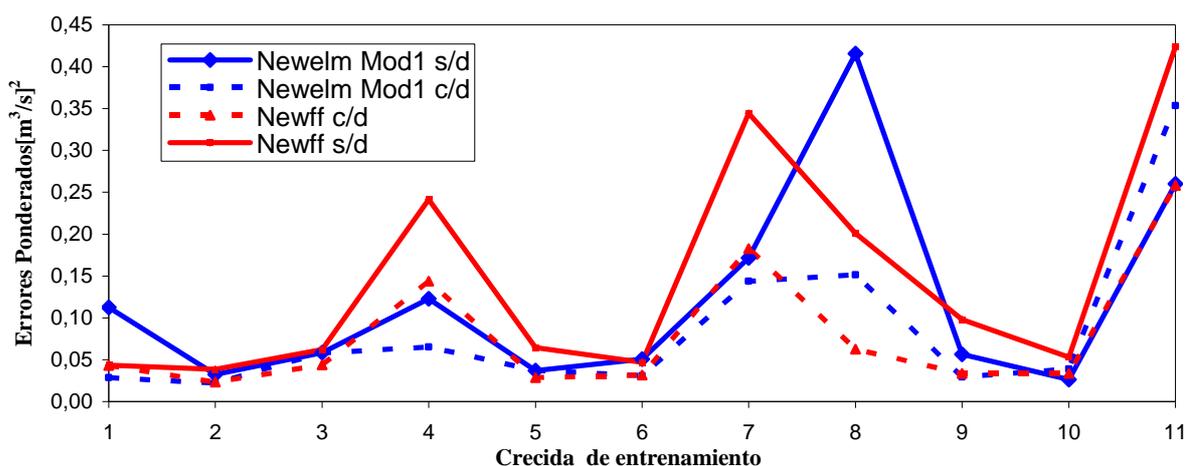
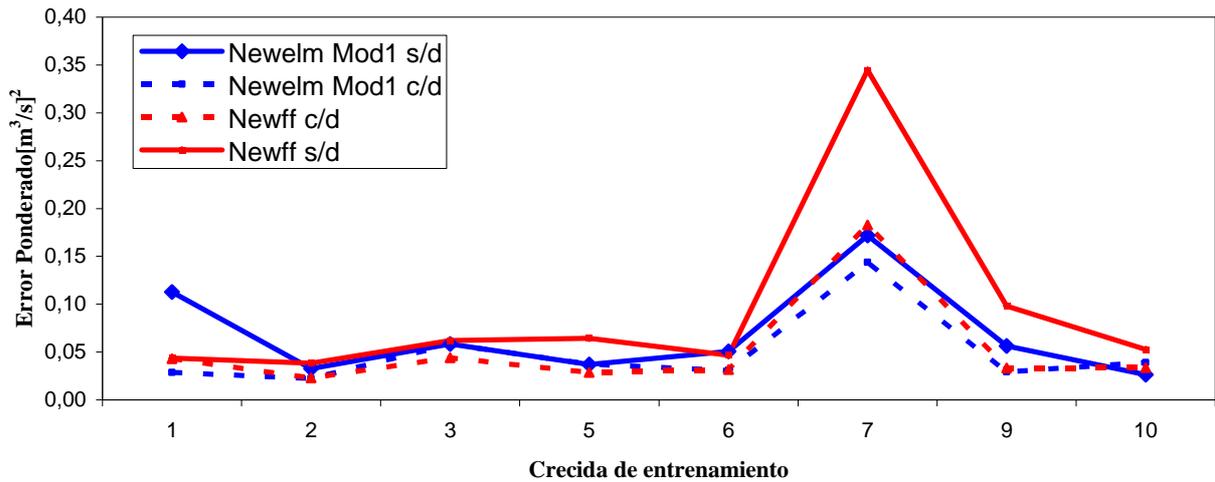


Figura 4.54: Curvas de error ponderado modelos NEWELM y NEWFF.



A continuación se presenta gráficamente el comportamiento de los índices descritos en el capítulo 2 en relación al alcance y pronóstico de máximos de caudal para el entrenamiento unitario. A modo de ejemplificar el comportamiento de dichos índices en validación se evalúan para la crecida 5, en cada modelo definido a través del tipo de red y crecida de entrenamiento.

En las figuras 4.55 a 4.60, se observa que tanto durante el entrenamiento como durante la validación los modelos 1 y 2 que consideran como variable de entrada el gradiente de caudal (sobre todo el n° 1) de NEWELM presentan los más bajos errores en el pronóstico del valor máximo de la crecida, ya que los índices EMEM y EMRMOBS se encuentran mucho más cercanos a un valor nulo que el resto de los modelos planteados (para cualquier crecida que se utilice para el entrenamiento del modelo) y en su mayoría por debajo de dicho valor, por lo que en general subestiman el caudal máximo.

Se observa que para un modelo NEWELM tipo 1 entrenado con los datos de la crecida 2, los índices EMEM y EMRMOBS son aprox. de un 2% de subestimación en el entrenamiento, valores similares se encuentran con un modelo de red prealimentada, pero en la validación de la crecida 5 la diferencia porcentual entre el máximo observado y pronosticado es de aprox. un 1% de sobrestimación en un modelo NEWELM y de un 1% de subestimación en un modelo NEWFF, prefiriéndose por lo tanto el primer modelo. También, existe un mejor pronóstico de caudal a la hora en que se produce el máximo observado para el modelo NEWELM que subestima el máximo en aprox. un 1% mientras que un modelo NEWFF en un 5%.

En cuanto al índice de desfase entre ocurrencia de máximos observados y simulados (TMEM), se observa que gracias a la incorporación de la variable gradiente y a las ventajas de las RRS este valor es en la mayor parte las crecidas de entrenamiento nulo para el entrenamiento (Figura 4.57). En la validación de la crecida 5 no se logra tal efecto, pero se obtiene el más bajo valor de

dicho índice para el modelo NEWELM tipo 1 que incorpora como variable de entrada \dot{Q}_t en relación al resto de los modelos entrenados (Figura 4.60).

El hecho de que existan valores de TMEM del orden de -24 hrs. ó -48 hrs. ó 24 hrs. se debe a que dado que el comportamiento de la variable caudal es cíclico tienden a haber máximos locales, los cuales pueden no diferir mucho del máximo global. Luego, cuando la red realiza el pronóstico de caudales puede ocurrir que subestime algunos valores como que sobreestime otros y por lo tanto el máximo pronosticado ya no es el máximo global de la serie de caudales observada y el desfase es entonces por al menos de 24 hrs. dada la periodicidad de la serie.

Al observar todos los índices llama la atención que el modelo formado a través de la crecida 3 (set de entrenamiento), que se encuentra dentro de las crecidas con una alta capacidad de generalización presente tan altos índices de error al pronosticar el máximo en relación a otros modelos. Esto se debe a la alta desviación estándar y baja pendiente máxima en los datos de dicha crecida, ya que una vez ajustados los pesos de la red y formado el modelo de pronóstico, la red ha tomado en cuenta que se producen bajos cambios de pendiente lo cual no necesariamente ocurre con las demás crecidas, razón por la cual tiende a subestimar en mayor medida los máximos que modelos formados por el entrenamiento con otras crecidas.

Se recomienda utilizar las crecidas 2, 5 y 6 como set de entrenamiento de la RNR de Elman con arquitectura 4-10-1.

Figura 4.55: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.

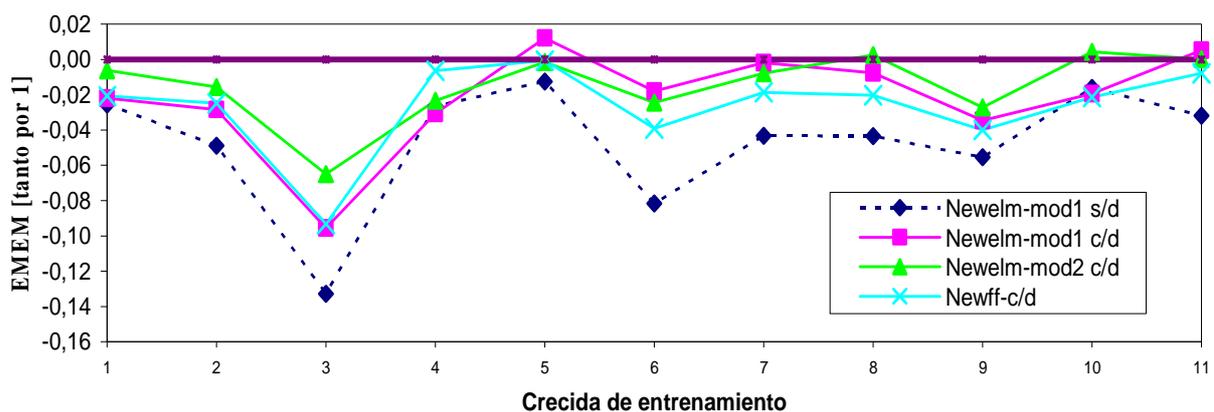


Figura 4.56: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.

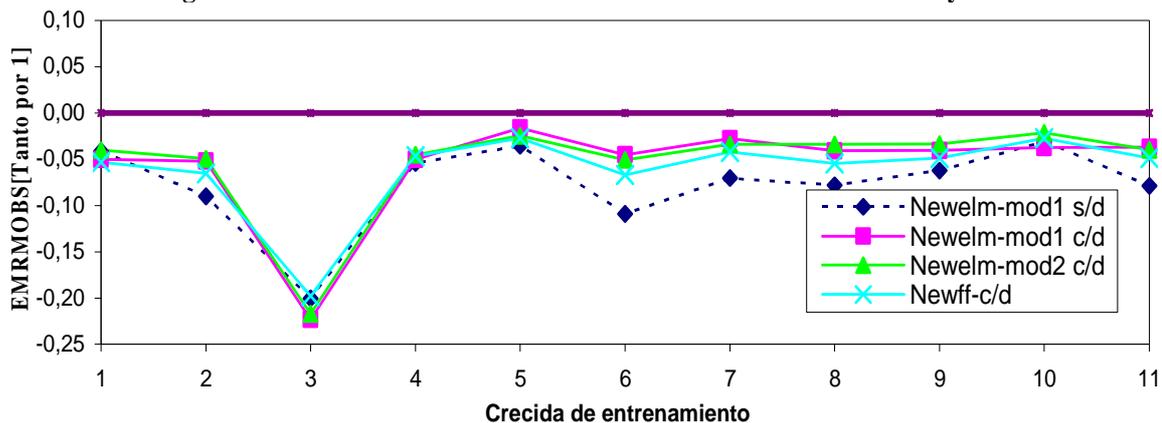


Figura 4.57: Índice TMEM entrenamiento. Modelos NEWELM y NEWFF.

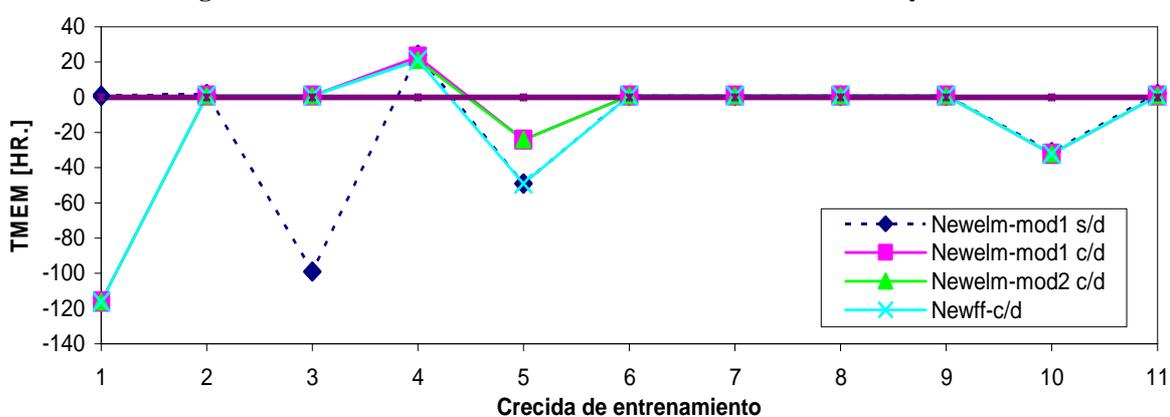


Figura 4.58: Índice EMEM validación crecida 5. Modelos NEWELM y NEWFF.

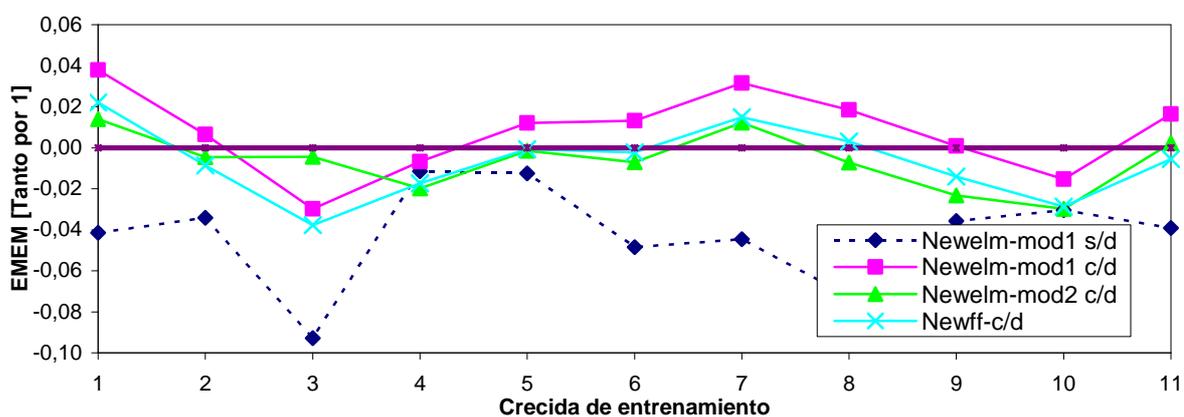


Figura 4.59: Índice EMRMOBS validación crecida 5. Mod. NEWELM y NEWFF.

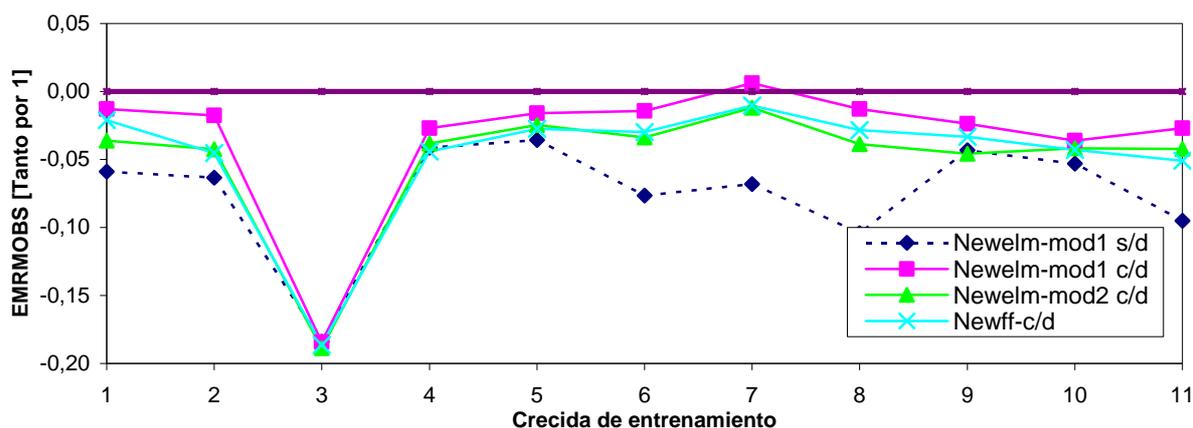
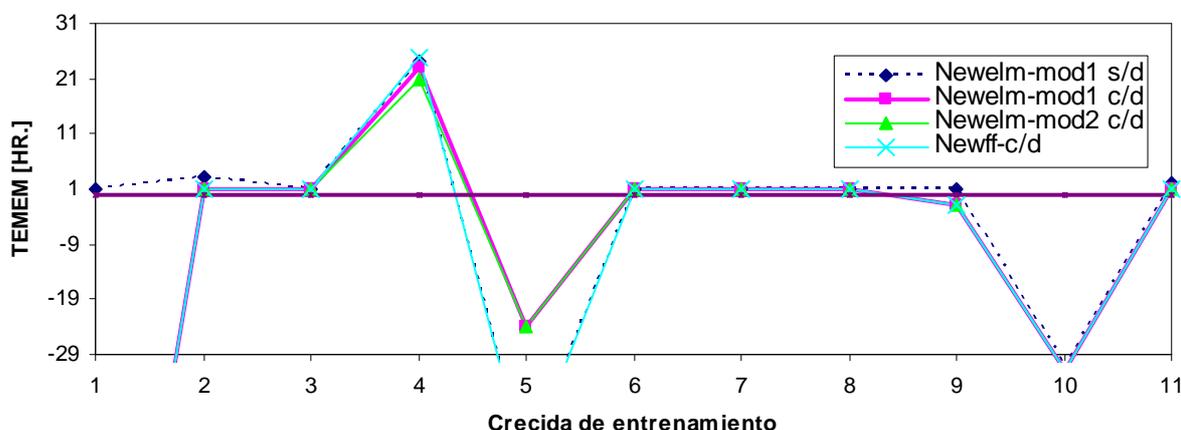


Figura 4.60: Índice TEMEM validación crecida 5. Modelos NEWELM y NEWFF.



4.4.1.2 Pronósticos mediante Redes Neuronales - Entrenamiento Continuo

4.4.1.2.1 Pronósticos mediante Redes Neuronales Recurrentes de Elman

El análisis continuo de los pronósticos a un paso de adelanto cuenta con un conjunto de entrenamiento de 5596 datos formado por las crecidas 1 a 6, un conjunto de validación de 3080 datos formado por las crecidas 5 a 7 y finalmente un conjunto de prueba de 1950 datos formado por las crecidas 8 a 11. Las características de cada conjunto de datos se presentan en la tabla 4.80.

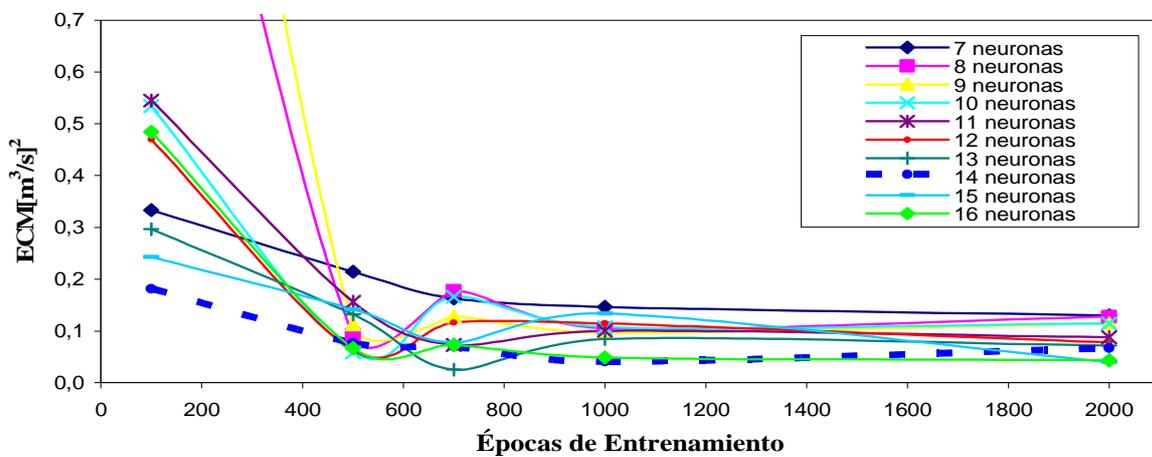
Tabla 4.80: Caracterización Conjunto de datos.

CONJUNTO	MEDIA [m ³ /s]	DESV ESTANDAR [m ³ /s]	Q Máx [m ³ /s]	QMin [m ³ /s]
ENTRENAMIENTO	5,1	2,7	17,4	2,1
VALIDACIÓN	3,9	1,1	7,4	2,6
PRUEBA	4,1	1,6	8,8	2,5

Se consideran como entradas las variables $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$ y se añade la Precipitación acumulada en 14 días antes de realizar el pronóstico, con las cuales se realizan diferentes configuraciones de entrada de tal manera de definir la importancia de cada una de ellas.

La figura 4.61 muestra el análisis de validación en función del número de neuronas en la capa oculta y de las épocas de entrenamiento, se deduce que 1200 épocas de entrenamiento son al menos necesarias y que 14 neuronas en la capa oculta forman la envolvente inferior de los errores.

Figura 4.61: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.



Desde ahora en adelante para este apartado un modelo NEWELM del tipo 5-X-1 es un modelo que involucra la precipitación acumulada además de las variables $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$ y un modelo tipo 4-X-1 es un modelo que no incorpora dicha variable.

Todos los modelos analizados tienen como parámetro de entrada a \dot{Q}_t , basado en que en el análisis unitario, modelos que incorporan dicha variable son más eficientes.

En la tabla 4.81 se presentan los errores obtenidos para distintas configuraciones neuronales considerando 1200 épocas de entrenamiento, se aprecia que el modelo óptimo se obtiene para una arquitectura de 14 neuronas en la capa oculta y que considera como variables de entrada: $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$

Luego, se elige como modelo óptimo de red neuronal recurrente de Elman un modelo NEWELM4-14-1.

Tabla 4.81: Errores modelo NEWELM Y-X -1.

Configuraciones	ECM [m ³ /s] ²			Ponderado [m ³ /s] ²	r ² entrenamiento	r ² validación	r ² prueba
	entrenamiento	Validación	Prueba				
	5596	3080	1950				
4-7-1	0,176	0,147	0,258	0,182	0,972	0,760	0,722
4-8-1	0,151	0,105	0,229	0,152	0,976	0,852	0,849
4-9-1	0,113	0,096	0,245	0,132	0,983	0,864	0,836
4-10-1	0,134	0,107	0,270	0,151	0,980	0,854	0,813
4-11-1	0,145	0,100	0,262	0,153	0,979	0,864	0,825
4-12-1	0,119	0,105	0,209	0,132	0,970	0,852	0,846
4-13-1	0,041	0,065	0,131	0,064	0,982	0,900	0,907
4-14-1	0,061	0,040	0,103	0,063	0,991	0,958	0,947
4-15-1	0,100	0,074	0,151	0,102	0,957	0,847	0,859
4-16-1	0,073	0,049	0,133	0,077	0,989	0,945	0,924
5-14-1	0,105	0,061	0,132	0,097	0,985	0,929	0,925
5-13-1	0,064	0,044	0,132	0,071	0,991	0,952	0,930

4.4.1.2.2 Pronósticos mediante Redes Neuronales Prealimentadas

Las entradas a la red quedan definidas por las siguientes 9 variables, las cuales se consideran representativas de una crecida de deshielo en un modelo neuronal prealimentado:

$$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}, \dot{Q}_t$$

La figura 4.62 muestra el análisis de los errores de validación en función del número de épocas de entrenamiento y neuronas de las capas ocultas para este tipo de red. Se observa que la envolvente mínima de errores considera 12 neuronas en cada capa oculta y un óptimo de 1200 épocas de entrenamiento.

Desde ahora en adelante para este apartado un modelo NEWFF del tipo 10-X-1 es un modelo que involucra la precipitación acumulada además de las variables “base” y un modelo tipo 11-X-1 es un modelo que no lo hace.

En la tabla 4.82 se presentan valores de los errores obtenidos en cada uno de los set de datos para distintas arquitecturas de RNR, donde se aprecia que el número óptimo de neuronas en la capa oculta es precisamente 12 neuronas en cada una, para una configuración 10-12-1.

Figura 4.62: Análisis de Errores de validación según número de neuronas y épocas de entrenamiento.

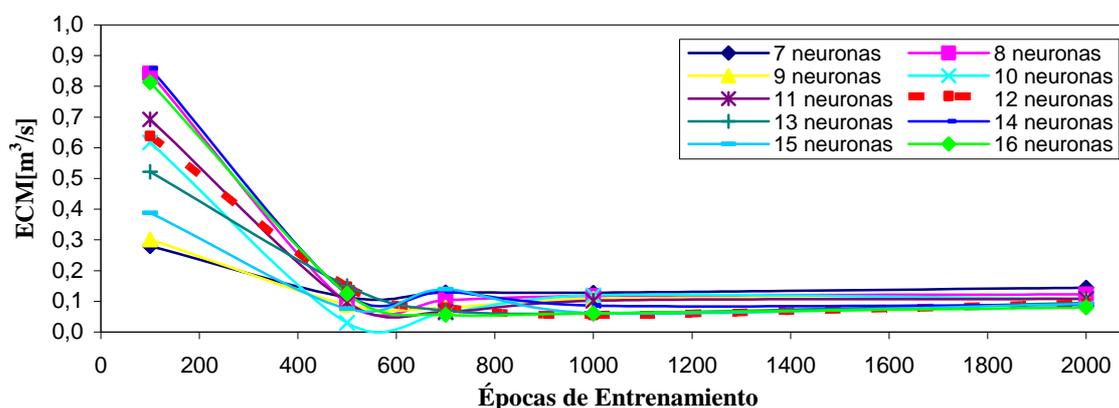


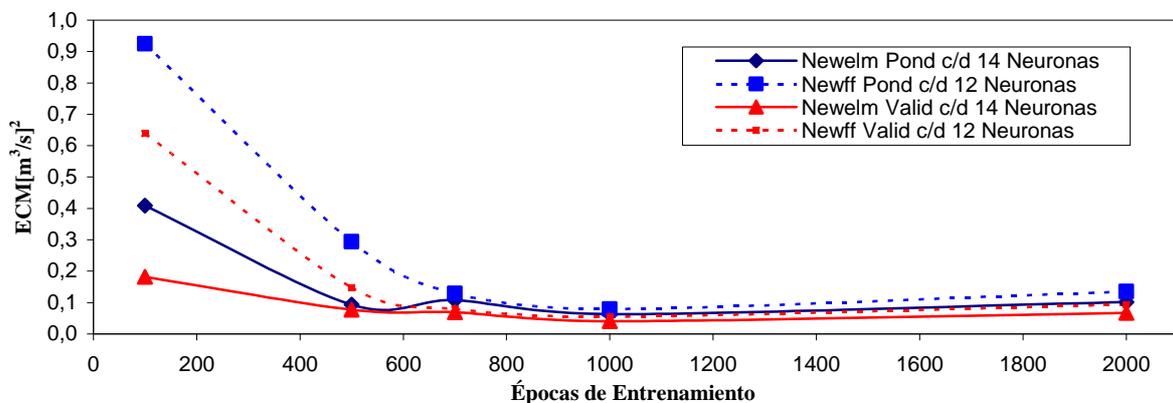
Tabla 4.82: Errores modelo NEWFF Y-X -1.

Configuraciones	ECM [m^3/s^2]			Ponderado [m^3/s^2]	r^2 entrenamiento	r^2 validación	r^2 prueba
	entrenamiento	Validación	Prueba				
10-7-1	5596	3080	1950	0,176	0,975	0,803	0,780
10-8-1	0,146	0,117	0,277	0,162	0,977	0,825	0,805
10-9-1	0,133	0,108	0,259	0,149	0,979	0,841	0,821
10-10-1	0,159	0,119	0,281	0,170	0,976	0,832	0,813
10-11-1	0,260	0,102	0,213	0,205	0,959	0,847	0,851
10-12-1	0,071	0,055	0,139	0,079	0,990	0,936	0,922
10-13-1	0,073	0,060	0,157	0,085	0,989	0,934	0,935
10-14-1	0,465	0,086	0,209	0,308	0,940	0,912	0,893
10-15-1	0,170	0,062	0,153	0,135	0,974	0,922	0,906
10-16-1	0,107	0,062	0,164	0,104	0,985	0,930	0,910
11-14-1	0,073	0,059	0,149	0,083	0,992	0,943	0,917
11-12-1	0,248	0,102	0,330	0,221	0,965	0,876	0,796

Al comparar los errores cometidos en la mejor arquitectura neuronal obtenida para el modelo de Elman y el modelo prealimentado, se observa que un modelo de Elman presenta errores más bajos que un modelo prealimentado. Ponderadamente la diferencia es de aprox. un 20% del valor del error obtenido para este último modelo, ventaja que se complementa además, con la cantidad de variables de entrada necesarias para realizar pronósticos.

En la figura 4.63, se presentan los errores ponderados en función del número de épocas de entrenamiento, para el mejor modelo neuronal de Elman y el mejor modelo neuronal prealimentado. Se considera la nomenclatura “c/d” a si las variables de entrada consideran la variable \dot{Q}_t y “s/d” a si no la consideran. Se desprende que no sólo el entrenamiento, sino también las validaciones y pruebas (a través del error ponderado) son mejores con un modelo neuronal de Elman.

Figura 4.63: Análisis de Errores Ponderados según el modelo de pronóstico.



En la tabla 4.83 se presentan los índices de eficiencia de pronóstico de caudal máximo, se desprende que la red de Elman entrega mejores pronósticos en cada uno de los conjuntos de datos. Esta red subestima los caudales máximos entre un 15 a un 18% del valor máximo observado y destacable es que tanto en el entrenamiento como en la validación la ocurrencia del caudal máximo pronosticado es en el tiempo real.

Tabla 4.83: Índices de eficiencia de pronósticos de caudal máximo.

ÍNDICES	NEWELM 4-14-1			NEWFF 10-12-1		
	Entrenamiento	Validación	Prueba	Entrenamiento	Validación	Prueba
EMEM [tanto por 1]	-0,184	-0,155	0,160	-0,193	-0,168	-0,181
EMRMOBS [tanto por 1]	-0,184	-0,155	0,175	-0,260	-0,178	-0,184
TMEM [hr.]	0	0	-24	1	-49	-29

4.4.1.3 Comparación de Modelos Neuronales de Pronóstico por crecidas

En la figura 4.64 se presentan 2 crecidas validadas con los modelos neuronales NEWELM y NEWFF elegidos como los mejores modelos tanto para un entrenamiento de tipo unitario como continuo. Se puede apreciar gráficamente las ventajas de realizar pronósticos mediante redes neuronales recurrentes de Elman entrenadas con los datos de la crecida 2 sobre modelos de redes prealimentadas entrenadas tanto unitariamente (crecida 2) como continuamente.

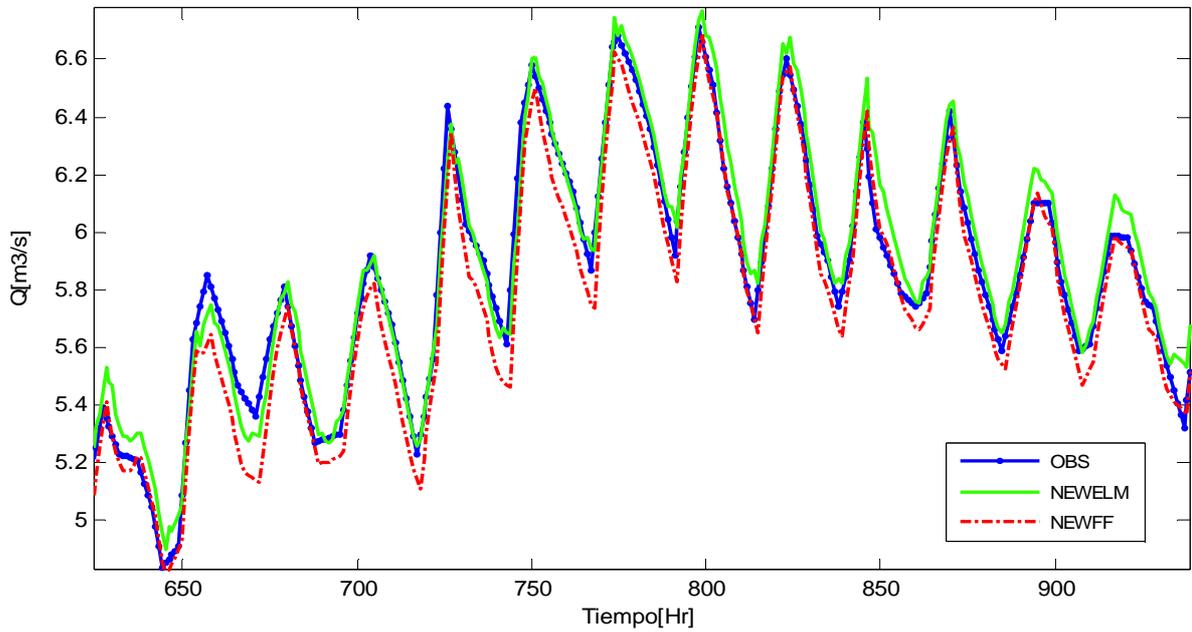
En la figura 4.64.1, se observa que en general la red de Elman en algunos caudales máximos sobrestima y en otros subestima si es que no es exacta, en cambio la red prealimentada en general los subestima si es que no los alcanza, pero lo más llamativo es el hecho de que la red prealimentada en general tiende a subestimar los mínimos notoriamente, lo que no sucede en una red de Elman donde se alcanzan casi con exactitud.

En la segunda figura 4.61, se observa una sobrestimación de los caudales máximos para una red prealimentada y el mismo fenómeno que en un entrenamiento de tipo unitario, el cual es una subestimación de los caudales mínimos, pero menos notoria. Se observa un comportamiento de la red de Elman no identificable con alguna tendencia, ya que en algunos puntos subestima y en otros sobrestima, llama la atención que en algunos caudales mínimos la red sobrestima sin ninguna explicación predecible. En términos generales para este tipo de entrenamiento y en particular para esta crecida se observa un buen pronóstico con la red de Elman pero al nivel de una red prealimentada

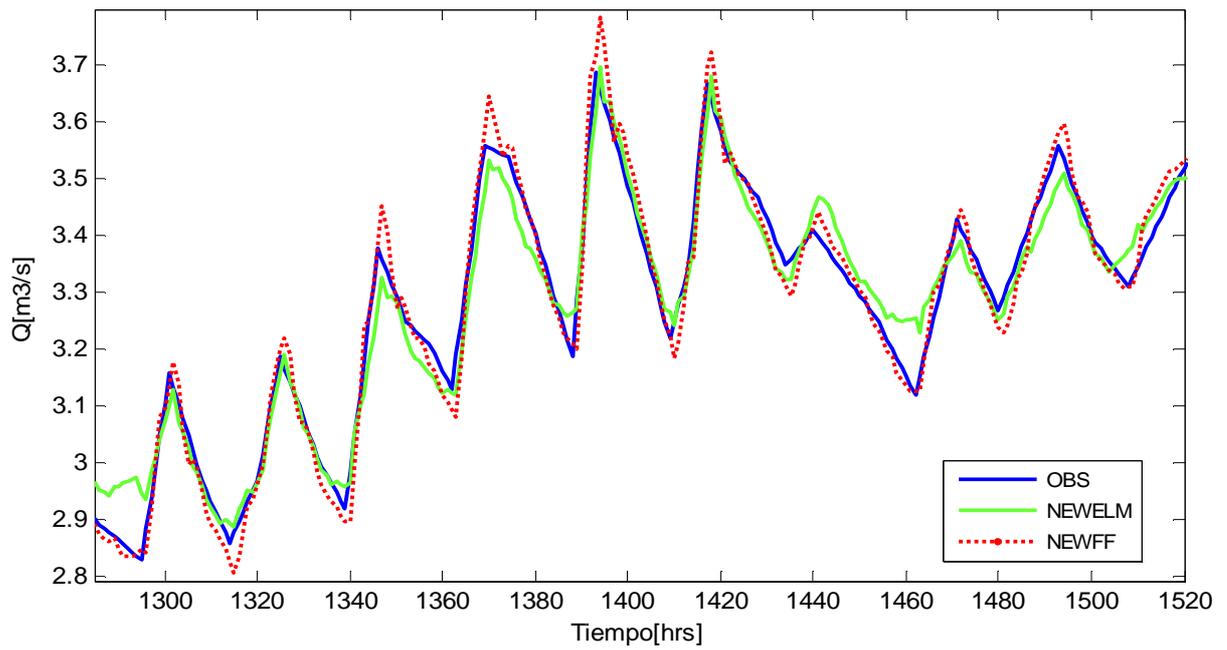
Luego, en base a las gráficas se deduce que una red recurrente de Elman de arquitectura 4-10-1 con configuración de entrada $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$, entrenada con los datos de la crecida 2, resulta ser el modelo de pronóstico más eficiente.

Figura 4.64: Validación de crecidas según modelo neuronal y tipo de entrenamiento.

4.64.1-Pronóstico a un paso de adelanto-Crecida 10 - Entre. Unitario.



4.64.2-Pronóstico a un paso de adelanto-Crecida 5 - Entre. Continuo.



4.4.1.4 Pronóstico mediante Modelos Paramétricos

Los tres conjuntos empleados en el análisis de tipo continuo se utilizan para realizar un análisis de pronóstico del mismo tipo mediante modelos paramétricos, ARMAX y RARMAX a una hora de adelanto.

La variable de salida es Q_{t+1} y las variables de entrada: Q_t, T_{T+1-D} y \dot{Q}_t

Los retardos de caudal (na), temperatura (nb) y gradiente de caudal se varían de 1 a 7 hrs. y el desfase inicial de temperatura (nk) (desfasada ya en T + 1-D) desde 1 a 2 hrs.

Al realizar el análisis de errores se observa que al igual que en el caso de crecidas pluviales, un modelo que involucra la variable gradiente de caudal como parámetro de entrada no es más eficiente que uno que no lo hace, por lo tanto en un modelo de tipo paramétrico usado para pronósticos de crecidas de deshielo dicha variable no aporta información adicional al modelo y resulta innecesaria.

En la tabla 4.84 se observan los mejores modelos tipo ARMAX y RARMAX, obtenidos de acuerdo a una comparación del error cuadrático medio ponderado, cuyos valores se observan en la tabla 4.85.

Los valores del ECM y el coeficiente de determinación asociado a cada conjunto de datos para modelos ARMAX y RARMAX se pueden observar en la tabla 4.86 y 4.87 respectivamente.

Se desprende que el mejor modelo de pronóstico es un modelo ARMAX que involucra 7 retardos de la variable caudal y 2 retardos de la variable temperatura a partir de t +1-D, donde D es el desfase entre los máximos de caudal y temperatura.

Tabla 4.84: Mejores Modelos ARMAX y RARMAX

Tabla 4.84.a: Modelos ARMAX.

Modelo	na	nb	nk
56	7	1	2
21	7	3	1
84	7	5	2
77	7	4	2
53	4	1	2
76	6	4	2
42	7	6	1

Tabla 4.84.b: Modelos RARMAX.

Modelo	na	nb	nk
50	1	1	2
43	1	7	1
36	1	6	1
71	1	4	2
8	1	2	1
22	1	4	1
78	1	5	2

Tabla 4.85: Error Cuadrático Medio Ponderado Modelo ARMAX y RARMAX.

Tabla 4.85.a: Modelo ARMAX.

Modelo	ECM ponderado [m ³ /s] ²	R ² ponderado
56	0,039	0,970
21	0,039	0,970
84	0,039	0,970
77	0,039	0,970
53	0,040	0,969
76	0,040	0,969
42	0,040	0,969

Tabla 4.85.b: Modelo RARMAX.

Modelo	ECM ponderado [m ³ /s] ²	R ² ponderado
50	0,040	0,969
43	0,040	0,969
36	0,041	0,969
71	0,041	0,969
8	0,041	0,968
22	0,041	0,968
78	0,041	0,968

Tabla 4.86: Error cuadrático medio Modelos ARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento [m ³ /s] ²	ECM Validación [m ³ /s] ²	ECM Prueba [m ³ /s] ²
56	0,905	0,875	0,862	0,017	0,054	0,052
21	0,905	0,875	0,862	0,017	0,054	0,052
84	0,905	0,875	0,862	0,017	0,054	0,052
77	0,905	0,875	0,862	0,017	0,055	0,052
53	0,905	0,874	0,861	0,018	0,055	0,053
76	0,904	0,874	0,861	0,018	0,055	0,053
42	0,904	0,874	0,861	0,018	0,055	0,053

Tabla 4.87: Error cuadrático medio Modelos RARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento [m ³ /s] ²	ECM Validación [m ³ /s] ²	ECM Prueba [m ³ /s] ²
50	0,905	0,874	0,862	0,018	0,053	0,056
43	0,905	0,874	0,862	0,018	0,053	0,056
36	0,905	0,874	0,862	0,018	0,054	0,057
71	0,905	0,874	0,862	0,018	0,054	0,057
8	0,904	0,871	0,861	0,018	0,055	0,057
22	0,904	0,871	0,861	0,019	0,050	0,058
78	0,904	0,871	0,861	0,019	0,050	0,058

Se desprende que un modelo ARMAX presenta mejores pronósticos que un modelo RARMAX, pero la diferencia es mínima y se debe principalmente al aumento de los errores tanto de entrenamiento como de prueba que se comenten en el segundo modelo con respecto al primero. Ambos modelos entregan errores similares y del mismo orden de magnitud, en términos del coeficiente de determinación, si bien es mejor un modelo ARMAX, ambos modelos son recomendables como modelos de pronóstico de deshielo.

ARMAX y RARMAX entregan errores de pronóstico del mismo orden que los errores que se obtienen por medio de redes neuronales tanto recurrentes como prealimentadas entrenadas continuamente, pero si se comparan los ECM obtenidos con los entregados por un modelo de red recurrente entrenada unitariamente, los errores son mayores y aproximadamente del doble.

Se realiza también, un entrenamiento unitario en este tipo de modelos, donde cada uno tiene la mejor configuración obtenida para ARMAX y RARMAX “continuo” y es entrenado de acuerdo a los datos de cada crecida, de tal manera que se forman 11 modelos. En la tabla 4.88 se presentan los errores ponderados obtenidos para cada modelo. Se observa que en modelos

ARMAX la tendencia a relacionar la capacidad de generalización con las crecidas con mayores amplitudes de frecuencias no es tan clara, aunque los modelos formados a partir de crecidas con destacadas amplitudes de frecuencia forman buenos modelos de pronóstico (5 y 10), aquella que mayor variabilidad y amplitudes (crecida n° 2) no entrega el error más bajo de pronóstico, pero viceversa sí, la crecida con más bajas amplitudes, n° 8, arroja el error ponderado más alto.

Tabla 4.88: Errores Ponderados Modelo ARMAX y RARMAX.

Nª Crecida	RARMAX	ARMAX
	Error Ponderado [m ³ /s] ²	
1	0,044	0,028
2	0,044	0,030
3	0,044	0,027
4	0,044	0,028
5	0,044	0,026
6	0,044	0,028
7	0,044	0,029
8	0,044	0,034
9	0,044	0,029
10	0,044	0,026
11	0,044	0,040

Se desprende que modelos RARMAX entrenados continuamente entregan mejores resultados que aquellos formados a través de entrenamientos de tipo unitario, en el caso de modelos ARMAX se observan errores menores para entrenamientos de este último tipo. Las razones que justifican dichos comportamientos son las mismas que se exponen en el caso de pronósticos de crecidas pluviales para ambas cuencas y que tienen relación con la base de operación de los modelos.

En general, para modelos ARMAX se obtienen errores similares usando cualquier crecida como set de entrenamiento, aunque el desempeño es mucho más equilibrado en un modelo RARMAX, donde cualquier crecida que sea utilizada como conjunto de datos de entrenamiento tiene el mismo efecto en términos de los errores.

A nivel de entrenamientos unitarios, un modelo de redes neuronales recurrentes NEWELM 4-10-1 con un entrenamiento unitario (crecida 2) entrega un 18% menos de error que el mejor modelo ARMAX, porcentaje aún mayor en el caso de ser comparado con modelos RARMAX (50%).

4.4.1.5 Resumen y análisis de modelos de pronósticos a un paso de adelanto

Esta sección tiene por objetivo, situar al lector en el desarrollo de los pronósticos realizados hasta el momento, a fin de continuar con pronósticos a varios pasos de adelanto.

Se realizan análisis entre los mejores modelos de pronóstico de cada tipo analizado. Para comparar modelos neuronales entrenados unitariamente se considera como crecida de entrenamiento la crecida 2 (recordar, crecida con mayor amplitud frecuencial y demostrada su mayor capacidad de generalización). Se comparan modelos de redes neuronales recurrentes de Elman y de redes prealimentadas a nivel de entrenamiento unitario y continuo y además, modelos paramétricos, ARMAX y RARMAX [na, nb, nk].

Para estos últimos modelos se incluyen sólo el modelo ARMAX unitario y el modelo RARMAX continuo, debido a las razones indicadas precedentemente (punto 4.4.1.4) que los señalan como los mejores modelos de pronóstico de este tipo.

Los parámetros de comparación utilizados son el ECM ponderado, el ECM de validación por crecida y los índices de error en el pronóstico de máximos. Los cuales se presentan en la tabla 4.89.a ,4.89.b y 4.89.c respectivamente.

En la tabla 4.89.a se cuantifica la eficiencia del modelo de pronóstico en base al ECM ponderado, a partir del cual se infiere que tanto modelos neuronales recurrentes de Elman, prealimentados y un modelo paramétrico ARMAX con entrenamientos unitarios resultan buenos modelos de pronóstico, siendo el primero de ellos el mejor. Se observan aumentos de los errores para entrenamientos de tipo continuo con respecto a los de tipo unitario, pero el orden de magnitud se mantiene. En términos generales resulta más eficiente realizar entrenamientos de tipo unitario con un modelo NEWELM 4-10-1.

Es destacable el hecho que a nivel continuo un modelo paramétrico resulta más eficiente (c/r al ECM) que un modelo neuronal.

Tabla 4.89.a: Errores ponderados según modelo de pronóstico.

ECM PONDERADO [m ³ /s] ²	Entrenamiento	
	Unitario	Continuo
NEWELM	0,022	0,063
NEWFF	0,023	0,079
ARMAX	0,026	0,039
RARMAX	0,044	0,040

En la tabla 4.89.b se presentan los errores de validación obtenidos para cada crecida, se pueden apreciar gráficamente en la figura 4.65. Desde este análisis de eficiencia segmentado por crecidas, se deduce que tanto un modelo ARMAX como un modelo neuronal ya sea recurrente

de Elman o prealimentado entrenado unitariamente es un buen modelo de pronóstico. Cabe destacar que el modelo ARMAX presenta errores mayores en aquellas crecidas que poseen mayor capacidad de generalización y que pueden ser también las más caudalosas y más difíciles de controlar.

A nivel de entrenamiento continuo para el modelo NEWELM y NEWFF, se observa que no existe una envolvente clara y si bien la red de Elman arroja menores errores en general, sobre todo para las primeras crecidas (se debe recordar que son parte del conjunto de entrenamiento), las crecidas 1, 6 y 11 presentan mejores validaciones en un modelo NEWFF y las 8 crecidas restantes son mejor pronosticadas con un modelo NEWELM.

En la tabla 4.89.c se presenta el error que se produce en el pronóstico de máximos por medio de los dos modelos neuronales analizados, estos índices resultan muy importantes de analizar, dada la eficiencia de tres modelos, en los ámbitos considerados anteriormente. El error es cuantificado para el máximo de cada crecida y luego ponderado, al igual que las tablas que presentan dichos índices precedentemente.

Se deduce que un modelo NEWELM 4-10-1 resulta el modelo más eficiente para pronosticar máximos mediante redes neuronales, debido a que presenta los menores errores. Subestimaciones de aprox. 2%, valor que crece a aprox. 5% en el momento de ocurrencia real del caudal máximo.

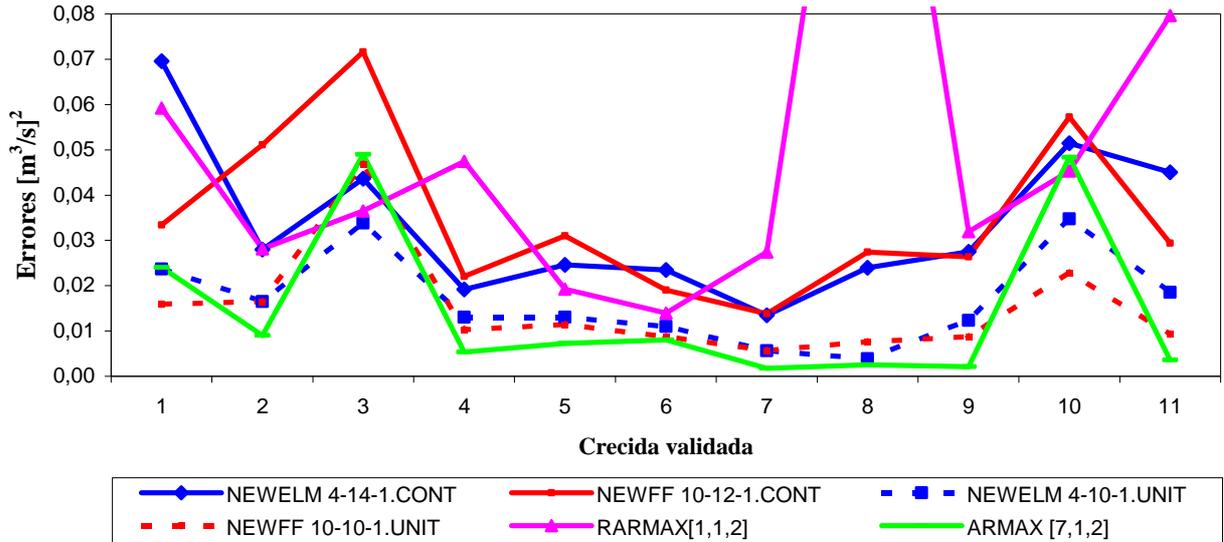
Tabla 4.89.b: Errores de validación por crecida según modelo de pronóstico.

Crecida	ECM de validación [m ³ /s] ²					
	Unitario		Continuo		ARMAX [7,1,2]	RARMAX[1,1,2]
	NEWELM 4-10-1	NEWFF 10-10-1	NEWELM 4-14-1	NEWFF10-2-1		
1	0,024	0,016	0,070	0,033	0,024	0,059
2	0,017	0,017	0,028	0,051	0,009	0,028
3	0,034	0,047	0,044	0,072	0,049	0,037
4	0,013	0,010	0,019	0,022	0,005	0,047
5	0,013	0,011	0,025	0,031	0,007	0,019
6	0,011	0,009	0,023	0,019	0,008	0,014
7	0,006	0,006	0,013	0,014	0,002	0,027
8	0,004	0,008	0,024	0,027	0,003	0,173
9	0,012	0,009	0,027	0,026	0,002	0,032
10	0,035	0,023	0,051	0,057	0,048	0,045
11	0,019	0,009	0,045	0,029	0,004	0,080

Tabla 4.89.c: Índices de error en el pronóstico de caudales máximos según modelo de pronóstico

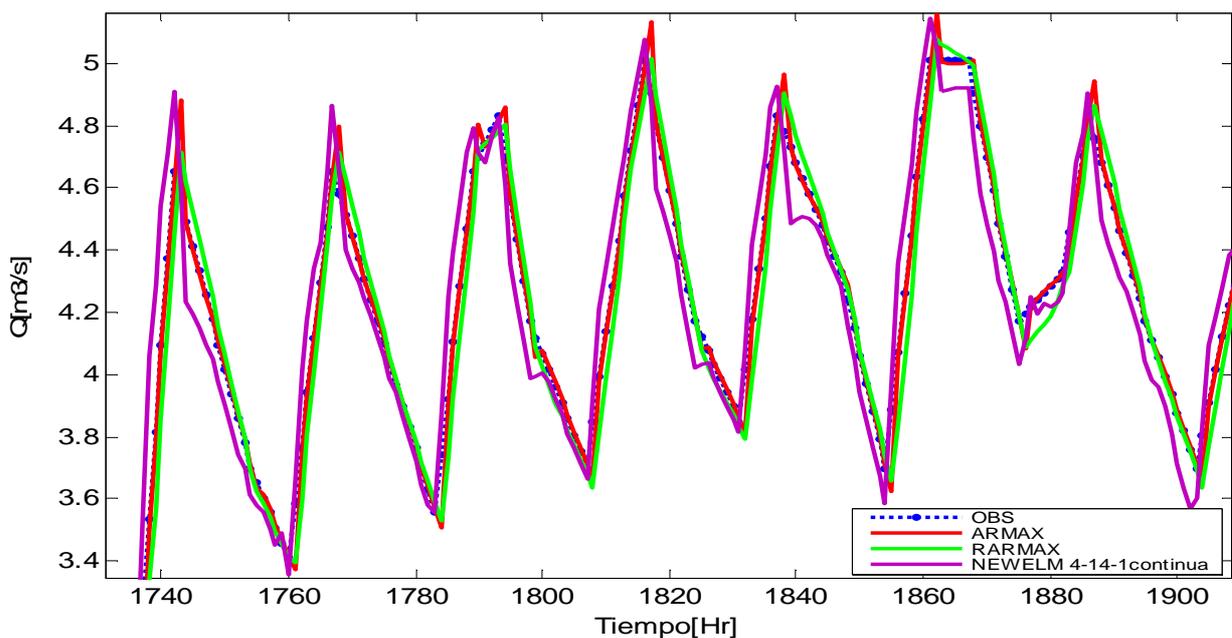
ÍNDICES DE ERROR	Entrenamiento					
	Unitario			Continuo		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]
NEWELM	-0,020	-0,046	1	-0,112	-0,110	-4,404
NEWFF	-0,028	-0,068	1	-0,184	-0,222	-18,998

Figura 4.65: Análisis de Errores de validación según el modelo de pronóstico.



En la figura 4.66 se observa una ventana de tiempo del conjunto de validación pronosticado según los diferentes modelos considerados, en general el modelo ARMAX, RARMAX y el modelo neuronal recurrente de Elman resultan eficientes y en general la curvas de caudales se superponen a la de caudales observados e incluso en algunos puntos el modelo paramétrico resulta más eficiente. Si se observa detalladamente se distingue que la curva de pronósticos de la red NEWELM se encuentra adelantada, exceptuando el máximo que lo pronostica al tiempo real, estos comportamientos pueden resultar muy útiles a la hora de realizar pronósticos.

Figura 4.66: Análisis de Modelos en el conjunto de validación.



Considerando todos los índices analizados el mejor modelo de pronóstico a un paso de adelanto es un modelo NEWELM 4-10-1 entrenado con los datos de la crecida 2. Cabe señalar que no se debe desmerecer la eficiencia de los otros métodos, que poseen teorías y funcionamientos más simples que requieren un menor tiempo de procesamiento de datos, lo cual los aventaja sobre la complejidad de la red recurrente de Elman.

Antes de continuar, cabe señalar que no se ha considerado realizar pronósticos mediante la red recurrente de Hopfield en este tipo de crecidas. Las razones radican en la alta eficiencia de los modelos estudiados hasta el momento, se ha logrado mejorar los pronósticos mediante el uso de la red recurrente de Elman, con respecto a los pronósticos entregados por modelos ya existentes, en forma notoria en la estimación de los valores máximos tanto en cantidad como en calidad.

Dado el comportamiento cíclico de éste tipo de crecidas, la gran cantidad de registros que las caracterizan y el hecho de que el modelo de Hopfield se base en la identificación de un gradiente o diferencia (otorga junto a las demás variables involucradas las condiciones iniciales de la crecida y del comportamiento de la cuenca), resulta difícil aplicar el modelo de Hopfield, tanto por la cantidad de información a almacenar, que debe ser codificada y que además genera una gran cantidad de patrones. Luego, garantizar el correcto funcionamiento de la red se ve poco factible, dado que el número de patrones almacenado no debe ser excesivamente mayor a la cantidad de neuronas de la red, ver punto 2.2.4.1, sumado al hecho de que la variabilidad de la crecida producto de su ciclicidad hace que el gradiente deje de ser representativo a varios pasos de adelanto.

4.4.1.6 Pronósticos a varios pasos de adelanto

Se realiza pronósticos de caudales a 6, 12, 24 y 48 horas considerando dos tipos de modelos uno que considera como conjunto de datos de entrenamiento sólo la crecida número 2 y otro que considera como dicho conjunto, aquel empleado en el análisis de tipo continuo.

Para elegir la arquitectura óptima de los modelos en función de minimizar el ECM, se prueban nuevamente diferentes números de neuronas en la capa oculta para cada modelo neuronal definido de acuerdo a la salida que se desea. Para ello, se parte de la base de que para pronosticar a un paso de adelanto se necesitan las arquitecturas anteriormente señaladas.

Las variables de entrada de los modelos neuronales son las mismas que se consideran en los mejores modelos elegidos para un paso de adelanto.

Luego, los modelos neuronales que se utilizan para realizar pronósticos a $t + x$ horas de adelanto son los siguientes:

Tabla 4.90: Arquitecturas óptimas de red.

horas	UNITARIO		CONTINUO	
	NEWELM	NEWFF	NEWELM	NEWFF
6	10-7-1	10-10-1	10-13-1	10-12-1
12	10-7-1	10-10-1	10-10-1	10-12-1
24	10-10-1	10-7-1	10-7-1	10-14-1
48	10-12-1	10-10-1	10-12-1	10-9-1

En la tabla 4.91 se presenta los errores ponderados de pronósticos realizados por medio de modelos neuronales de Elman y prealimentados considerando entrenamientos de tipo continuo. Se desprende que si bien los errores ponderados para una red de tipo recurrente no son significativamente menores a los de una red prealimentada, son lo suficiente como para indicar que la RRS entrega un mejor modelo de pronóstico.

En la tabla 4.92 se presenta los errores de entrenamiento y ponderados obtenidos de realizar pronósticos de acuerdo al análisis unitario por medio de la crecida 2. Se observa que los errores de entrenamiento y ponderados (reflejan el comportamiento de los errores de validación) son menores utilizando un modelo de red recurrente, lo que es producto del beneficio que otorgan gracias al retardo de la información en el fenómeno individual.

Comparando ambos modelos de acuerdo a sus tipos de entrenamiento, se infiere que un modelo de Elman entrega levemente menores errores a nivel de entrenamiento unitario, en cambio un modelo prealimentado no presenta una tendencia clara a otorgar mejores pronósticos con alguno de los tipos de entrenamiento, por ejemplo a 6 hrs. de adelanto el modelo es mejor a nivel unitario y a 20 hrs. es mejor a nivel de entrenamiento continuo.

Es importante destacar el hecho de que una red de Elman entrenada continuamente presenta prácticamente igual desempeño que una entrenada unitariamente, esta última es levemente mejor según se observa, lo cual tiene su explicación en que se produce una mejora a nivel de “crecidas pegadas”, dada por el hecho de que las crecidas de deshielo son más duraderas que las crecidas pluviales y por lo tanto los parámetros de la red logran ajustarse mejor puesto que se produce un mayor procesamiento de la información propia de la crecida durante el ajuste de dichos parámetros descartando más la información antecedente.

El error de pronósticos es bastante bajo tanto para modelos recurrentes de Elman como para modelos prealimentados hasta 48 horas de adelanto y si bien en términos generales un modelo neuronal de Elman entrega pronósticos más eficientes, realizar pronósticos con cuales quiera de los dos modelos es recomendable.

Tabla 4.91: ECM entrenamiento y Ponderado. Modelos NEWELM y NEWFF. Análisis Continuo.

Modelo (Continuo)	Error Ponderado $[m^3/s]^2$			
	t+6	t+12	t+24	t+48
Newelm 4-14-1	0,240	0,411	0,537	0,923
Newff 10-12-1	0,295	0,524	0,586	1,016

Tabla 4.92: ECM entrenamiento y Ponderado. Modelos NEWELM y NEWFF. Análisis unitario.

Pronóstico	Error de Entrenamiento $[m^3/s]^2$		Error Ponderado $[m^3/s]^2$	
	Newelm	Newff	Newelm	Newff
t+6	0,221	0,213	0,240	0,226
t+12	0,407	0,456	0,418	0,474
t+24	0,493	0,503	0,538	0,619
t+48	1,049	1,071	0,912	1,508

En la figura 4.67 se presenta la correlación existente entre el caudal pronosticado y observado (análisis continuo) de las crecidas que forman parte del conjunto de validación y en la tabla 4.943 se presentan los caudales máximos observados y pronosticados para una red de Elman entrenada con los datos de la crecida 2. Se deduce la factibilidad de realizar pronósticos hasta 48 hrs.

El error promedio cometido en pronósticos a 48 hrs. por medio de la red NEWELM es de aprox. del 12% del caudal medio máximo de todas las crecidas analizadas.

Figura 4.67: Caudal Pronosticado y observado (Validación continua). Modelo NEWELM.

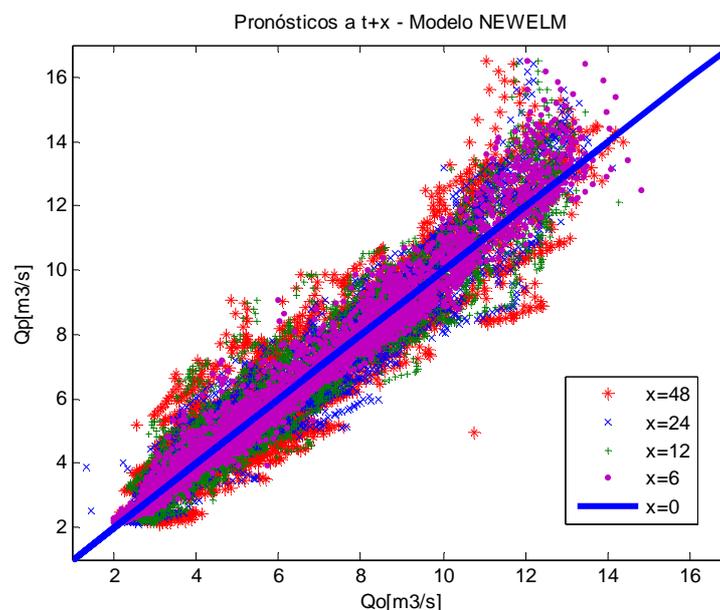


Tabla 4.93: Caudal máximo observado y pronosticado a t+x hrs. Modelo NEWELM. Análisis unitario.

Crecida	Q máx Observado [m ³ /s]	Q máx simulado [m ³ /s]			
		t+6	t+12	t+24	t+48
1	6,0	6,0	5,7	5,8	5,7
2	9,9	9,5	8,7	9,4	8,9
3	17,4	16,5	15,8	16,8	16,1
4	5,1	4,9	4,3	4,7	4,5
5	7,4	7,0	6,4	7,0	6,7
6	7,3	7,1	6,8	6,9	6,0
7	4,1	4,0	3,7	4,0	4,0
8	4,2	4,2	4,2	4,4	4,4
9	5,4	5,4	5,2	5,3	7,4
10	8,8	8,3	8,1	8,0	7,3
11	4,1	4,0	3,6	4,0	3,4

Para su comparación con modelos de tipo paramétrico, se utiliza el mejor modelo de ellos, es decir un modelo ARMAX de configuración [7,1,2] entrenado unitariamente para pronósticos a un paso de adelanto. Los valores de los ECM obtenidos para distintos tiempos de adelanto se presentan en la tabla 4.94. Se observa que los errores obtenidos por medio de estos modelos son mayores que los entregados por un modelo neuronal recurrente de Elman y prealimentado, pero destacándose que son del mismo orden de magnitud.

Tabla 4.94: ECM Ponderado. Modelos ARMAX.

Modelo ARMAX (unitario)	Error Ponderado [m ³ /s] ²			
	t+6	t+12	t+24	t+48
	0,340	0,761	0,852	1,210

En las tablas 4.95 y 4.96 se presentan los índices de eficiencia de pronósticos de caudal máximo hasta 48 horas. Se deduce que un modelo neuronal de Elman realiza mejores pronósticos de máximos en cuanto a valor. El índice de desfase en sí no tiene mucha validez dado que se trata de un fenómeno cíclico y por lo tanto el hecho de que existan máximos locales cercanos al máximo global y sus pronósticos sobrepasen el valor pronosticado en el tiempo en que ocurre el máximo global, lleva a que los desfases sean en general grandes.

Tabla 4.95: Índices de pronóstico de caudal máximo. Análisis Continuo.

ÍNDICE	NEWELM	NEWFF
EMEM [%]	-0,054	-0,078
EMRMMOBS[%]	-0,068	-0,089
TMEM[HR]	-8,400	-22,200

Tabla 4.96: Índices de pronóstico de caudal máximo. Análisis unitario.

ÍNDICE	NEWELM	NEWFF
EMEM [%]	-0,047	-0,058
EMRMMOBS[%]	-0,070	-0,100
TMEM[HR]	-7,200	-24,000

4.4.2 Cuenca del Mapocho en Los Almendros

Se cuenta con un total de 10 crecidas (13145 datos) cuyas características se presentan en la tabla 4.97. Al observar los desfases entre la variable temperatura y caudal para los valores máximos se deduce que existe un desfase de aprox. 6 horas, razón por la cual se utiliza un modelo neuronal de una capa oculta.

Tabla 4.97: Caracterización de Crecidas.

CRECIDA	INICIO	TERMINO	DURACIÓN [HR]	MEDIA [m ³ /s]	DESV ESTANDAR [m ³ /s]	Q Máx [m ³ /s]	QMin [m ³ /s]	Tº[ºC]
1	6-9-82 16:00	17-9-82 21:00	269	19,2	4,1	32,4	12,4	10,5
2	7-10-82 20:00	10-2-83 0:00	3004	27,0	6,7	47,1	14,4	19,3
3	3-10-83 5:00	30-11-83 17:00	1404	17,5	4,6	30,5	9,8	17,9
4	6-10-84 9:00	7-12-84 5:00	1484	21,7	5,2	42,4	11,3	16,4
5	23-9-87 19:00	7-1-88 10:00	2535	32,1	9,8	63,8	14,4	18,5
6	3-10-89 19:00	27-11-89 2:00	1303	17,1	2,6	24,6	11,5	16,7
7	11-9-91 19:00	16-9-91 22:00	123	18,4	3,0	25,9	13,7	11,7
8	18-10-91 19:00	8-12-91 21:00	1226	19,6	3,2	29,6	13,8	17,7
9	10-10-92 15:00	23-10-92 19:00	316	14,6	3,0	24,3	10,4	16,1
10	1-11-97 0:00	12-1-98 0:00	1728	24,3	4,4	39,8	16,0	19,4

Se aplica la transformada de Fourier a la serie de caudales que compone cada crecida, lo cual queda representado en la figura 4.68 y 4.69. Se deduce que las crecidas con mayor amplitud de frecuencias son las crecidas número 5, 2, 4 y 10, las que presentan menor amplitud y menor variabilidad frecuencial son las crecidas número 7, 8, 9 y 1.

Figura 4.68: Transformada de Fourier de la serie de caudales –tiempo.

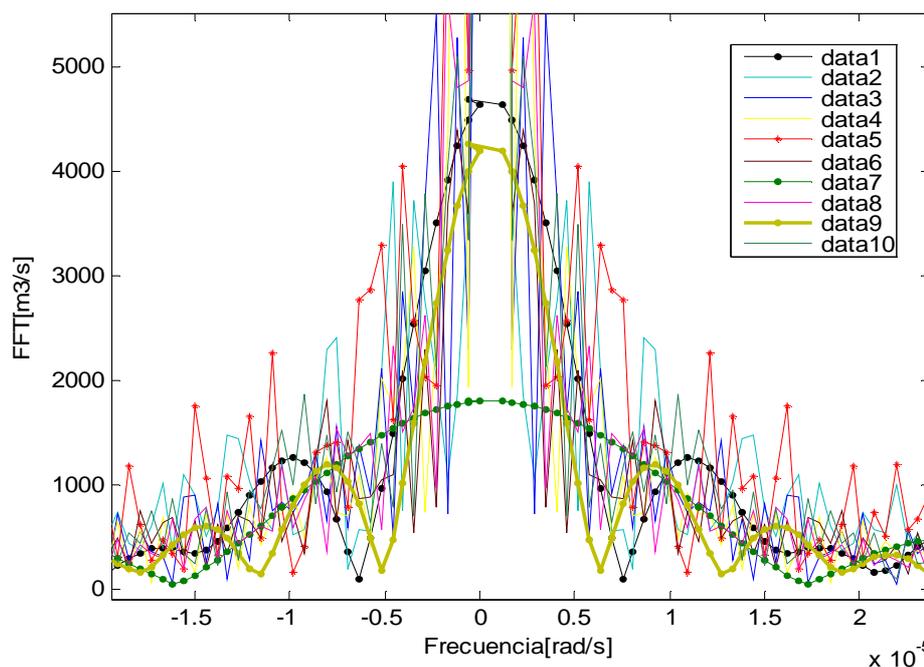
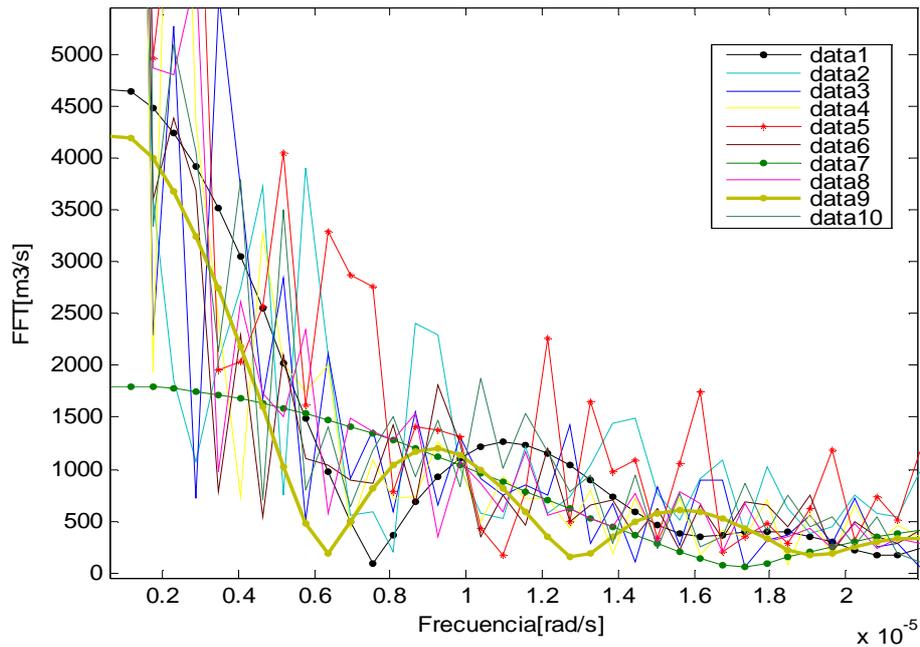


Figura 4.69: Transformada de Fourier de la serie de caudales –tiempo.



4.4.2.1 Pronóstico mediante Redes Neuronales – Entrenamiento Unitario

Primeramente y en este apartado se realizan pronósticos a un paso de adelanto.

De acuerdo al planteamiento general, cada crecida es usada como set de datos de entrenamiento, formando un modelo neuronal que es validado con las crecidas restantes, por lo tanto, se obtienen 10 modelos independientes. Esto crea una matriz de errores en que el error cuadrático medio de validación de la crecida de entrenamiento es el mismo error que se obtiene de su entrenamiento.

4.4.2.1.1 Pronóstico mediante Redes Neuronales Recurrentes de Elman

Las configuraciones planteadas involucran como parámetros de entrada Q_t , Q_{t-23} o Q_{t-24} , T_{t+1-D} o T_{t-D} , y \dot{Q}_t , donde D es el desfase existente entre la ocurrencia de la temperatura máxima y el caudal máximo (6 hrs.).

Se consideran las siguientes:

- Q_t , Q_{t-23} , T_{t+1-D} .(Configuración 0).
- Q_t , Q_{t-23} , T_{t+1-D} , \dot{Q}_t (configuración 1).
- Q_t , Q_{t-24} , T_{t-D} , \dot{Q}_t (configuración 2).

A. Configuraciones de entrada 0, 1, 2

Se modelan diferentes arquitecturas neuronales, con 3 ó 4 neuronas en la capa de entrada dependiendo de las variables de entrada a la red, el número de neuronas en la capa oculta es variable, de 4 a 16 neuronas y una neurona de salida (pronóstico).

Por medio de las configuraciones de entrada planteadas se analiza si el hecho de considerar como variables de entrada el gradiente con respecto al instante anterior y el caudal producido aproximadamente a la hora de pronóstico el día anterior aporta información relevante para el pronóstico.

En la figura 4.70 se presenta la comparación de errores ponderados promedio de todas las crecidas de entrenamiento en función del número de neuronas en la capa oculta y de las épocas de entrenamiento. Se observa que para 13 neuronas en la capa oculta se tiene la envolvente inferior de los errores, se escoge 1500 épocas como número óptimo de épocas de entrenamiento.

Por lo tanto, se elige un modelo neuronal del tipo X (3 ó 4) neuronas en la capa de entrada, 10 neuronas en la capa oculta y una neurona de salida, arquitectura que se utiliza para analizar los errores correspondientes a distintas configuraciones de entrada.

En la tabla 4.98 se presenta la matriz de errores con cada una de las crecidas entrenadas y validadas para el modelo 0, se observa que tanto los errores de entrenamiento como de validación son del orden de $10^{-1} [\text{m}^3/\text{s}]^2$ para los mejores modelos, que se forman a partir del entrenamiento con las crecidas 2 y 6.

En la tabla 4.99 se presenta la matriz de errores para el modelo 1, se observa que el error de entrenamiento es del orden de 1 y los errores de validación en general son del orden de $10^{-1} [\text{m}^3/\text{s}]^2$ para los mejores modelos formados a partir del entrenamiento con las crecidas 2 y 5.

Finalmente, en la tabla 4.100 se presenta la matriz de errores para el modelo 2, se desprende que los mejores modelos se forman a partir del entrenamiento con las crecidas 10, 4 y 2, cuyos errores de entrenamiento y validación son del orden de $10^{-1} [\text{m}^3/\text{s}]^2$.

Si bien, se obtiene un error de entrenamiento de un orden mayor para el modelo 1 formado a partir del entrenamiento con los datos de la crecida 5, los errores de validación en el resto de las crecidas son menores que para los mejores modelos 0 y 2 y consecuentemente el error ponderado lo es también. Luego, el modelo 1 entrenado con los datos de la crecida número 5 es el mejor modelo de pronóstico.

En términos generales no se observan claramente las ventajas de incorporar la variable gradiente de caudal, ni el caudal desfasado en 24 horas con respecto al instante de pronóstico, exceptuando los modelos formados a partir del entrenamiento con la crecida 5 y 3.

Se observa que un 50% de las crecidas presentan mejores validaciones con un modelo que incorpora la variable gradiente de caudal con respecto al instante anterior de pronóstico, el resto mantiene o aumenta sus errores con respecto a un modelo que no lo hace, además la consideración del caudal desfasado en 24 horas con respecto al instante de pronóstico en general no entrega errores menores en comparación a un modelo que no considera dicha variable (Conf.1).

Analizando los errores de acuerdo a la crecida de entrenamiento, se desprende que el análisis de Fourier presenta validez en crecidas de deshielo, donde el comportamiento es cíclico y por lo tanto las componentes de frecuencia relevantes son identificables más claramente, por ende las crecidas con mejor capacidad de generalización (destacadas en amarillo).

Figura 4.70: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.

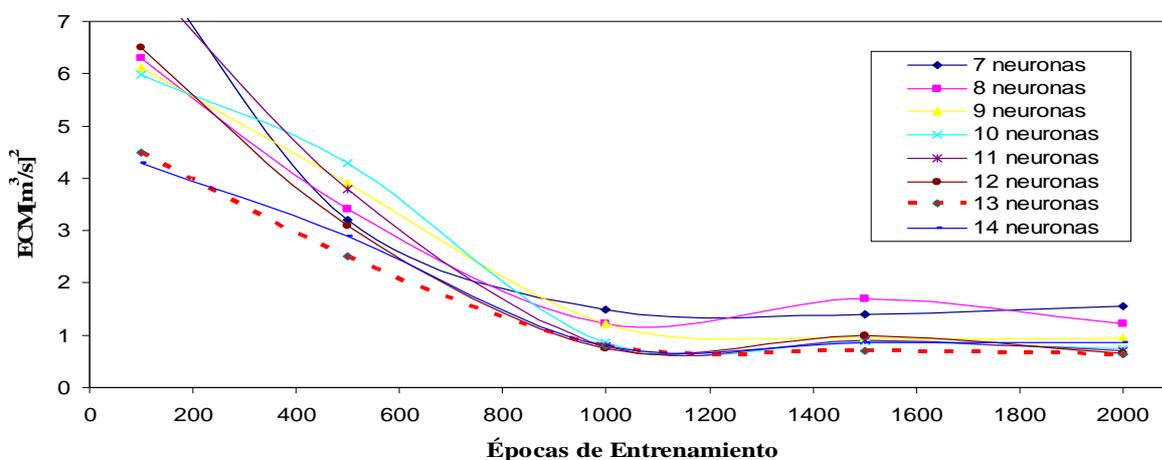


Tabla 4.98: Errores de modelo NEWELM 3-13-1. Conf. 0.

crecida	Nº de datos	Error Entrenam. [m³/s]²	Error Validación [m³/s]²										Error Ponderado [m³/s]²
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	
1	244	0,147	0,147	0,776	0,366	0,751	1,789	0,232	0,870	0,339	0,431	0,528	0,780
2	2979	0,653	0,579	0,653	0,368	0,604	1,300	0,253	1,797	0,405	0,632	0,503	0,667
3	1382	0,547	1,084	1,120	0,547	0,996	1,969	0,398	1,783	0,480	0,580	0,825	1,034
4	1459	0,623	1,051	1,141	0,610	0,623	1,583	0,610	2,716	0,566	0,767	0,717	0,955
5	2510	1,054	1,738	1,798	1,106	1,362	1,054	0,935	2,256	1,153	1,203	1,623	1,358
6	1278	0,126	0,198	0,816	0,311	0,663	1,387	0,126	0,613	0,248	0,410	0,540	0,678
7	98	0,386	2,409	4,621	1,872	4,080	11,163	0,457	0,386	0,690	1,185	1,943	4,261
8	1202	0,319	0,639	1,516	0,667	1,722	5,548	0,206	1,032	0,319	0,518	0,811	1,849
9	291	0,469	0,530	1,631	0,654	1,135	2,592	0,382	1,057	0,401	0,469	1,039	1,296
10	1703	0,495	0,495	0,823	0,536	0,886	1,613	0,345	1,988	0,424	0,626	0,495	0,823

Tabla 4.99: Errores de Entrenamiento y Validación. NEWELM 4-13-1. Conf. 1.

crecida	Nº de datos	Error Entrenam. [m ³ /s] ²	Error Validación [m ³ /s] ²										Error Ponderado [m ³ /s] ²
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	
1	244	0,242	0,242	0,802	0,359	2,943	2,113	4,390	1,860	3,290	2,637	3,513	2,209
2	2979	0,653	0,466	0,653	0,506	0,705	1,125	0,451	1,206	0,528	0,726	0,549	0,691
3	1382	0,361	1,329	1,074	0,361	1,172	1,323	0,706	1,103	0,522	0,714	1,227	0,988
4	1459	0,748	1,008	1,817	0,592	0,748	2,260	0,674	1,490	0,646	0,834	0,970	1,287
5	2510	1,105	0,416	0,696	0,420	0,557	1,105	0,292	1,005	0,394	0,537	0,562	0,639
6	1278	0,206	0,393	0,862	0,462	0,939	1,638	0,206	0,921	0,309	0,584	0,696	0,826
7	98	0,400	1,691	2,642	0,781	2,335	5,715	0,604	0,400	0,773	1,235	1,903	2,469
8	1202	0,263	0,660	0,916	0,401	0,971	2,690	0,231	0,873	0,263	0,520	0,639	1,031
9	291	0,372	0,564	0,834	0,745	0,851	3,665	0,361	0,805	0,436	0,372	1,019	1,293
10	1703	0,531	0,662	0,858	0,695	1,581	1,226	0,562	2,510	0,873	0,908	0,531	0,931

Tabla 4.100: Errores de Entrenamiento y Validación. NEWELM 4-13-1. Conf. 2.

crecida	Nº de datos	Error Entrenam. [m ³ /s] ²	Error Validación [m ³ /s] ²										Error Ponderado [m ³ /s] ²
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	
1	244	0,145	0,145	12,906	4,155	4,516	17,064	4,758	6,218	5,794	3,336	5,692	8,973
2	2979	0,774	0,279	0,774	0,650	0,926	1,487	0,451	0,979	0,606	0,757	0,677	0,847
3	1382	0,390	0,947	1,157	0,390	0,938	1,655	0,471	1,160	0,447	0,828	1,225	1,013
4	1459	0,599	0,209	1,045	0,435	0,599	1,318	0,512	1,097	0,536	0,644	0,626	0,807
5	2510	1,031	1,214	0,743	0,772	0,993	1,031	0,694	2,065	0,689	0,714	0,855	0,852
6	1278	0,166	0,144	0,679	0,313	0,637	1,156	0,166	0,595	0,298	0,442	0,507	0,604
7	98	0,434	0,974	2,097	1,099	2,872	4,399	0,496	0,434	0,636	1,039	1,742	2,126
8	1202	0,341	0,480	1,000	0,470	0,931	1,849	0,279	0,750	0,341	0,524	0,778	0,918
9	291	0,436	0,433	1,449	0,661	1,131	2,733	0,376	0,929	0,456	0,436	0,879	1,262
10	1703	0,511	0,469	0,751	0,496	0,659	1,289	0,453	2,327	0,592	0,712	0,511	0,748

B. Análisis y comparación de Modelos Recurrentes de Elman

En general los errores obtenidos usando este tipo de red neuronal son pequeños y del orden de $10^{-1}[\text{m}^3/\text{s}]^2$ para los mejores modelos, se considera necesaria la incorporación de la variable \dot{Q}_t , dado que el modelo que mejores pronósticos entrega, aquel entrenado con los datos de la crecida 5, presenta los menores errores de pronóstico cuando se considera una configuración de entrada que considera dicha variable.

En la figura 4.71 se presentan los errores ponderados obtenidos de acuerdo a la crecida de entrenamiento, se han unido mediante rectas con el fin de mostrar al modelo Newelm tipo 1 como la envolvente inferior de los errores. La nomenclatura utilizada indica que un modelo Newelm-mod.0 (1 y 2) es aquel que posee una configuración de entrada (0, 1 y 2).

En la figura 4.72 se presentan los errores de validación obtenidos para cada modelo, cada uno de los cuales está formado a partir de la crecida de entrenamiento que entrega los menores errores para cada configuración. Se observan las ventajas de considerar un modelo de este tipo con la configuración de entrada 1.

En la figura 4.73 se presenta esquemáticamente el comportamiento de los errores de validación obtenidos según la crecida de entrenamiento utilizada y para cada una de las configuraciones indicadas. Nuevamente, la curva en sí no posee ningún significado más que facilitar el entendimiento del comportamiento de los errores de validación obtenidos según la crecida con la cual el modelo es entrenado.

La línea punteada un modelo de configuración de entrada 1, línea continua indica un modelo de configuración 2 y línea jaspeada indica un modelo de configuración 0, curvas del mismo color indican la misma crecida validada en uno u otro modelo.

Se desprende que las validaciones son menores en un modelo tipo 1 (Conf.1) entrenado con la crecida 5 y mayores en un modelo tipo 0(Conf.0) entrenado con la crecida 7, en general para todos los modelos (a partir de todas las crecidas de entrenamiento) se observa que el modelo tipo 0 presenta altos errores de validación con respecto a modelos tipo 1 y 2.

Figura 4.71: Errores Ponderados modelos NEWELM.

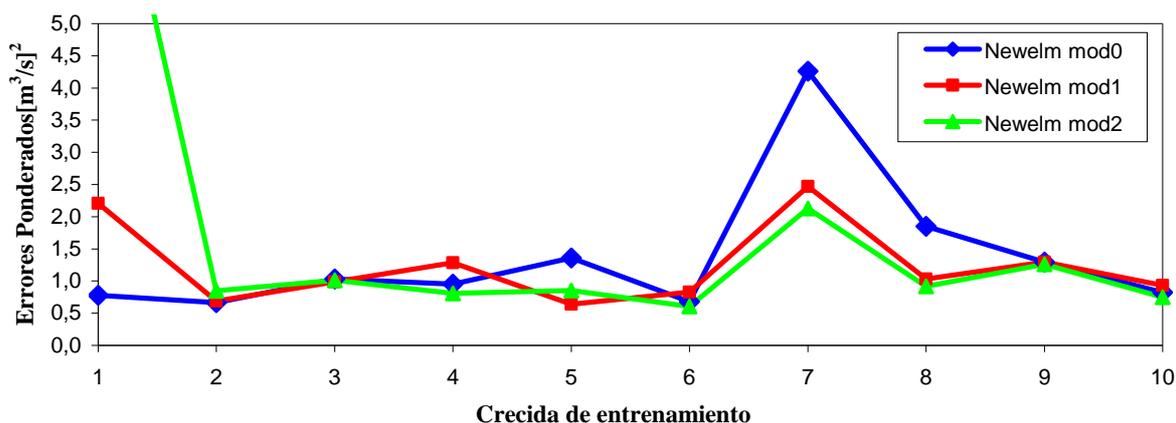


Figura 4.72: Errores de validación modelos NEWELM.

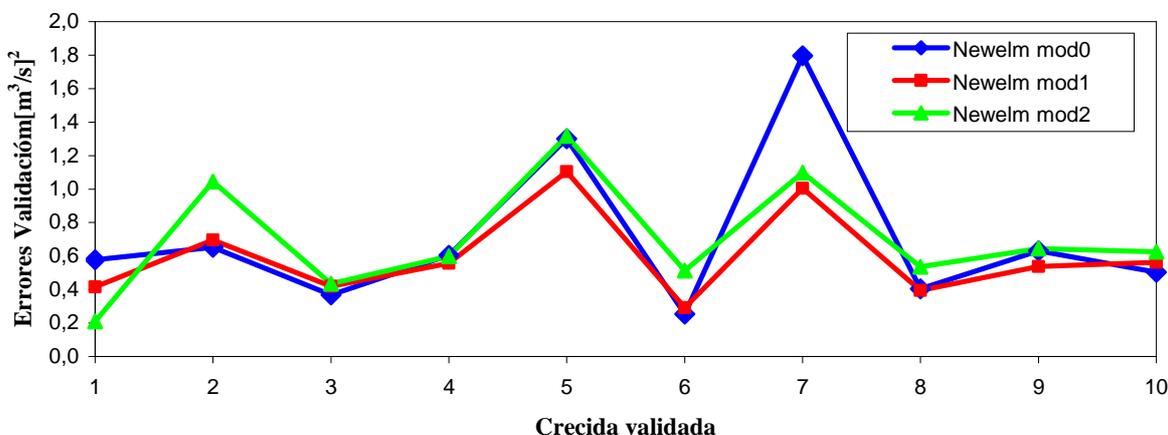
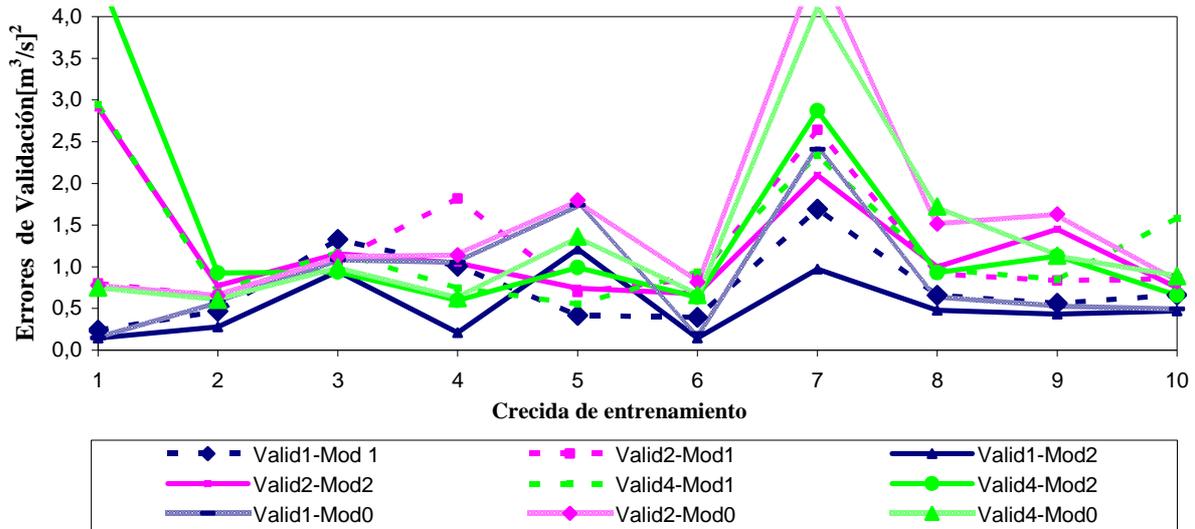


Figura 4.73: Errores de validación modelos NEWELM.



En la tabla 4.101 se presentan los coeficientes de determinación ponderados obtenidos para cada uno de los modelos planteados, se observa que todos los modelos presentan coeficientes sobre 0,9, exceptuando los modelos 1 y 2 entrenados con la crecida 1. No se aprecia que los mejores coeficientes de determinación se tengan para todos los modelos tipo 1 y no existe una tendencia de mejores coeficientes en ningún tipo de modelo, pero para el modelo escogido NEWELM 4-13-1 tipo 1 entrenado con la crecida 5 se obtiene el más alto valor de dicho índice de eficiencia de pronóstico.

Tabla 4.101: Coeficientes de determinación modelos NEWELM.

Modelo(crecida)	r ² ponderado		
	Newelm 3-13-1 Modelo 0	Newelm 4-13-1 Modelo1	Newelm 3-13-1 Modelo 2
1	0,974	0,745	0,543
2	0,974	0,968	0,963
3	0,962	0,947	0,957
4	0,954	0,944	0,962
5	0,915	0,979	0,951
6	0,979	0,971	0,980
7	0,860	0,912	0,932
8	0,948	0,967	0,968
9	0,958	0,953	0,957
10	0,966	0,953	0,963

4.4.2.1.2 Pronóstico mediante Redes Neuronales Prealimentadas

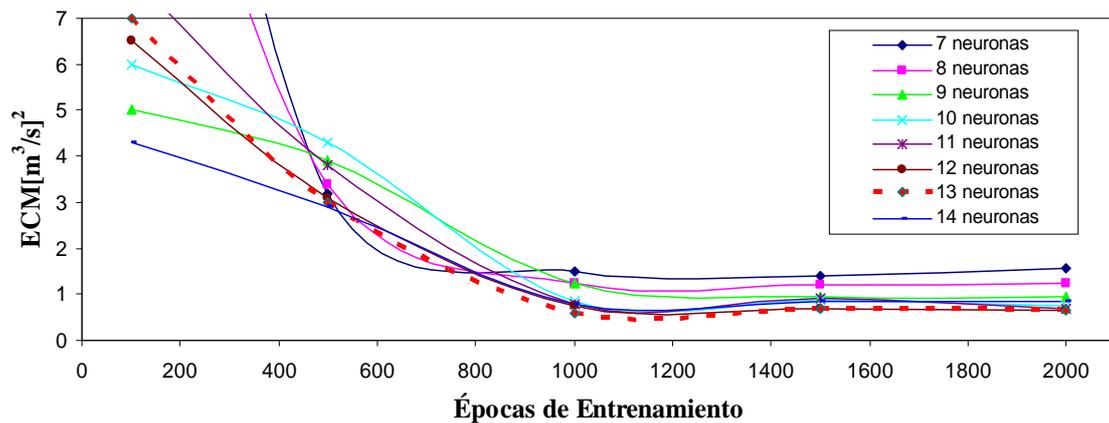
Se plantean modelos neuronales prealimentados con el objetivo de comparar su eficiencia al realizar pronósticos con respecto a modelos neuronales recurrentes.

Las variables de entrada representativas del fenómeno de deshielo y los retardos que se consideran de dichas variables son los siguientes:

- 9 neuronas de entrada $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}$ (Conf. 0).
- 10 neuronas de entrada $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}, \dot{Q}_t$ (Conf. 1).

Se realiza el mismo procedimiento empleado anteriormente para encontrar el número de neuronas y épocas de entrenamiento óptimas, se obtiene al igual que para una RRS un óptimo de 13 neuronas en la capa oculta y 1500 épocas de entrenamiento, según se observa en la figura 4.74.

Figura 4.74: ECM ponderado según número de neuronas en cada capa oculta y épocas de entrenamiento.



A. Configuración de entrada 0

Se obtiene la matriz de errores para la configuración indicada, en la cual se observa que los errores de entrenamiento y validación son del orden de 10^{-1} a $1 \text{ [m}^3/\text{s}]^2$ para los mejores modelos, definidos por el uso de las crecidas 10, 9, 2 y 4 (Tabla 4.102).

Se observa que los mejores modelos siguen siendo los que se obtienen de usar como datos de entrenamiento las crecidas con mayor amplitud de frecuencia (destacadas en amarillo) y que los que mayores errores son obtenidos para aquellos modelos formados a partir de crecidas con pocas y bajas componentes de frecuencia (destacados en verde). En los primeros los errores son del orden de $1 \text{ [m}^3/\text{s}]^2$, aunque la crecida 5 que es la crecida que mayores amplitudes de frecuencias presenta no entrega el más bajo ECM ponderado para este tipo de red y configuración.

El modelo que menor error ponderado presenta y por lo tanto escogido como óptimo para esta configuración de entrada es el modelo formado a partir del entrenamiento con los datos de la crecida 10 (se encuentra dentro de las crecidas con mayor amplitud de frecuencias).

Tabla 4.102: Matriz de errores modelo NEWFF 9-13-1.

crecida	Nº de datos	Error Entrenam. [m ³ /s] ²	Error Validación [m ³ /s] ²										Error Ponderado [m ³ /s] ²
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	
1	244	1,456	1,456	1,187	1,240	1,562	2,445	0,770	1,457	1,007	1,282	1,176	1,425
2	2979	1,342	0,489	1,342	0,796	1,093	1,758	0,551	1,340	0,674	0,933	0,839	1,108
3	1382	0,623	0,706	1,194	0,623	1,128	2,172	0,360	0,928	0,469	0,701	0,971	1,115
4	1459	0,917	0,676	1,226	0,761	0,917	1,894	0,644	1,659	0,700	0,870	0,913	1,110
5	2510	2,245	0,755	1,126	0,887	1,281	2,245	0,585	1,407	0,751	0,948	0,998	1,220
6	1278	1,561	1,414	5,427	1,628	2,099	4,512	6,561	2,837	1,934	1,376	3,614	3,856
7	98	0,664	2,429	5,538	0,854	5,750	6,814	0,355	0,664	0,502	0,762	4,855	4,060
8	1202	0,702	0,644	1,367	0,719	0,874	1,993	0,616	1,625	0,702	0,704	1,054	1,163
9	291	0,538	0,497	1,384	0,619	0,734	1,672	0,645	1,444	0,732	0,538	0,968	1,066
10	1703	0,661	0,728	0,998	0,608	0,857	1,998	0,413	1,676	0,513	0,734	0,861	1,007

B. Configuración de entrada 1

En la tabla 4.103 se presenta la matriz de errores obtenida de considerar un modelo neuronal prealimentado con dicha configuración de entrada y 13 neuronas en la capa oculta, al igual que para la primera configuración el orden de los errores tanto de entrenamiento y como de validación se encuentra en un rango de 10^{-1} a 1 [m³/s]².

No se observa una disminución de los errores de acuerdo a la crecida utilizada en el entrenamiento, por lo tanto, la variable \dot{Q}_t utilizada como parámetro de entrada a la red no contribuye a aminorar los errores de cada modelo.

A nivel general, se observa que el mejor modelo para esta última configuración sigue el comportamiento del análisis de Fourier. Debido a que el modelo formado a partir del entrenamiento con los datos de la crecida 5 presenta el más bajo error ponderado y menor en un muy pequeño porcentaje al modelo elegido para una configuración que no involucra el gradiente de caudal.

Tabla 4.103: Matriz de errores modelo NEWFF 10-13-1.

crecida	Nº de datos	Error Entrenam. [m³/s]²	Error Validación [m³/s]²										Error Ponderado [m³/s]²
			Valid 1	Valid 2	Valid 3	Valid 4	Valid 5	Valid 6	Valid 7	Valid 8	Valid 9	Valid 10	
1	244	1,352	1,352	2,050	1,180	2,113	4,154	0,900	0,928	1,161	1,267	1,606	2,078
2	2979	1,060	0,575	1,060	0,943	1,247	2,039	0,650	1,487	0,798	1,073	0,984	1,176
3	1382	0,606	0,494	1,247	0,606	0,749	1,826	0,577	1,304	0,602	0,669	0,873	1,036
4	1459	0,984	0,967	1,510	0,837	0,984	2,236	0,704	1,865	0,758	0,979	1,204	1,314
5	2510	2,356	0,653	1,155	0,987	1,257	0,704	0,779	1,836	0,949	1,051	1,166	1,002
6	1278	0,360	0,465	1,173	0,663	1,038	2,194	0,360	1,277	0,449	0,675	0,839	1,088
7	98	0,706	2,466	5,377	1,910	5,367	6,856	0,936	0,706	1,519	1,611	3,155	4,049
8	1202	0,500	1,161	1,751	0,815	1,397	3,183	0,519	1,494	0,500	0,846	1,444	1,580
9	291	0,558	0,730	1,639	0,750	0,953	2,374	0,702	1,640	0,800	0,558	1,128	1,335
10	1703	0,657	0,419	0,992	0,679	0,956	1,995	0,469	1,231	0,594	0,800	0,857	1,029

En la tabla 4.104, se presenta el valor del coeficiente de determinación obtenido para cada modelo (formado a partir de la crecida de entrenamiento) y configuración de entrada a la red, no se observa una tendencia clara de valores mejores en uno u otro modelo (Conf. 1 ó 2), pero mayormente, al contrario del resto de los análisis presentados, el modelo tipo 0 posee levemente mejores valores de dicho índice. Para el modelo de interés, aquel entrenado con la crecida 5, el valor de dicho coeficiente permanece estable independientemente de si el modelo considera o no \dot{Q}_t .

Se desprende que la variable gradiente de caudal con respecto al instante anterior de pronóstico no aporta información adicional al modelo prealimentado en términos de medir su eficiencia bajo el índice r^2 , sin embargo, en términos del error cuadrático medio de un modelo entrenado con la crecida que posee mayor amplitud de frecuencias se observan ventajas en la incorporación de dicha variable y considerando que el valor del coeficiente de determinación se mantiene estable, se prefiere su incorporación.

Luego, se elige como modelo NEWFF óptimo un modelo tipo 1 entrenado con la crecida 5.

Tabla 4.104: Coeficiente de determinación modelos NEWFF.

Modelo(crecida)	r² ponderado	
	Newff 9-13-1 Modelo 0	Newff 10-13-1 Modelo1
1	0,937	0,915
2	0,954	0,946
3	0,960	0,954
4	0,945	0,935
5	0,946	0,946
6	0,873	0,960
7	0,825	0,826
8	0,948	0,938
9	0,953	0,943
10	0,962	0,958

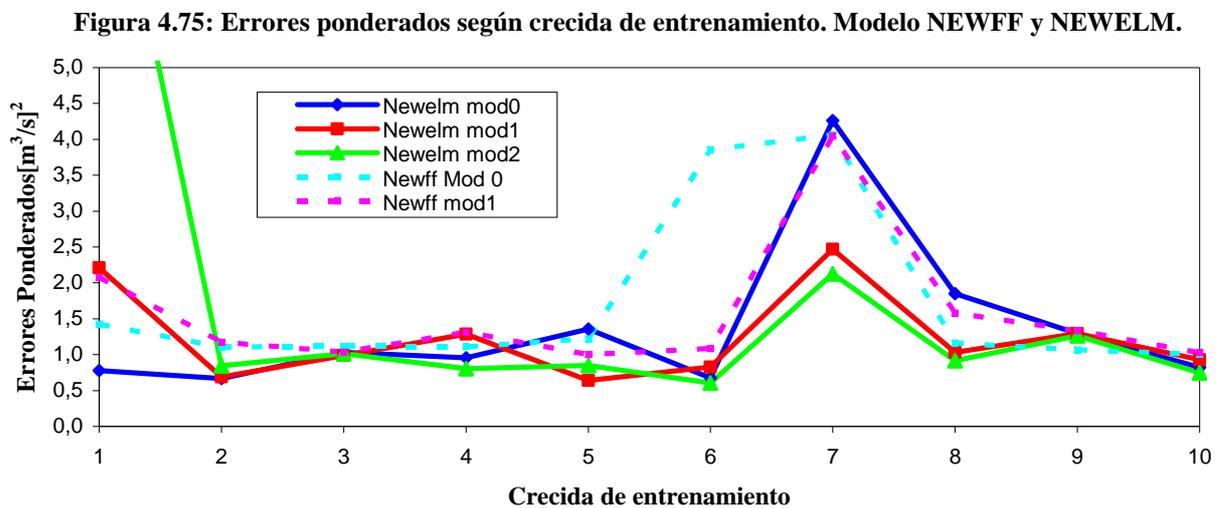
4.4.2.1.3 Análisis y comparación de Modelos Neuronales – Entrenamiento Unitario

A continuación se esquematiza el comportamiento de los errores de ponderados y de validación, así como también de ciertos índices que cuantifican la eficiencia del modelo de pronóstico en el valor máximo.

En la figura 4.75 se observan los errores ponderados de los 4 modelos analizados, la serie en sí unida y representada como una curva no tiene ningún significado físico, sólo es una esquematización de los errores obtenidos.

Se observa, según se desprendió unitariamente que las crecidas 5, 2, 3 y 4 son crecidas que sirven como set de entrenamiento de redes neuronales de Elman para realizar pronósticos, en el caso de redes prealimentadas se añade además, la crecida 10. En cambio, las crecidas 7 y 1, no resultan eficientes, puesto que en todos los modelos planteados a partir de ellas, los errores de validación o pronóstico son superiores al resto.

Se observa que las curvas definidas por los modelos NEWELM 1 y 2 definen la envolvente mínima de los errores, exceptuando el comportamiento que se obtiene al usar la crecida 1 como crecida de entrenamiento. Los modelos NEWFF y NEWELM tipo 0 presentan los más altos errores sobre todo en el caso de ser entrenados con la crecida número 7.



A continuación se presenta gráficamente el comportamiento de los índices descritos en el capítulo 2 en relación al alcance y pronóstico de máximos de caudal para cada uno de los modelos entrenados unitariamente. A modo de ejemplificar el comportamiento de dichos índices en validación se evalúan para el pronóstico de la crecida 6, por crecida entrenada.

En las figuras 4.76 a 4.81, se observa que tanto durante el entrenamiento como durante la validación los modelos 1 y 2 que consideran como variable de entrada el gradiente de caudal (sobre todo el n° 1) de NEWELM presentan los más bajos errores en el pronóstico del valor máximo de la crecida, y son de aproximadamente un 2 % en promedio tanto para el entrenamiento (con todas las crecidas) como en particular para la validación de la crecida 5.

En cuanto al índice de desfase entre ocurrencia de máximos observados y simulados (TMEM), se observa que gracias a la incorporación de la variable gradiente y a las ventajas de las RRS, este valor es en todas las crecidas de entrenamiento aprox. nulo para un modelo NEWELM tipo 1 (Figura 4.78). En la validación de la crecida 6 no se logra tal efecto (Figura 4.81) y se obtienen desfases del orden de -24 hrs. ó -48 hrs. ó 24 hrs. dado que como se explica en el caso de pronósticos de deshielo en la cuenca del Estero Arrayán, el comportamiento de la variable caudal es cíclico y por lo tanto, tienden a haber máximos locales o varios máximos globales (a distintos tiempos con desfases de 24 hrs. dada la periodicidad de la serie).

Se recomienda utilizar las crecidas 5, 3 y 2 como set de entrenamiento de la RNR de Elman con arquitectura 4-13-1., modelos que en promedio presentan índices EMEM, EMRMOBS y TMEM iguales a un 3 %, un 6 % y 2,36 hrs. respectivamente.

Figura 4.76: Índice EMEM entrenamiento. Modelos NEWELM y NEWFF.

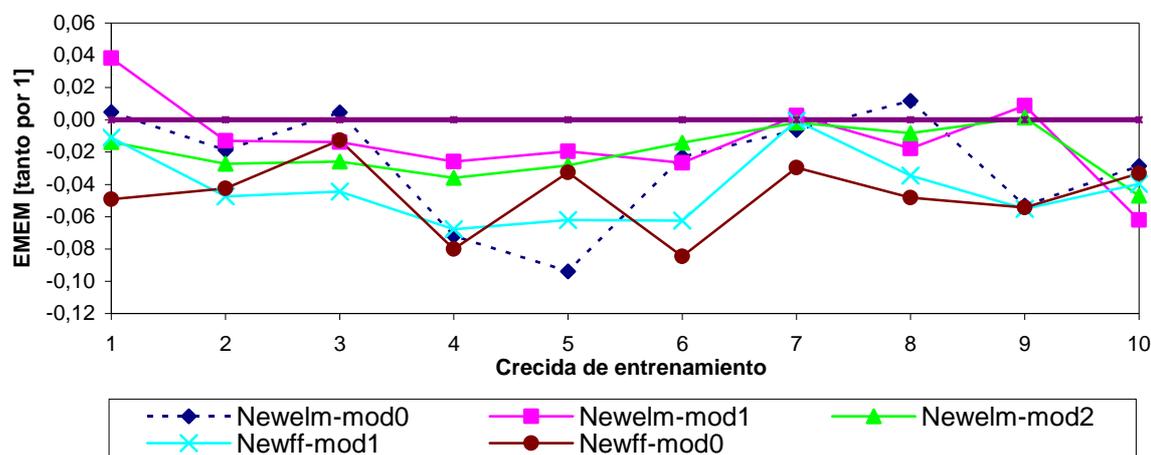


Figura 4.77: Índice EMRMOBS entrenamiento. Modelos NEWELM y NEWFF.

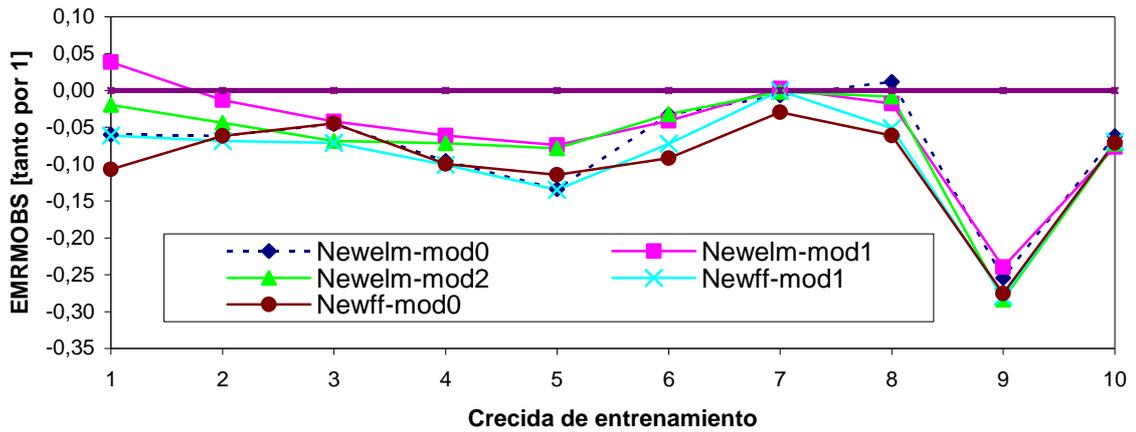


Figura 4.78: Índice TMEM [Hr.] entrenamiento. Modelos NEWELM y NEWFF.

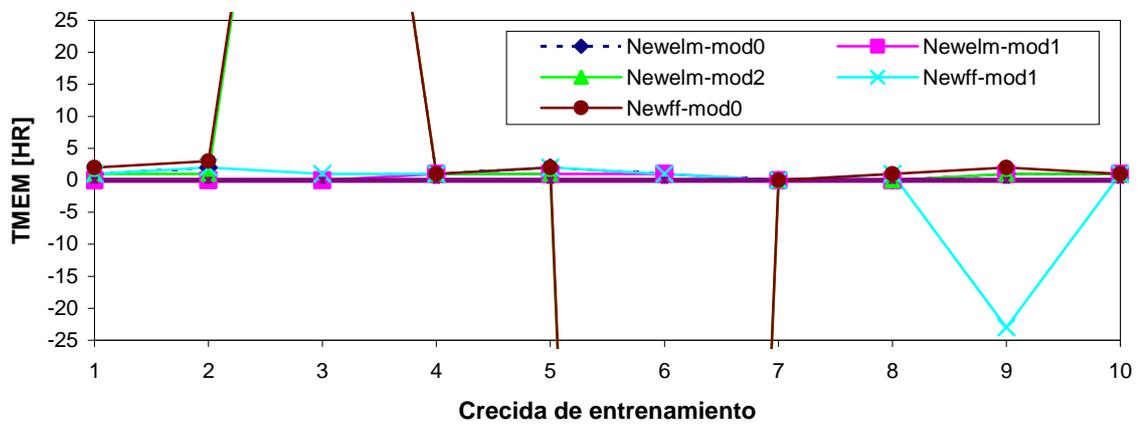


Figura 4.79: Índice EMEM validación crecida 6. Modelos NEWELM y NEWFF.

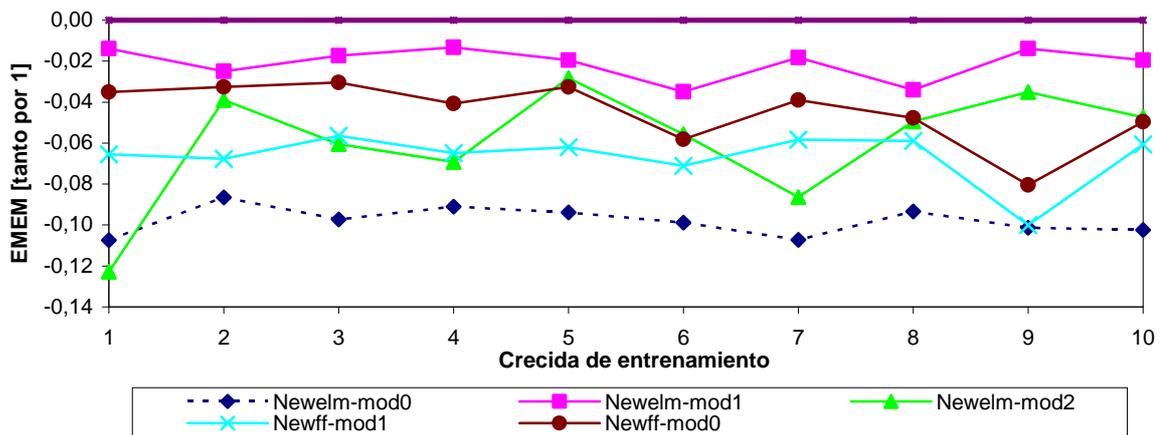


Figura 4.80: Índice EMRMOBS validación crecida 6. Modelos NEWELM y NEWFF.

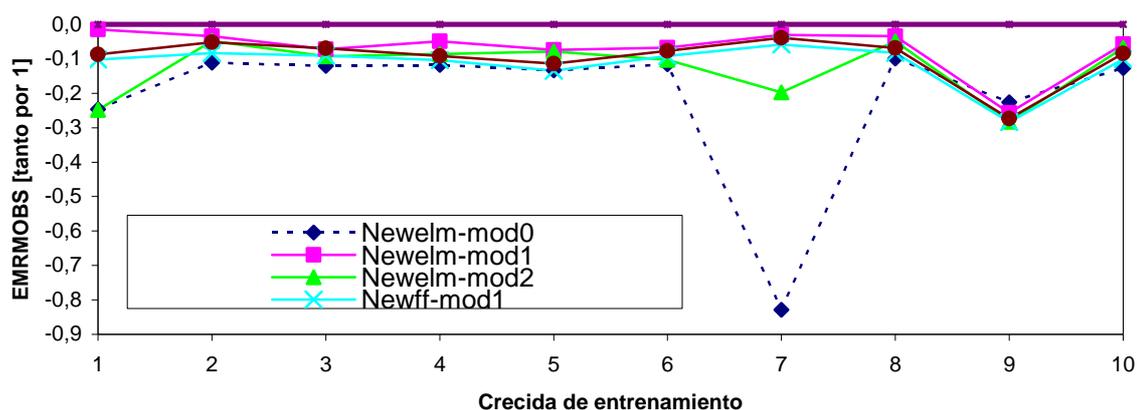
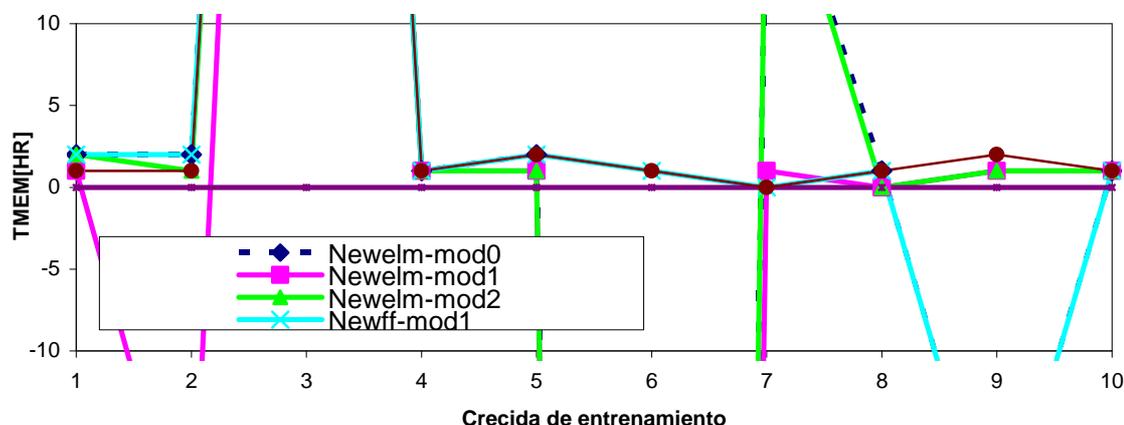


Figura 4.81: Índice TMEM [Hr.] validación crecida 6. Modelos NEWELM y NEWFF.



4.4.2.2 Pronóstico mediante Redes Neuronales – Entrenamiento Continuo

4.4.2.2.1 Pronóstico mediante Redes Neuronales Recurrentes de Elman

Para el análisis continuo de los pronósticos a un paso de adelanto se cuenta con un conjunto de entrenamiento de 6064 datos formado por las crecidas 1 a 4, un conjunto de validación de 3886 datos formado por las crecidas 5 a 7 y finalmente un conjunto de prueba de 3196 datos formado por las crecidas 8 a 10.

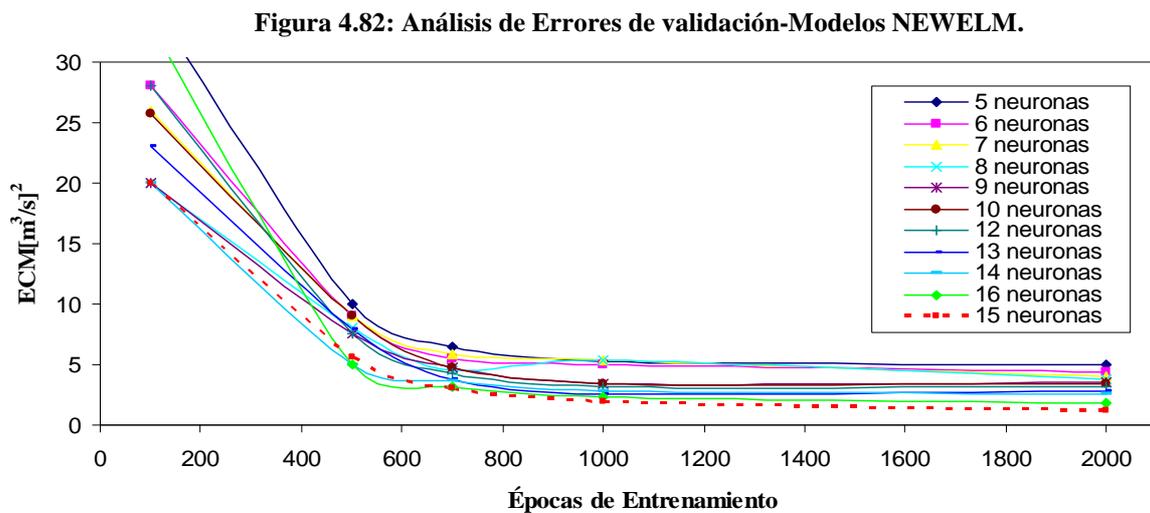
Las características de cada conjunto de datos se presentan en la tabla 4.105.

Tabla 4.105: Caracterización Conjunto de datos.

CRECIDA	MEDIA [m ³ /s]	DESV ESTANDAR [m ³ /s]	Q Máx [m ³ /s]	QMín [m ³ /s]
ENTRENAMIENTO	21,3	7,0	38,1	12,0
VALIDACIÓN	22,5	10,8	38,1	13,2
PRUEBA	19,5	4,9	31,2	13,4

Se consideran como entradas las variables $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$ y además, la precipitación acumulada 14 días antes del pronóstico. Recordando que la naturaleza del fenómeno define modelos de una capa oculta.

En la figura 4.82 se presenta el análisis de validación en función del número de neuronas en la capa oculta y de las épocas de entrenamiento, se deduce que 1500 épocas de entrenamiento son al menos necesarias y que 15 neuronas en la capa oculta forman la envolvente inferior de los errores.



Desde ahora en adelante para este apartado, un modelo NEWELM del tipo 5-X-1 es un modelo que involucra la precipitación acumulada además de las variables $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$, un modelo tipo 4-X-1 es un modelo que no incorpora dicha variable y un modelo 3-X-1 es un modelo que no involucra la variable gradiente de caudal con respecto al instante anterior de pronóstico.

En la tabla 4.106 se presentan los errores obtenidos para distintas configuraciones neuronales, se aprecia que el modelo óptimo se obtiene para una arquitectura de 15 neuronas en la capa oculta y que considera como variables de entrada: $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$

No se aprecia el efecto de incorporación de la variable precipitación acumulada 14 días antes del tiempo de pronóstico si bien el error es uno de los más bajos para los modelos considerados, no supera a un modelo que no incorpora dicha variable en ninguno de los conjuntos considerados. Por otra parte, se aprecia que se obtienen errores mayores para un modelo que no considera la variable \dot{Q}_t en los tres conjuntos considerados.

Luego, se elige como modelo óptimo de red neuronal recurrente un modelo NEWELM4-15-1.

Tabla 4.106: Errores modelo NEWELM Y-X -1.

Configuraciones	ECM [m ³ /s] ²			Ponderado [m ³ /s] ²	r ² entrenamiento	r ² validación	r ² prueba
	entrenamiento	Validación	Prueba				
	6064	3886	3196				
4-4-1	3,589	5,054	2,479	3,752	0,898	0,947	0,839
4-5-1	3,076	4,359	2,143	3,228	0,919	0,958	0,872
4-6-1	2,760	4,003	1,935	2,927	0,930	0,963	0,890
4-7-1	2,553	3,809	1,791	2,739	0,938	0,966	0,901
4-8-1	2,365	3,499	1,702	2,539	0,943	0,969	0,908
4-9-1	2,193	3,410	1,553	2,397	0,949	0,970	0,918
4-10-1	2,166	3,183	1,592	2,327	0,949	0,972	0,917
4-11-1	1,898	2,817	1,382	2,044	0,956	0,976	0,929
4-12-1	1,624	2,569	1,154	1,789	0,964	0,978	0,943
4-13-1	1,716	2,541	1,291	1,856	0,961	0,978	0,936
4-14-1	1,147	1,829	0,882	1,284	0,975	0,984	0,959
4-15-1	0,770	1,272	0,694	0,900	0,984	0,989	0,968
4-16-1	1,050	1,548	0,733	1,120	0,977	0,987	0,966
3-15-1	0,944	1,412	0,862	1,062	0,980	0,987	0,962
3-16-1	1,125	1,467	0,967	1,188	0,978	0,987	0,966
5-15-1	0,954	1,375	0,775	1,035	0,98	0,985	0,967

4.4.2.2.2 Pronóstico mediante Redes Neuronales Prealimentadas

Las entradas a la red quedan definidas por las siguientes 11 variables, representativas de una crecida de deshielo en un modelo neuronal prealimentado:

$$Q_t, Q_{t-1}, Q_{t-2}, Q_{t-25}, Q_{t-24}, Q_{t-23}, T_{t+1-D}, T_{t-D}, T_{t-1-D}, \dot{Q}_t, P_{p14 \text{ días}}$$

Desarrollando la misma metodología de arquitecturas neuronales v/s número de épocas de entrenamiento, se obtiene al igual que antes que el óptimo considera 15 neuronas en cada capa oculta y un óptimo de 1200 épocas de entrenamiento. Figura 4.83.

Desde ahora en adelante y para este apartado un modelo NEWFF 11-X-1 es un modelo que involucra la precipitación acumulada además de las variables “base”, un modelo tipo 10-X-1 es un modelo que no lo hace y un modelo NEWFF 9-X-1 es un modelo que no incorpora como variable de entrada el gradiente de caudal.

En la tabla 4.107 se presentan los errores obtenidos en cada uno de los set de datos para distintas arquitecturas de RNR, donde se aprecia que el número óptimo de neuronas en la capa oculta es precisamente 15 y que el modelo óptimo es un modelo NEWFF 10-15-1.

Figura 4.83: Análisis de Errores de validación-Modelos NEWFF.

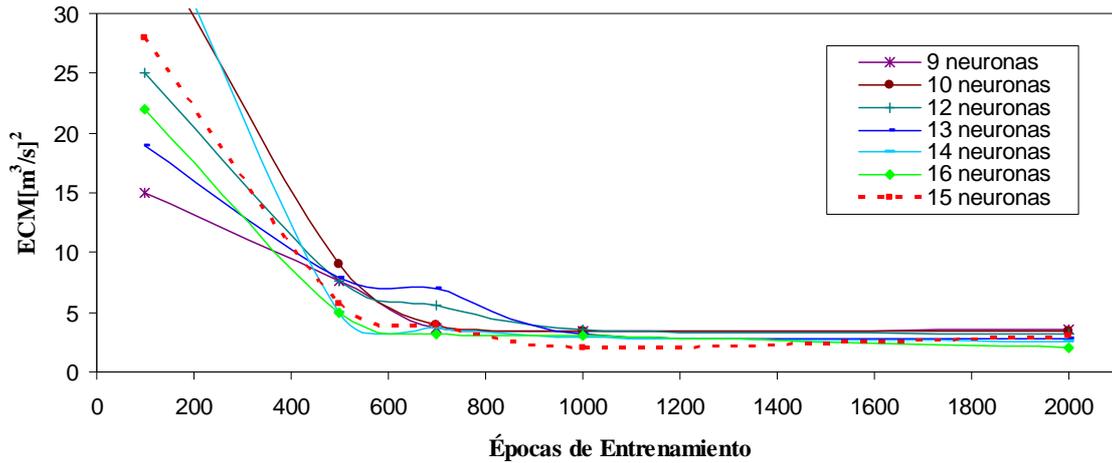


Tabla 4.107: Errores modelo NEWFF Y-X -1.

Configuraciones	ECM [m^3/s^2]			Ponderado [m^3/s^2]	r^2 entrenamiento	r^2 validación	r^2 prueba
	entrenamiento	Validación	Prueba				
	6064	3886	3196				
10-4-1	3,065	4,269	2,189	3,208	0,918	0,958	0,869
10-5-1	2,735	3,864	1,967	2,882	0,930	0,963	0,887
10-6-1	2,493	3,557	1,801	2,639	0,938	0,967	0,900
10-7-1	2,333	3,381	1,688	2,486	0,943	0,969	0,908
10-8-1	2,233	3,389	1,607	2,422	0,947	0,970	0,914
10-9-1	2,076	3,086	1,509	2,237	0,951	0,973	0,920
10-10-1	2,051	3,259	1,466	2,266	0,952	0,971	0,924
10-11-1	1,892	2,862	1,377	2,054	0,956	0,975	0,929
10-12-1	1,834	2,798	1,328	1,996	0,958	0,976	0,932
10-13-1	1,790	2,877	1,280	1,987	0,959	0,975	0,935
10-14-1	1,961	2,745	1,478	2,075	0,955	0,976	0,924
10-15-1	1,633	2,566	1,177	1,798	0,963	0,978	0,941
10-16-1	1,673	2,779	1,193	1,883	0,962	0,976	0,941
9-15-1	1,760	2,911	1,249	1,976	0,960	0,975	0,937
9-16-1	1,640	2,658	1,172	1,827	0,963	0,977	0,941
11-15-1	2,012	2,755	1,525	2,113	0,954	0,976	0,924

Al observar los errores que se obtienen para los mejores modelos a nivel de entrenamientos de tipo continuo de RNR de Elman y RN prealimentadas, se infiere, comparando el valor del error ponderado obtenido para un modelo neuronal de Elman 4-15-1 ($0,9 [m^3/s^2]$) que es aprox. la mitad del error obtenido para un modelo neuronal prealimentado “equivalente” 10-15-1 ($1,8[m^3/s^2]$). Por lo tanto, un modelo NEWELM es un mejor modelo de pronóstico además de a nivel unitario a nivel de entrenamiento continuo.

En la tabla 4.108 se presentan los índices de eficiencia de pronóstico de caudal máximo, se desprende que la red de Elman entrega en general mejores pronósticos de máximos en cada uno de los conjuntos de datos, dicha red subestima los máximos entre un 11 a un 14% del valor máximo observado, en cambio un modelo prealimentado subestima el caudal máximo en aproximadamente un 15% a un 19% del valor máximo observado. Estos valores tanto para la red de Elman como para la red prealimentada son superiores a los obtenidos mediante

entrenamientos de tipo unitario, específicamente mediante el entrenamiento con los datos de la crecida 5.

Tabla 4.108: Índices de eficiencia de pronósticos de caudal máximo.

ÍNDICES	NEWELM 4-15-1			NEWFF 10-15-1		
	Entrenamiento	Validación	Prueba	Entrenamiento	Validación	Prueba
EMEM [tanto por 1]	-0,118	-0,121	-0,119	-0,152	-0,149	-0,151
EMRMOBS [tanto por 1]	-0,129	-0,148	-0,133	-0,158	-0,192	-0,173
TMEM [hr.]	2	0	1	2	2	1

4.4.2.3 Comparación de Modelos Neuronales de Pronóstico por crecidas

En la figura 4.84, a modo de ejemplo, se presentan 2 crecidas validadas con los modelos neuronales NEWELM y NEWFF, elegidos como los mejores modelos. Se puede apreciar gráficamente las ventajas de realizar pronósticos mediante redes neuronales recurrentes de Elman entrenadas con los datos de la crecida 5 sobre modelos de redes prealimentadas entrenadas tanto unitariamente (crecida 5) como continuamente.

En la primera figura 4.84, correspondiente a modelos entrenados unitariamente, se observa que en general la red de Elman pronostica con mayor eficiencia los valores máximos en comparación a una red prealimentada y los alcanza al tiempo observado ya que prácticamente no existen desfases entre los caudales pronosticados y observados, en cambio una red prealimentada tiende a subestimar los caudales máximos, pero igualmente los alcanza en el tiempo observado.

En la segunda figura 4.84, correspondiente a modelos entrenados continuamente, se observa una subestimación considerable de los caudales máximos y mínimos para una red prealimentada, lo cual no sucede para una red recurrente de Elman, en la cual los máximos y mínimos son alcanzados o sobrestimados.

Cabe destacar que dichos resultados son propios de las crecidas analizadas, ya que en promedio para modelos formados a partir del entrenamiento tanto continuo como unitario se tiende a subestimar los caudales máximos, según se destacó anteriormente, aunque dicho porcentaje (máximo 10 % para modelos NEWELM entrenados unitariamente) puede resultar poco visible.

A nivel general por medio de las gráficas, se desprende que el mejor modelo de pronóstico es un modelo neuronal recurrente de Elman de arquitectura 4-15-1 con configuración de entrada $Q_t, Q_{t-23}, T_{t+1-D}, \dot{Q}_t$, entrenada con los datos de la crecida 5.

Figura 4.84: Validación de crecidas según modelo neuronal y tipo de entrenamiento.

Figura 4.84.1: Pronóstico a un paso de adelanto – Crecida 2. Modelos NEWELM y NEWFF- Entre. Unitario.

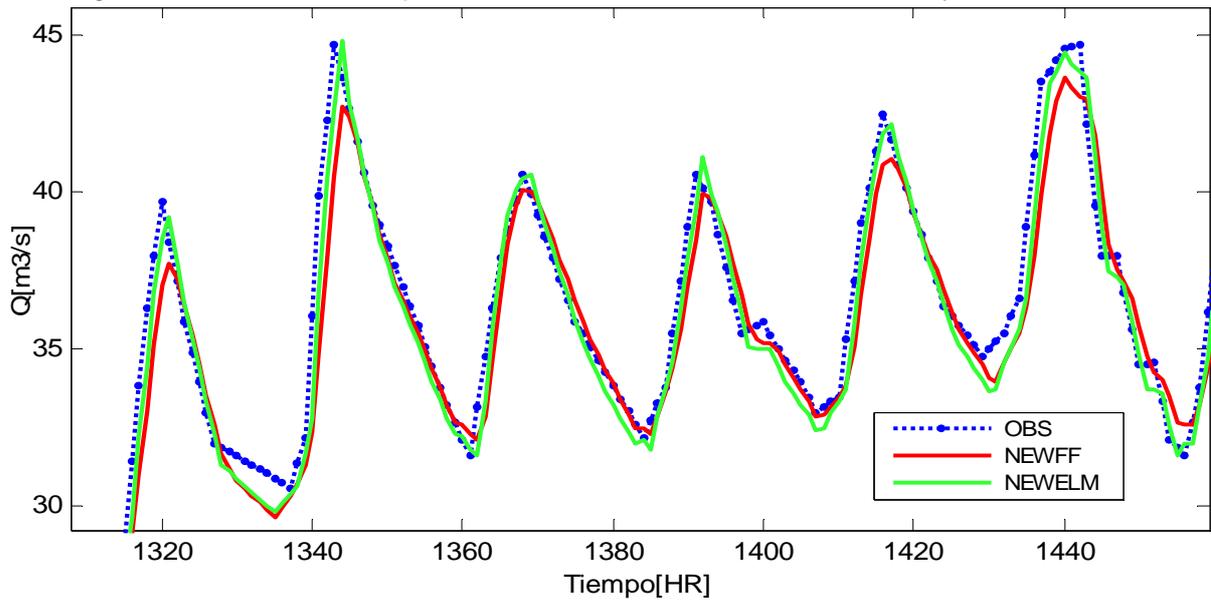
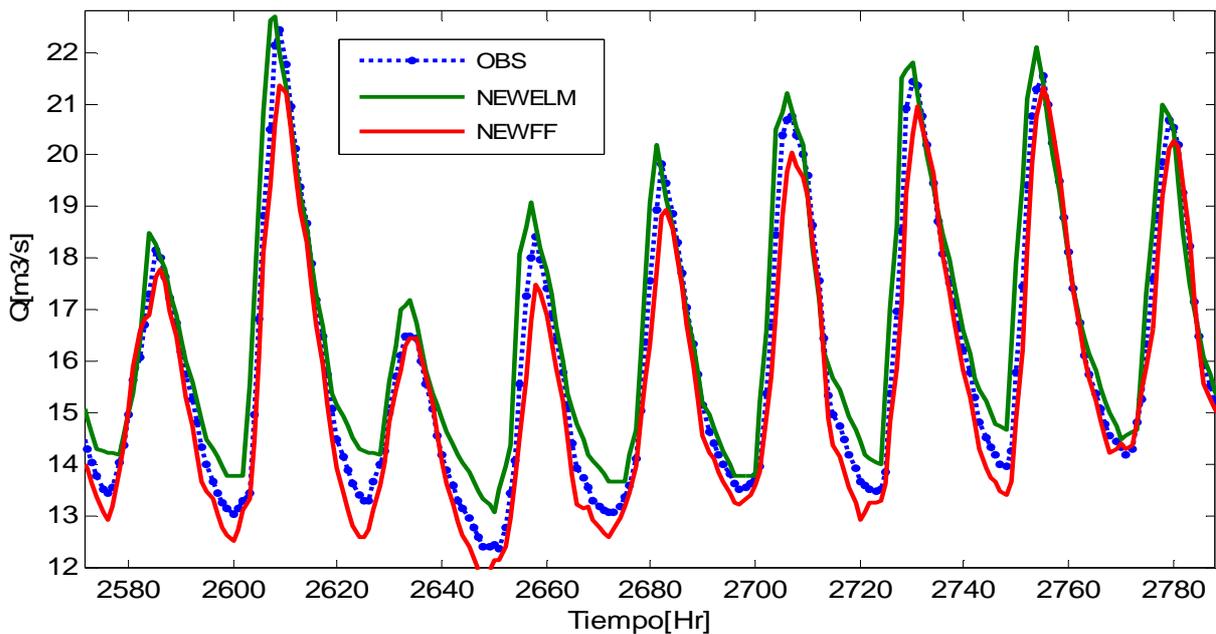


Figura 4.84.2: Pronóstico a un paso de adelanto - Crecida 6. Modelos NEWELM y NEWFF- Entre. Continuo.



4.4.2.4 Pronóstico mediante Modelos Paramétricos

Los tres conjuntos empleados anteriormente se utilizan para realizar un análisis de pronóstico mediante modelos paramétricos tipo ARMAX y RARMAX a una hora de adelanto.

La variable de salida es Q_{t+1} y las variables de entrada: Q_t, T_{T+1-D} y \dot{Q}_t

Los retardos de caudal (na), temperatura (nb) y gradiente de caudal se varían de 1 a 7 hrs. (nb) y el desfase inicial de temperatura (nk) (desfasada ya en $T + 1-D$) desde 1 a 2 hrs.

Al igual que en el caso de crecidas pluviales, mediante pruebas de las variables a considerar se desprende que un modelo que involucra la variable gradiente de caudal como parámetro de entrada no es más eficiente que uno que no lo hace, por lo tanto, en un modelo de tipo paramétrico usado para pronósticos de crecidas de deshielo dicha variable no aporta información adicional al modelo y resulta innecesaria.

En la tabla 4.109 se observan los mejores modelos tipo ARMAX y RARMAX respectivamente, obtenidos de acuerdo a una comparación del error cuadrático medio ponderado, cuyos valores se observan en la tabla 4.110.

Los valores del ECM y el coeficiente de determinación asociado a cada conjunto de datos para ambos modelos se pueden observar en la tabla 4.111 y 4.112 respectivamente.

Se desprende que el mejor modelo de pronóstico es un modelo ARMAX que involucra 7 retardos de la variable caudal y 4 retardos de la variable temperatura a partir de $t + 1 - D$, donde D es el desfase entre los máximos de caudal y temperatura.

Tabla 4.109: Mejores Modelos ARMAX.

Tabla 4.109.a: Modelo ARMAX

Modelo	na	nb	nk
28	7	4	1
49	7	7	1
83	6	5	2
54	5	1	2
84	7	5	2
75	5	4	2
76	6	4	2

Tabla 4.109.b: Modelo RARMAX

Modelo	na	nb	nk
16	2	3	1
22	1	4	1
15	1	3	1
23	2	4	1
37	2	6	1
44	2	7	1
8	1	2	1

Tabla 4.110: Error Cuadrático Medio Ponderado

Tabla 4.110.a: Modelo ARMAX.

Modelo	ECM ponderado $[\text{m}^3/\text{s}]^2$	R^2 ponderado
28	0,93	0,99
49	0,93	0,99
83	0,93	0,99
54	0,94	0,99
84	0,94	0,99
75	0,94	0,99
76	0,94	0,99

Tabla 4.110.b: Modelo RARMAX.

Modelo	ECM ponderado $[\text{m}^3/\text{s}]^2$	R^2 ponderado
16	0,95	0,98
22	0,95	0,98
15	0,95	0,98
23	0,98	0,97
37	0,98	0,97
44	0,98	0,97
8	1,04	0,97

Tabla 4.111: Error cuadrático medio Modelos ARMAX.

Modelo	R^2 entrenamiento	R^2 validacion	R^2 prueba	ECM entrenamiento $[\text{m}^3/\text{s}]^2$	ECM Validación $[\text{m}^3/\text{s}]^2$	ECM Prueba $[\text{m}^3/\text{s}]^2$
28	0,99	0,97	0,99	0,91	0,90	0,98
49	0,99	0,96	0,98	0,91	0,90	0,98
83	0,99	0,99	0,98	0,91	0,91	0,98
54	0,99	0,99	0,98	0,91	0,91	0,98
84	0,99	0,99	0,98	0,91	0,91	0,98
75	0,99	0,99	0,98	0,91	0,91	1,00
76	0,99	0,99	0,98	0,91	0,91	1,00

Tabla 4.112: Error cuadrático medio Modelos RARMAX.

Modelo	R ² entrenamiento	R ² validacion	R ² prueba	ECM entrenamiento [m ³ /s] ²	ECM Validación [m ³ /s] ²	ECM Prueba [m ³ /s] ²
16	0,98	0,98	0,98	0,91	0,95	0,99
22	0,98	0,98	0,98	0,92	0,95	0,99
15	0,98	0,98	0,98	0,92	0,95	0,99
23	0,97	0,97	0,97	0,93	0,97	1,03
37	0,97	0,97	0,97	0,93	0,97	1,03
44	0,97	0,97	0,97	0,93	0,97	1,03
8	0,96	0,97	0,96	0,93	0,99	1,19

Un modelo ARMAX presenta mejores pronósticos que un modelo RARMAX, pero la diferencia es mínima y se debe principalmente al aumento de los errores tanto de entrenamiento como de validación y prueba que se comenten en el segundo modelo con respecto al primero. Ambos modelos entregan errores similares y del mismo orden de magnitud, en términos del coeficiente de determinación, si bien es mejor un modelo ARMAX, ambos modelos son recomendables como modelos de pronóstico de deshielo (r^2 prácticamente 1).

ARMAX y RARMAX entregan errores de pronóstico del mismo orden que los errores que se obtienen por medio de redes neuronales tanto recurrentes como prealimentadas entrenadas continuamente, pero si se comparan los ECM obtenidos con los entregados por un modelo de red recurrente entrenada unitariamente, los errores son mayores aprox. en un 40 %.

Se realiza también un análisis unitario, donde cada modelo tiene la mejor configuración obtenida para ARMAX y RARMAX “continuo” y es entrenado de acuerdo a los datos de cada crecida, de tal manera que se forman 10 modelos. En la tabla 4.113 se presentan los errores ponderados obtenidos para cada modelo. Los mejores de cada tipo se destacan en rosado y los peores en amarillo.

Tabla 4.113: Errores Ponderados Modelo ARMAX y RARMAX.

Crecida	Error Ponderado [m ³ /s] ²	
	ARMAX	RARMAX
1	0,727	0,978
2	0,716	0,978
3	0,713	1,068
4	0,641	1,173
5	0,640	0,999
6	0,642	0,978
7	0,658	0,978
8	0,695	0,978
9	0,657	1,085
10	0,613	0,980

Se observa que las crecidas con mayor amplitud de frecuencias forman los mejores modelos ARMAX (5, 4 y 10), en cambio, las crecidas con menores amplitudes frecuenciales (7, 8 y 1) forman los mejores modelos RARMAX, las razones son las mismas que se han justificado para

crecidas pluviales en esta cuenca, punto 4.3.1.4, concernientes al método de ajuste de los parámetros de cada modelo.

Se desprende que modelos RARMAX entrenados continuamente entregan resultados similares a aquellos obtenidos a través de entrenamientos de tipo unitario, en el caso de modelos ARMAX se observan errores menores para entrenamientos de tipo unitario.

En general para modelos ARMAX se obtienen errores similares usando cualquier crecida como set de entrenamiento, aunque el desempeño es mucho más equilibrado en un modelo RARMAX, donde cualquier crecida que sea utilizada como conjunto de datos de entrenamiento tiene el mismo efecto en término de los errores.

En cuanto a los valores del ECM, un modelo de redes neuronales recurrentes NEWELM 4-13-1 con entrenamiento unitario (crecida 5) entrega errores del mismo orden que un modelo ARMAX, pero en el caso de ser comparado con modelos RARMAX, el modelo de redes recurrentes es un mejor modelo de pronóstico en aprox. un 40 % (porcentaje calculado sobre el valor de los errores ponderados de cada modelo).

4.4.2.5 Resumen y análisis de modelos de pronóstico a un paso de adelanto

Esta sección tiene por objetivo, situar al lector en el desarrollo de los pronósticos realizados hasta el momento, a fin de continuar con pronósticos a varios pasos de adelanto.

Se realizan análisis entre los mejores modelos de pronóstico de cada tipo analizado. Para comparar modelos neuronales entrenados unitariamente se considera como crecida de entrenamiento la crecida 5 (recordar, crecida con mayor amplitud frecuencial y demostrada su mayor capacidad de generalización). Se comparan modelos de redes neuronales recurrentes de Elman y de redes prealimentadas a nivel de entrenamiento unitario y continuo y además, modelos paramétricos, ARMAX y RARMAX [na, nb, nk].

Para estos últimos modelos se incluyen sólo el modelo ARMAX unitario y el modelo RARMAX continuo, debido a las razones indicadas precedentemente (punto 4.4.2.4) que los señalan como los mejores modelos de pronóstico de este tipo.

Los parámetros de comparación utilizados son el ECM ponderado, el ECM de validación por crecida y los índices de error en el pronóstico de máximos. Los cuales se presentan en la tabla 4.114.a ,4.114.b y 4.114.c respectivamente.

De la tabla 4.114.a se infiere que tanto modelos neuronales recurrentes de Elman y un modelo paramétrico ARMAX con entrenamientos unitarios resultan los mejores modelos de pronóstico.

Se observan aumentos de los errores para entrenamientos de tipo continuo con respecto a los de tipo unitario, exceptuando el modelo RARMAX, dichos aumentos son leves y tales que el orden de magnitud se mantiene. En términos generales resulta más eficiente realizar entrenamientos de tipo unitario (crecida 5) con un modelo NEWELM 4-13-1.

Es destacable el hecho de que a nivel continuo un modelo paramétrico pueda resultar más eficiente (c/r al ECM) que un modelo neuronal prealimentado.

Tabla 4.114.a: Errores ponderados según modelo de pronóstico.

ECM PONDERADO	Entrenamiento	
	Unitario	Continuo
NEWELM	0,639	0,900
NEWFF	1,007	1,798
ARMAX	0,64	0,930
RARMAX	0,980	0,950

En la tabla 4.114.b se presentan los errores de validación obtenidos para cada crecida, se pueden apreciar gráficamente en la figura 4.85. Desde este análisis de eficiencia segmentado por crecidas, se deduce que tanto un modelo ARMAX como un modelo neuronal recurrente de Elman y también un modelo prealimentado entrenado unitariamente son buenos modelos de pronóstico. Cabe destacar que el modelo ARMAX presenta errores mayores, en relación al modelo de Elman, en aquellas crecidas que poseen mayor capacidad de generalización y que pueden ser también las más caudalosas y más difíciles de controlar. Además, con respecto al modelo prealimentado con un entrenamiento del mismo tipo, se observa que el modelo de Elman sigue siendo el más eficiente.

En la figura 4.85 la curva en sí no posee ningún significado y sólo pretende encontrar gráficamente una envolvente inferior de errores. Analizando los modelos neuronales entrenados continuamente, se observa una envolvente clara dada por el modelo NEWELM 4-15-1, debido a que arroja menores errores en todas las crecidas validadas con respecto al modelo NEWFF 10-15-1. Por lo tanto, una red de Elman entrenada continuamente es evidentemente un mejor modelo de pronóstico frente a un modelo neuronal prealimentado con un entrenamiento del mismo tipo.

En la tabla 4.114.c se presenta el error que se produce en el pronóstico de máximos por medio de los dos modelos neuronales analizados. El error es cuantificado para el máximo de cada crecida y luego ponderado, al igual que las tablas que presentan dichos índices precedentemente.

Se deduce que un modelo NEWELM 4-13-1 resulta el modelo más eficiente para pronosticar máximos mediante redes neuronales, debido a que presenta los menores errores.

Subestimaciones de aprox. 3%, valor que crece a aprox. 8% en el momento de ocurrencia real del caudal máximo.

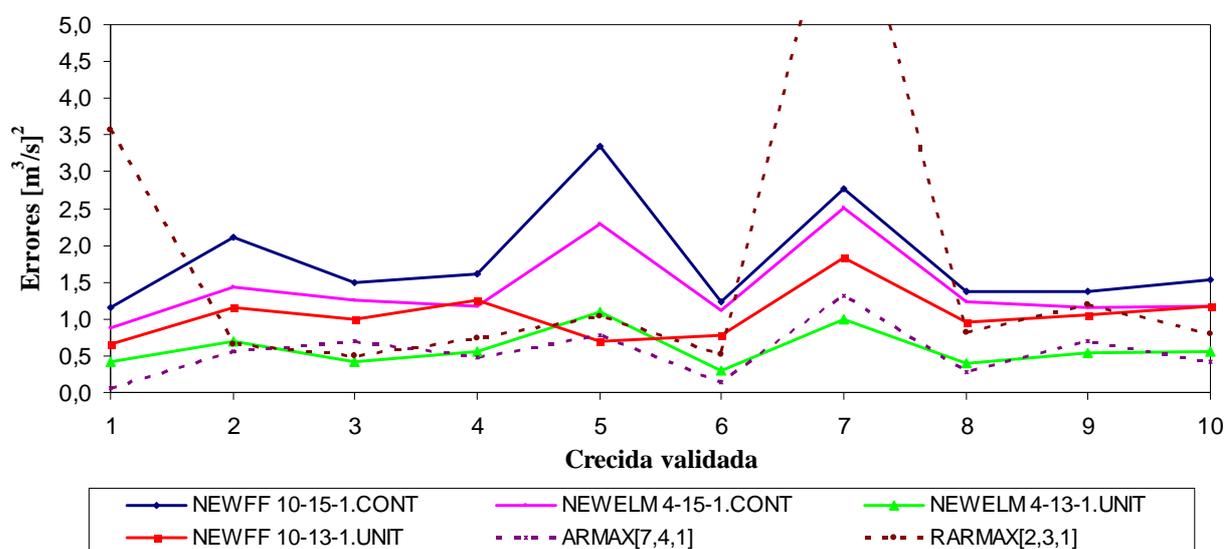
Tabla 4.114.b: Errores de validación por crecida según modelo de pronóstico.

Crecida	ECM de validación [m ³ /s] ²					
	Unitario		Continuo		ARMAX [7,4,1]	RARMAX[2,3,1]
	NEWELM 4-13-1	NEWFF 10-13-1	NEWELM 4-15-1	NEWFF 10-15-1		
1	0,416	0,653	0,878	1,163	0,061	3,571
2	0,696	1,155	1,428	2,102	0,566	0,657
3	0,420	0,987	1,257	1,493	0,697	0,502
4	0,557	1,257	1,174	1,609	0,473	0,730
5	1,105	0,704	2,283	3,338	0,781	1,045
6	0,292	0,779	1,113	1,243	0,146	0,526
7	1,005	1,836	2,518	2,761	1,320	7,540
8	0,394	0,949	1,243	1,370	0,275	0,809
9	0,537	1,051	1,164	1,380	0,689	1,202
10	0,562	1,166	1,178	1,525	0,423	0,795

Tabla 4.114.c: Índices de error en el pronóstico de caudales máximos según modelo de pronóstico

ÍNDICES DE ERROR	Entrenamiento					
	Unitario			Continuo		
	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]	EMEM [tanto por 1]	EMRMOBS [tanto por 1]	TMEM [Hr.]
NEWELM	-0,030	-0,077	0,721	-0,119	-0,136	1,166
NEWFF	-0,045	-0,106	1,900	-0,151	-0,172	1,757

Figura 4.85: Análisis de Errores de validación según el modelo de pronóstico.

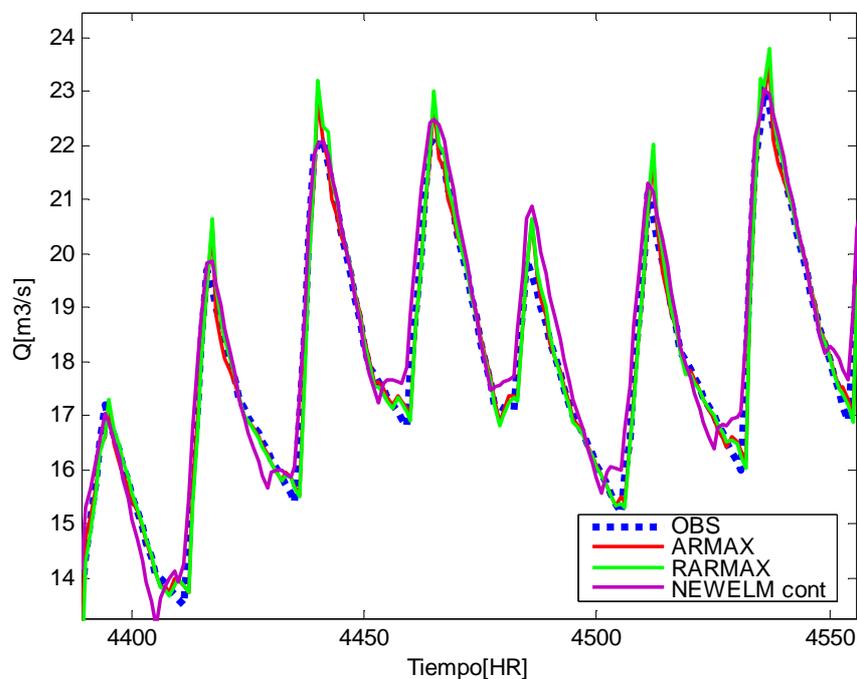


En la figura 4.86 se observan las ventajas del modelo neuronal recurrente frente a modelos de tipo paramétrico ARMAX y RARMAX, los tres formados a partir de entrenamientos de tipo continuo. Se presenta el comportamiento de los pronósticos en un sector característico del conjunto de validación.

Se observa que el modelo NEWELM 4-15-1 entrenado continuamente pronostica casi con exactitud el caudal máximo y al tiempo en que se produce realmente, lo cual lo aventaja sobre

cualquiera de los otros métodos, pero sobrestima los mínimos de caudal. Un modelo tipo RARMAX en cambio, presenta grandes sobrestimaciones del caudal máximo y en menor cantidad un modelo ARMAX, que a su vez presenta un buen pronóstico general.

Figura 4.86: Análisis de Modelos en el conjunto de validación.



Considerando todos los índices analizados el mejor modelo de pronóstico a un paso de adelanto es un modelo NEWELM 4-13-1 entrenado con los datos de la crecida 5. Cabe señalar que no se debe desmerecer la eficiencia de los otros métodos, que poseen teorías y funcionamientos más simples que requieren un menor tiempo de procesamiento de datos, como es el caso del modelo ARMAX, lo cual los aventaja sobre la complejidad de la red recurrente de Elman.

Antes de continuar, al igual que se indicó en pronósticos de deshielo en la cuenca del Estero Arrayán, cabe señalar que no se ha considerado realizar pronósticos mediante la red recurrente de Hopfield en este tipo de crecidas. Las razones radican en la alta eficiencia de los modelos estudiados hasta el momento, se ha logrado mejorar los pronósticos mediante el uso de la red recurrente de Elman, con respecto a modelos ya existentes y en forma notoria en la estimación de los valores máximos tanto en cantidad como en calidad.

Dado el comportamiento cíclico de éste tipo de crecidas, la gran cantidad de registros que las caracterizan y el hecho de que el modelo de Hopfield se base en la identificación de un gradiente, resulta difícil aplicar el modelo de Hopfield, tanto por la cantidad de información a almacenar, que debe ser codificada y que además genera una gran cantidad de patrones, lo que

hace difícil garantizar el correcto funcionamiento de la red, dado que el número de patrones almacenado no debe ser excesivamente mayor a la cantidad de neuronas de la red, ver punto 2.2.4.1, sumado al hecho de que la variabilidad de la crecida hace que el gradiente deje de ser representativo a varios pasos de adelanto.

4.4.2.6 Pronósticos a varios pasos de adelanto

Se realiza pronósticos de caudales a 6, 12, 24 y 48 horas considerando dos tipos de modelos uno que considera como conjunto de datos de entrenamiento sólo la crecida número 5 y otro que considera como dicho conjunto, aquel empleado en el análisis de tipo continuo.

Para elegir la arquitectura óptima de los modelos en función de minimizar el ECM, se prueban nuevamente diferentes números de neuronas en la capa oculta para cada modelo neuronal, definido de acuerdo a la salida que se desea. Las variables de entrada de los modelos neuronales son las mismas que se consideran en los mejores modelos elegidos para un paso de adelanto.

Luego, los modelos neuronales y las correspondientes arquitecturas que se utilizan para realizar pronósticos a $t + x$ horas de adelanto se presentan en la figura 4.115.

Tabla 4.115: Arquitecturas óptimas de red.

Modelo	t+6		t+12		t+24		t+48	
	Continuo	unitario	Continuo	unitario	Continuo	unitario	Continuo	unitario
NEWELM	4-15-1	4-16-1	4-15-1	4-13-1	4-16-1	4-13-1	4-16-1	4-15-1
NEWFF	10-15-1	4-16-1	10-15-1	4-16-1	4-16-1	4-13-1	4-16-1	4-16-1

En la tabla 4.116 se presenta los errores de entrenamiento y ponderados de pronósticos realizados por medio de modelos neuronales de Elman y prealimentados considerando entrenamientos de tipo continuo. Se desprende que los errores ponderados obtenidos para una red de tipo recurrente son menores a los de una red prealimentada, y suficientes para indicar que la RRS entrega un mejor modelo de pronóstico.

En la tabla 4.117 se presenta los errores de entrenamiento y ponderados obtenidos de realizar pronósticos de acuerdo al análisis unitario por medio de la crecida 5. Se observa que los errores, al igual que a nivel continuo son menores utilizando un modelo de red recurrente de Elman.

Se observa que los errores de entrenamiento aumentan a nivel unitario con respecto a un entrenamiento continuo, pero ponderadamente los errores para ambos tipos de entrenamiento son

muy similares, por lo tanto, un modelo entrenado unitariamente presenta una capacidad de generalización más clara.

Un resultado importante que se infiere de las tablas 4.116 y 4.117 es que en el caso de pronósticos de deshielo para la cuenca en cuestión, redes de Elman entrenadas unitaria y continuamente presentan desempeños ponderados muy similares. La explicación radica en que las crecidas de deshielo son más duraderas que las pluviales y por lo tanto, los parámetros de la red logran ajustarse mejor, puesto que se produce un mayor procesamiento de la información de la crecida, aunque la serie sea de “crecidas pegadas”, de esta manera se descarta más la información antecedente en aquellos puntos de unión de crecidas.

Llama la atención el encontrar errores menores para pronósticos a 24 horas en relación a pronósticos a 12 horas, lo cual puede deberse al comportamiento cíclico de período 24 horas.

Tabla 4.116: ECM entrenamiento y Ponderado. Modelos NEWELM y NEWFF. Análisis Continuo.

Pronóstico	Entrenamiento continuo			
	Modelo NEWELM		Modelo NEWFF	
	Error de entrenamiento	Error Ponderado [m ³ /s] ²	Error de entrenamiento	Error Ponderado [m ³ /s] ³
t+6	6,647	7,601	7,429	8,267
t+12	10,770	13,904	13,417	14,343
t+24	5,941	6,841	6,975	7,548
t+48	12,590	11,871	13,377	12,961

Tabla 4.117: ECM entrenamiento y Ponderado. Modelos NEWELM y NEWFF. Análisis unitario.

Pronóstico	Entrenamiento unitario			
	Modelo NEWELM		Modelo NEWFF	
	Error de entrenamiento	Error Ponderado [m ³ /s] ²	Error de entrenamiento	Error Ponderado [m ³ /s] ³
t+6	10,659	7,318	11,557	7,600
t+12	16,777	13,788	20,750	13,988
t+24	11,444	6,585	12,707	7,423
t+48	19,471	10,749	21,642	11,161

Para la comparación en términos de eficiencia de modelos neuronales con modelos de tipo paramétrico, se utiliza el mejor modelo de estos últimos, es decir un modelo ARMAX entrenado con los datos de la crecida 5 y con la configuración óptima encontrada para un paso de adelanto. Los valores de los ECM obtenidos para distintos tiempos de adelanto se presentan en la tabla 4.118. Se observa que los errores obtenidos por medio de este modelo son mayores que los entregados por un modelo neuronal recurrente de Elman, pero destacándose que son del mismo orden de magnitud.

Tabla 4.118: ECM Ponderado. Modelo ARMAX.

Modelo ARMAX (unitario)	Error Ponderado [m ³ /s] ²			
	t+6	t+12	t+24	t+48
	8,456	14,710	8,670	15,730

En la tabla 4.119 se presentan los caudales máximos observados y pronosticados para una red de Elman entrenada con los datos de la crecida 5. Se deduce la factibilidad de realizar pronósticos hasta 48 hrs.

El error promedio cometido en pronósticos a 48 hrs. por medio de la red NEWELM es de aprox. el 17% del caudal medio máximo de todas las crecidas analizadas.

Tabla 4.119: Caudal máximo observado y pronosticado a t+x hrs. Modelo NEWELM. Análisis unitario.

Crecida	Q pronosticado[m ³ /s]				Q observado [m ³ /s]
	t+6	t+12	t+24	t+48	
1	24,7	19,7	22,4	19,4	24,4
2	41,5	41,6	42,7	38,4	47,1
3	27,3	24,7	28,0	24,4	30,5
4	36,2	32,3	38,6	33,4	42,4
5	57,3	53,3	58,0	52,4	63,8
6	21,7	22,5	22,2	19,7	24,6
7	19,7	16,7	22,6	18,6	24,0
8	25,2	24,6	27,4	22,9	29,6
9	20,6	20,7	21,5	19,4	24,3
10	36,4	35,4	35,9	31,9	39,8

En las tablas 4.120 y 4.121 se presentan los índices de eficiencia de pronósticos de caudal máximo hasta 48 horas. Se deduce que un modelo neuronal de Elman realiza mejores pronósticos de máximos en cuanto a valor y sobre todo para el modelo unitario que considera como set de entrenamiento la crecida número 5. El índice de desfase en sí no tiene mucha validez dado que se trata de un fenómeno cíclico y por lo tanto existen máximos locales cercanos al máximo global o varios máximos globales.

Tabla 4.120: Índices de pronóstico de caudal máximo. Análisis Continuo.

ÍNDICES	Análisis continuo	
	NEWELM	NEWFF
EMEM [tanto por 1]	-0,145	-0,176
EMRMOBS [tanto por 1]	-0,195	-0,293
TMEM [hr.]	24	-2,525

Tabla 4.121: Índices de pronóstico de caudal máximo. Análisis unitario.

ÍNDICES	Análisis unitario	
	NEWELM	NEWFF
EMEM [tanto por 1]	-0,110	-0,128
EMRMOBS [tanto por 1]	-0,183	-0,267
TMEM [hr.]	24	-3

Capítulo 5

5 Conclusiones y Recomendaciones

En términos generales, para pronosticar crecidas pluviales y de deshielo las redes neuronales recurrentes de Elman resultan más eficientes que las redes prealimentadas y que los modelos paramétricos, lo cual es más notorio en pronósticos pluviales. Las razones que justifican que los modelos RRS sean más eficientes son considerables, la primera y fundamental viene dada por el hecho de almacenar información de los patrones de entrada, de tal manera que no se hace necesaria la incorporación de variables retardadas, lo cual disminuye la complejidad en cuanto a tratamiento de datos y arquitectura de la red. Además, mediante el algoritmo de retropropagación del errores la red es capaz de “cortar” el almacenamiento sobre lo ocurrido en pasados muy remotos, dicho de otro modo, considera un umbral de truncamiento para el almacenamiento de información, entregando validez a análisis que consideran entrenamientos de la red con series de crecidas pegadas.

En el ámbito de la arquitectura de las redes, tanto las redes neuronales recurrentes como las prealimentadas deben poseer el mismo número de capas ocultas para modelar un mismo problema, la razón de ello es la naturaleza del fenómeno en sí. En el caso de fenómenos complejos o con una rápido tiempo de respuesta como son las crecidas pluviales que se producen en las cuencas analizadas, se necesita más de una capa oculta, puesto que se realiza un mayor procesamiento de la información con la finalidad de que la red pueda “entenderlo” y modelarlo mejor, en cambio modelamientos de fenómenos con un moderado tiempo de respuesta como es el caso de las crecidas de deshielo requieren sólo de una capa oculta.

Continuando con los aspectos arquitectónicos de las redes, una red recurrente tipo Elman no necesita ni más ni menos neuronas en las capas ocultas para modelar un fenómeno en relación a modelarlo con una red prealimentada. La cantidad de neuronas en las capas ocultas de cada red debe ser escogida de acuerdo a la realización de análisis que están basados en pruebas que asocian el número de épocas de entrenamiento y la cantidad de neuronas, de tal manera de formar curvas descriptivas de errores, encontrando una envolvente inferior característica para cada número de neuronas en la(s) capa(s) ocultas. Estos análisis pueden indicar que se necesita más neuronas para una red prealimentada, así como también, pueden indicar que se

necesita menos o la misma cantidad que en el caso de una red recurrente, no existe teoría que ligue dichas cantidades.

Con respecto a los algoritmos de entrenamiento, un algoritmo de retropropagación de errores resulta muy útil y si se considera como medida de error a minimizar el ECM regularizado, como el que se ocupó para los pronósticos, se obtienen resultados aún mejores. El número de épocas necesarias de entrenamiento es en general superior a 1000 con un óptimo entre 1200 a 2000 épocas, es en este ámbito y en cuanto a rapidez de convergencia a los pesos definitivos de la red, donde una red prealimentada presenta ventajas, debido a que se necesita alrededor del triple de tiempo para realizar pronósticos con RRS en relación a redes prealimentadas.

Cabe destacar, que en el proceso de entrenamiento una RRS supera considerablemente a una red prealimentada, sobre todo en el caso de crecidas pluviales y de análisis que consideran un entrenamiento con los datos de cada crecida, a partir de las cuales se forman modelos independientes. Las razones son una vez más las ventajas que se obtienen gracias al almacenamiento de la información contenida en los patrones de entrada a la red.

En este ámbito, se destacan los análisis de tipo unitario y continuo realizados en esta memoria, a través de los cuales se infiere que no se necesita un gran set de datos de entrenamiento para crear un buen modelo de pronóstico, sino que sólo una crecida que posea variabilidad frecuencial con grandes amplitudes de frecuencia y una duración promedio que caracterice a un determinado período, según el tipo de pronósticos que se desee (pluviales o de deshielo). Una crecida de este tipo resulta en algunos casos un mejor set de datos de entrenamiento que un conjunto de crecidas pegadas, la razón de ello está en que tanto las redes recurrentes de Elman como las prealimentadas, tratan de “entender” el fenómeno de crecidas, es decir, de ver como responde el sistema “cuenca” ante eventos de precipitación o temperatura. Luego, un set de crecidas pegadas es una serie de respuestas concatenadas, sin ninguna relación entre una y otra, ya que su ocurrencia puede estar separada tanto por días, meses o años y por lo tanto, cuando la red se encuentra con la unión de tales crecidas, y trata de modelarlo o interpretarlo en vez de agregar información agrega ruido al modelo ajustado, en el sentido de los pesos que otorga a las variables de entrada y de interconexión de neuronas.

Antes de continuar, cabe señalar que en el caso de una red prealimentada, en la cual no existe relación entre un patrón de entrada y otro, cuando la red es entrenada mediante un algoritmo, por ejemplo retropropagación de errores, los datos se tratan independientemente, pero el ajuste de pesos es en función de toda la serie de datos que ha sido entregada en el entrenamiento, de esta manera la red es capaz de generalizar más que en el caso de que se

considere una sola crecida como set de entrenamiento, pero dicha generalización no es tal a nivel de validación por crecidas, debido a que se ve afectada por datos de determinadas crecidas que no aportan información y que contribuyen a distorsionar el valor de los pesos de la red. Por ejemplo, si se valida una determinada crecida en un modelo que ha considerado un set de crecidas pegadas que la incluye v/s validarla en un modelo que ha sido entrenado sólo con ella, es lógico que el error de validación o pronóstico de esa crecida es menor en el modelo que sólo ha considerado dicha crecida para el entrenamiento. Estos puntos pretenden entonces, señalar que en el caso de redes neuronales prealimentadas si bien los algoritmos tratan la información independientemente (incluso puede entregarse sin ningún orden), los pesos de la red deben converger hasta ser representativos de todo el set de patrones independientes que se considere en el entrenamiento, algunos de los cuales pueden no ser relevantes o más bien no contribuyen. Lo que justifica que se encuentren mejores pronósticos en modelos de este tipo entrenados unitariamente.

Este efecto se manifiesta con mayor razón en el caso de RRS de Elman, donde si existe dependencia entre los patrones de entrada, debido a retardos de la información, de tal manera de almacenarla y ser utilizada como una variable de entrada útil para el pronóstico. Las variables de entrada son retardadas mediante el funcionamiento de la red y por lo tanto, cada vez que existe un salto en la serie de tiempo (paso de una crecida a otra) la red almacena información irrelevante para la crecida actual, incorporando factores de ruido al modelo ajustado, o dicho de otra forma, el modelo se desajusta cada vez que encuentra un salto en la serie de tiempo. Sin embargo, una serie de crecidas pegadas puede ser utilizada como conjunto de datos de entrenamiento, validación y prueba en una red recurrente de Elman y formar un buen modelo de pronóstico, debido a que en las cuencas estudiadas el almacenamiento de la información considera sólo 5 retardos de las variables relevantes en el caso de crecidas pluviales y aprox. 10 a 15 en el caso de crecidas de deshielo, retardos que son suficientes como para considerar los tiempos de respuesta de las cuencas y no demasiado grandes como para ocasionar ruidos en el ajuste de pesos, considerando la duración promedio de las crecidas pluviales y de deshielo.

En conclusión, los pronósticos obtenidos para entrenamientos de tipo unitario y continuo en una red neuronal recurrente de Elman son mejores que los obtenidos mediante el uso de una red prealimentada considerando ambos tipos de entrenamiento, tanto en valor como en calidad (tiempo de ocurrencia de los pronósticos con respecto a los valores observados).

Uno de los principales resultados de esta memoria se encuentra en el pronóstico de caudales máximos tanto en crecidas pluviales como de deshielo, marcadamente en las pluviales, el cual

se refiere a que una red recurrente simple tipo Elman que considera como variable de entrada el gradiente de caudal en t para pronósticos a $t + x$ hrs. , con x variable, entrega pronósticos más cercanos al valor y con menos desfases que una red prealimentada, lo cual es óptimo en términos de prevenir impactos y tomar medidas de mitigación.

Dicho beneficio, es extensivo a crecidas de deshielo en el sentido de que la RRS de Elman entrega pronósticos de caudales máximos más cercanos al valor que una red prealimentada, pero, en cuanto a los desfases de ocurrencia de los valores observados y pronosticados la mejoría no es tan clara, dado el comportamiento cíclico (cada 24 hrs.) de la variable a pronosticar, se presentan máximos locales o un solo máximo global de período aprox. de 24 hrs. Luego, el hecho de encontrar desfases en el pronóstico de máximos de 24hrs. o 48 o incluso más horas dependiendo de la duración de la crecida es una opción factible y que no afecta el modelo de pronóstico en cuanto a su eficiencia.

En esta memoria se ha comparado modelos de redes neuronales recurrentes de Elman con modelos neuronales prealimentados y además, con modelos paramétricos, ARMAX y RARMAX. Un modelo neuronal prealimentado y un modelo ARMAX son no recursivos o más bien estáticos, en los cuales se supone que los parámetros que identifican el fenómeno a modelar no varían bruscamente; esta hipótesis se libera al usar redes dinámicas como son las redes recurrentes o un modelo RARMAX, por lo tanto, implícitamente al comparar los errores de los modelos de pronóstico se está observando el comportamiento de los parámetros que definen la respuesta de la cuenca. En el caso de los análisis realizados, se desprende que si bien el comportamiento de los parámetros es dinámico, dicha variabilidad no es demasiado grande en el caso de crecidas de deshielo, especialmente en la cuenca del Estero Arrayán, donde los pronósticos y errores en modelos prealimentados y modelos ARMAX no difieren notoriamente de los encontrados en un modelo neuronal de Elman.

En el campo de los modelos paramétricos, los modelos ARMAX otorgan resultados de pronóstico satisfactorios, pero retrasados, lo cual es extensivo a modelos tipo RARMAX. En modelos del primer tipo no existe una tendencia marcada de sobrestimación o subestimación de los caudales, en cambio un modelo tipo RARMAX que considera un cierto factor de olvido (0,98) en general se causa grandes sobrestimaciones de caudal, sobre todo en los máximos.

El modelo RARMAX entrega un desempeño similar y estable en términos del valor del ECM de pronóstico de crecidas que se obtiene, independiente del conjunto de entrenamiento utilizado, lo que no sucede en un modelo ARMAX. La razón radica en que el modelo RARMAX ajusta los parámetros del modelo y los corrige punto a punto tanto en entrenamiento, validación y prueba (con la condición inicial de los parámetros de

entrenamiento), de tal manera que para los primeros datos de la crecida, el modelo comienza estimando muy mal y por lo tanto suma un gran error, pero a medida que procesa los datos de la crecida, el ajuste va mejorando. Un modelo ARMAX, en cambio, ajusta los parámetros característicos del modelo una sola vez, durante el proceso de entrenamiento, y dicho modelo se hace extensivo para validar y probar.

En el caso de modelos RARMAX entrenados unitariamente, los modelos formados a partir del entrenamiento con los datos de las crecidas con menores amplitudes de frecuencia presentan los menores errores, la razón radica en el último punto señalado anteriormente, que indica que el ajuste se realiza punto a punto tanto en el entrenamiento como en la validación de crecidas. Luego, para una crecida que contenga muchas componentes de frecuencia y sea más variable, el ajuste se hace más difícil por medio de un modelo polinómico o paramétrico recursivo, lo cual fija el valor inicial de los parámetros característicos del modelo y por lo tanto, es menos extensivo a crecidas más simples, entregando de esta manera un mayor error. En cambio, para modelos ARMAX entrenados con los datos de la crecida con mayor amplitud de frecuencia los parámetros se ajustan una sola vez en función del comportamiento general de la crecida, permitiendo una mayor capacidad de generalización.

Entonces, si bien el modelo entrega errores mayores a los de un modelo ARMAX tanto en el entrenamiento como en la validación de crecidas con grandes componentes de frecuencia, frente a una crecida con bajas y pocas componentes de frecuencia un modelo tipo ARMAX entrega pronósticos más bien erróneos que son mejores en un modelo RARMAX.

El último modelo de pronóstico analizado corresponde a un modelo creado a partir de la red neuronal recurrente de identificación de Hopfield. Está basado en dos etapas, una de identificación del gradiente con respecto al tiempo de pronóstico y otra de estimación del caudal. En general, el modelo es competitivo en términos de eficiencia de pronósticos con los demás modelos neuronales y también con modelos paramétricos, pero de antemano con la desventaja de que la información debe ser codificada en 1s y -1s, lo que hace que el procesamiento de datos sea engorroso y con la incapacidad de almacenar demasiada información, ya que no se garantiza su correcto funcionamiento si el número de patrones almacenados en la red es demasiado grande en comparación a la cantidad de neuronas que posee (por demasiado grande debe entenderse mayor o igual en aprox. 1orden de magnitud).

5.1 Pronósticos Pluviales

- Las configuraciones óptimas consideran el volumen de precipitación, debido a que de esta manera se permite variar la influencia de la línea de nieve en el tiempo y durante cada crecida, además, consideran el caudal y el gradiente de este último. La variable gradiente de caudal no posee una interpretación física y fundamenta el hecho de encontrar mejores pronósticos debido a que la derivada de la función (tiempo, caudal) define su tendencia o forma, lo cual es muy relevante a la hora de realizar pronósticos, ya que permite modelar mejor las curvas de ascenso y descenso así como también obtener un mejor ajuste de valores máximos.
- Las arquitecturas óptimas de red involucran dos capas ocultas dado el rápido tiempo de respuesta de las cuencas analizadas ante eventos de precipitación.
- El número de neuronas necesarias y óptimas en las capas ocultas de un modelo de Elman es aproximadamente el mismo que para una red prealimentada y no depende del tipo de entrenamiento utilizado. Esto no tiene una justificación proveniente de la teoría de redes neuronales, la cantidad de neuronas en dichas capas en una red prealimentada y en una recurrente podría ser cualquiera y diferente.
- Los menores ECM y en consecuencia mejores pronósticos de crecidas pluviales son obtenidos para un modelo neuronal recurrente de Elman, frente a modelos prealimentados, recurrentes de Hopfield y paramétricos.
- Modelos entrenados unitariamente con la crecida que presenta mayor cantidad de componentes de frecuencia y amplitud de ellas presentan menores ECM y por lo tanto son un mejor modelo de pronóstico.
- Los índices de eficiencia de pronóstico de máximos muestran que un modelo neuronal de Elman presenta mejores pronósticos de caudales máximos que un modelo prealimentado, tanto en modelos entrenados unitariamente como continuamente.
- La eficiencia del pronóstico con respecto al desfase en la ocurrencia del caudal máximo observado y pronosticado, se debe en gran parte a la incorporación de la variable gradiente de caudal, pero no tiene el mismo efecto en ambos tipos de entrenamiento, siendo mejor en un análisis de tipo unitario.
- Modelos paramétricos pronostican con un ECM cercano a los entregados por las redes neuronales o a veces incluso un orden de magnitud de mayor que un modelo neuronal recurrente, además de presentar un mayor retraso y desfase.

- El modelo neuronal de Hopfield es una opción viable para realizar pronósticos .En general, los errores de pronóstico entregados por este modelo son del mismo orden que los obtenidos para modelos neuronales de pronóstico tanto prealimentados como recurrentes de Elman, aunque mayores que los obtenidos para modelos unitarios y en términos generales bien similares a los obtenidos para un modelo paramétrico tipo ARMAX y pero inferiores a los de un modelo RARMAX.

La mayor parte del tiempo existe sobre-estimación del caudal máximo, pero, en el momento de ocurrencia real de dicho valor se presenta más bien una subestimación con respecto al valor observado. Esto se debe a que el modelo de pronóstico trabaja en base a la identificación del gradiente en el instante de pronóstico conocida la tendencia precedente.

- Un aspecto relevante y aportante de error a los modelos, es el considerar los registros de precipitación medidos en la estación meteorológica Cerro Calán , la cual se ubica fuera de las cuencas definidas por las estaciones fluviográficas Estero Arrayán en La Montosa y Mapocho en Los Almendros, lo cual ocasiona un desfase temporal en cantidad, tiempo y espacio. Esto fue deducido anteriormente por González, 2003.

5.1.1 Cuenca del Estero Arrayán en La Montosa

El análisis de Fourier aplicado a las 21 crecidas seleccionadas, entrega a la crecida 4 como la crecida con mayores amplitudes de frecuencia.

Con respecto a modelos neuronales propiamente de pronóstico, las arquitecturas óptimas de red involucran dos capas ocultas con un óptimo de 7 neuronas en cada capa, tanto en modelos neuronales recurrentes de Elman y en modelos prealimentados, entrenados unitaria ó continuamente.

El mejor modelo paramétrico es un modelo ARMAX de configuración [5, 5, 1] ([retardo de caudal, volumen de precipitación, desfase inicial de precipitación], entrenado con los datos de la crecida 4, los errores de pronóstico son mayores que los entregados por un modelo neuronal tanto recurrente de Elman como prealimentado.

El modelo de identificación de Hopfield, posee 21 neuronas, 210 interconexiones y 161 patrones dados por los datos de la crecida 4. Entrega errores de pronóstico al nivel de un modelo ARMAX.

El modelo que entrega los menores errores de pronóstico a 4, 8, 12, 16 y 20 horas de adelanto es un modelo de Elman de arquitectura 3-7-7-1 entrenado con la crecida 4. Por lo tanto, se considera el mejor modelo de pronósticos considerando los modelos analizados.

El error promedio cometido en el pronóstico de caudales máximos para el modelo óptimo, indica subestimaciones de aprox. un 5 % del valor observado, valor que aumenta a un 17 % al tiempo de ocurrencia real (tiempo en que ocurre el máximo observado) y un retraso en el pronóstico que en promedio es de 3 hrs., para obtener pronósticos de hasta 20 hrs. de adelanto mediante un análisis unitario.

Se encuentran pronósticos “certeros” hasta 12 horas de adelanto los cuales en ciertas ocasiones pueden ser extensivos a 16 horas.

5.1.2 Cuenca Mapocho en Los Almendros

El análisis de Fourier aplicado a las 27 crecidas seleccionadas, entrega a la crecida 11 como la crecida con mayores amplitudes de frecuencia.

Con respecto a modelos neuronales propiamente de pronóstico, las arquitecturas óptimas de red involucran dos capas ocultas con un óptimo que varía según los pasos de adelanto a los que se quiere pronosticar, la cantidad oscila entre 5 a 10 neuronas en cada capa para pronósticos a 4, 8, 12, 16 y 20 hrs. tanto en modelos neuronales recurrentes de Elman y en modelos prealimentados, entrenados unitaria ó continuamente.

El mejor modelo paramétrico es un modelo ARMAX de configuración [3, 5, 1] ([retardo de caudal, volumen de precipitación, desfase inicial de precipitación], entrenado con los datos de la crecida 11. Los errores de pronóstico son mayores que los entregados por un modelo neuronal recurrente de Elman, pero del mismo orden de magnitud. Las diferencias entre ambos modelos (ARMAX y neuronales) disminuyen a medida que se realizan pronósticos a más horas de adelanto.

El modelo de identificación de Hopfield, posee 30 neuronas, 435 interconexiones y 101 patrones definidos por los datos que caracterizan la crecida 4. El error de pronóstico se encuentra al nivel de un modelo neuronal prealimentado entrenado continuamente.

El modelo óptimo de pronósticos a 4, 8, 12, 16 y 20 horas de adelanto es un modelo de Elman entrenado con la crecida 11, de arquitectura neuronal 3-5-5-1.

Existen subestimaciones del caudal máximo en promedio de un 6 %, valor que aumenta a un 24 % a la hora de ocurrencia del máximo y un retardo que en promedio es de 7 hrs. para obtener pronósticos de hasta 20 hrs. mediante un análisis unitario.

Se encuentran pronósticos aceptables hasta 6 horas de adelanto.

5.2 Pronósticos de Deshielo

- Las configuraciones óptimas consideran el caudal desfasado en 24 horas con respecto a un instante $t + 1$, el caudal en un instante t , la temperatura retrasada en el “tiempo de desfase” de ocurrencia de máximos de caudal y temperatura y además, el gradiente de caudal con respecto a un instante t , todas ellas para realizar pronósticos a $t+x$ horas con x variable. Al igual que para crecidas pluviales, la variable gradiente de caudal no posee una interpretación física y fundamenta el hecho de encontrar mejores pronósticos debido a que la derivada de la función define su tendencia o forma, aunque no es tan relevante como en el caso de crecidas pluviales, dada la naturaleza cíclica del fenómeno relacionada con el comportamiento de la temperatura y por lo tanto predecible para la red.
- Las arquitecturas óptimas de red involucran una capa oculta dado el tiempo de respuesta de las cuencas analizadas ante aumentos de temperatura, desfases de aprox. 6 a 7 horas.
- El número de neuronas óptimas en la capa oculta en una red de Elman es aproximadamente el mismo que para una red prealimentada y no depende del tipo de entrenamiento utilizado, pueden existir casos en que se necesite más o menos neuronas en un modelo que en otro, pero tal caso no tiene una interpretación proveniente de la teoría de redes neuronales.
- Los errores obtenidos son bastante pequeños en relación a los entregados en pronósticos de crecidas pluviales, la razón de ello radica en que se trata de un fenómeno cíclico gracias al comportamiento diario de la temperatura y por lo tanto más fácil de entender y simular para la red neuronal.
- El punto anterior justifica también que el error que se comete en el pronóstico de máximos es bastante pequeño. El desfase entre los tiempos de ocurrencia del caudal máximo observado y pronosticado puede ser bastante alto, del orden de 24 ò 48 hrs. o

incluso más dada la periodicidad del fenómeno, lo cual no afecta la eficiencia del modelo de pronóstico.

- Los menores errores se obtienen para modelos entrenados unitariamente, pero en general, una red de Elman entrenada continuamente presenta un desempeño bastante eficiente y similar (en orden de magnitud) a una entrenada unitariamente, dado que las crecidas de deshielo son más duraderas que las crecidas pluviales y por lo tanto, los parámetros de la red logran ajustarse mejor producto del mayor procesamiento de la información de la crecida, descartando la información antecedente cuando se encuentre con un salto en el caso de series “pegadas”.
- Modelos ARMAX resultan eficientes en el sentido que otorgan errores similares (pero, igualmente mayores) a los obtenidos mediante redes neuronales recurrentes y pueden por lo tanto resultar buenos modelos de pronóstico con la ventaja de ser más sencillos que modelos neuronales. No así modelos RARMAX que en general presentan errores mayores a los modelos ARMAX y con una mayor sobrestimación del máximo.
- La red de Hopfield no se aplica, puesto que existe gran cantidad de registros que caracterizan las crecidas y por lo tanto, gran cantidad de información a almacenar y codificar. Luego, garantizar el correcto funcionamiento de la red es poco factible, dado que el número de patrones almacenados excede en gran cantidad a las neuronas de la red, lo cual se suma al hecho de que la variabilidad de la crecida producto de su comportamiento cíclico hace que el gradiente deje de ser representativo a varios pasos de adelanto.

5.2.1 Cuenca del Estero Arrayán en La Montosa

El análisis de Fourier aplicado a las 11 crecidas seleccionadas, entrega a la crecida 2 como la crecida con mayores amplitudes de frecuencia.

Tanto en modelos neuronales recurrentes de Elman y en modelos prealimentados, entrenados unitaria ó continuamente, las arquitecturas óptimas de red involucran 1 capa oculta con una cantidad de neuronas variable entre 7 y 14, dependiendo del paso de adelanto de los pronósticos.

Los errores de pronóstico ponderados obtenidos para un modelo neuronal recurrente de Elman no son significativamente menores a los que se obtiene con un modelo prealimentado, pero suficientes como para indicar que la RRS entrega un mejor modelo de pronóstico.

El mejor modelo paramétrico es un modelo ARMAX de configuración [7, 1, 2] ([retardo de caudal, retardo de temperatura, desfase inicial de temperatura (a partir del retardo anterior)], entrenado con los datos de la crecida 2, los errores de pronóstico son mayores que los entregados por un modelo neuronal recurrente de Elman y prealimentado, pero del mismo orden de magnitud.

El error de pronósticos es bastante bajo hasta 48 horas de adelanto y si bien en términos generales un modelo neuronal de Elman entrega pronósticos más eficientes, realizar pronósticos con ambos modelos es recomendable. Finalmente, se elige como modelo óptimo de pronóstico un modelo neuronal recurrente de Elman entrenado con los datos de la crecida 2.

El ECM cometido en pronósticos a 48 hrs. por medio de la red recurrente de Elman es de aprox. un 12% del caudal medio máximo de todas las crecidas analizadas.

En cuanto al pronóstico de caudales máximos se producen subestimaciones de, en promedio, un 6% del caudal máximo observado y de un 10% al momento de ocurrencia real para pronósticos hasta 48 hrs.

Se considera una opción factible realizar pronósticos hasta 48 hrs. garantizando la obtención de pronósticos certeros.

5.2.2 Cuenca del Mapocho en Los Almendros

El análisis de Fourier aplicado a las 10 crecidas seleccionadas, entrega a la crecida 5 como la crecida con mayores amplitudes de frecuencia.

Tanto en modelos neuronales recurrentes de Elman y en modelos prealimentados, entrenados unitaria ó continuamente, las arquitecturas óptimas de red involucran 1 capa oculta con una cantidad de neuronas variable entre 13 y 16, dependiendo del paso de adelanto de los pronósticos.

Los errores ponderados obtenidos para el mejor modelo de red neuronal recurrente de Elman son suficientes como para indicar que dicha red entrega un mejor modelo de pronóstico frente a modelos neuronales prealimentados.

El mejor modelo paramétrico es un modelo ARMAX de configuración [7, 4, 1] ([retardo de caudal, retardo de temperatura, desfase inicial de temperatura (a partir del retardo anterior)], entrenado con los datos de la crecida 5, los errores de pronóstico son similares a los entregados por un modelo neuronal recurrente de Elman.

Se presentan dos modelos similarmente buenos a nivel de eficiencia medida por medio del valor del ECM. El modelo neuronal de Elman y el modelo paramétrico ARMAX.

Los errores de pronóstico son bastante bajos en relación a los obtenidos para crecidas pluviales en la misma cuenca, tanto para modelos recurrentes como para modelos prealimentados hasta 48 horas de adelanto y si bien, en términos generales un modelo neuronal de Elman entrega pronósticos más eficientes, realizar pronósticos con ambos modelos neuronales es recomendable, a los cuales se añade el modelo ARMAX. Finalmente, se elige como mejor modelo de pronóstico un modelo neuronal de Elman entrenado con los datos de la crecida 5.

El ECM cometido en pronósticos a 48 hrs. por medio de la red recurrente de Elman es de aprox. un 17 % del caudal medio máximo de todas las crecidas analizadas.

Se producen subestimaciones en el pronóstico de caudales máximos de cada crecida para pronósticos hasta 48 horas, en promedio de un 11 % del caudal máximo observado, valor que crece a un 18 % al momento de ocurrencia del máximo observado.

Se considera una opción factible realizar pronósticos hasta 48 hrs. garantizando la obtención de pronósticos aceptables.

5.3 Recomendaciones

En el campo de las redes neuronales recurrentes, la investigación de su eficiencia en realizar pronósticos de crecidas podría continuarse si se utiliza como algoritmo de entrenamiento el filtro de Kalman Extendido desacoplado, que utiliza de un modo más elaborado que el descenso por gradiente las derivadas de la función de error y que cualquiera de los algoritmos descritos en esta memoria.

Normalmente, algoritmos como el descenso por gradiente son menos rápidos de lo deseado debido a que utilizan la última estimación del gradiente: Las derivadas de la función de error sólo toman en cuenta la distancia entre la salida actual y la correspondiente salida deseada sin usar a la hora de actualizar los parámetros de la red ninguna información sobre la historia anterior del entrenamiento.

El filtro de Kalman extendido desacoplado (Haykin 1999; Puskorius y Feldkamp 1994), se basa en el filtro de Kalman extendido para superar la limitación anterior y considerar el entrenamiento como un problema de filtrado óptimo en el que se encuentra recursivamente

una solución al problema de mínimos cuadrados. En todo momento se utiliza toda la información suministrada a la red hasta el instante actual, incluidas todas las derivadas calculadas desde la primera iteración del proceso de aprendizaje. Sin embargo, el algoritmo funciona de tal manera que sólo es necesario almacenar explícitamente los resultados de la última iteración.

Luego, se propone programar el algoritmo de filtro de Kalman extendido desacoplado cuya teoría puede ser visualizada (Pérez Ortiz, 2002), en pos de mejorar la rapidez de entrenamiento y la estimación de los parámetros de la red.

La investigación de pronósticos de crecidas utilizando RNR, puede además, continuarse en el ámbito de las redes neuronales de Hopfield. Se propone entonces, utilizar dicha red neuronal para el ajuste de los pesos óptimos de la red de Elman, ya que la red es capaz de encontrar los parámetros para un problema de mínimos cuadrados sin restricciones. La teoría puede ser visualizada (Altes, R. 2007. Unconstrained Minimum Mean-Square Error Parameter Estimation with Hopfield Networks).

A fin de continuar los pronósticos de crecidas utilizando diferentes tipos de redes neuronales con el fin de mejorar su eficiencia, se propone implementar un modelo neuro-difuso, el cual ha resultado bastante satisfactorio en pronósticos realizados en la cuenca del río Magdalena, principal arteria fluvial de Colombia. Resulta interesante entonces, aplicar la teoría de un modelo neuro-difuso para pronósticos de caudal en la cuenca definida por la estación Mapocho en Los Almendros, ya que los obtenidos mediante redes recurrentes de Elman si bien mejoran en relación a los entregados por modelos prealimentados y modelos paramétricos, no son lo suficientemente efectivos para pronosticar con más de 6 horas de antelación (crecidas pluviales), que pueden llegar a ser necesarias para tomar medidas de prevención y mitigación de impactos. .

La teoría y desarrollo del modelo pueden ser visualizados (Jaimes G. 2006. Modelo Inteligente Para Soportar Decisiones de Navegación en el Río Magdalena).

Finalmente, se propone programar la red y el algoritmo de retropropagación de errores modificado en un programa computacional tipo MATLAB, para ser implementado en la cuenca del Río Mapocho.

En el ámbito hidrológico, se propone aplicar algún análisis de frecuencia, tipo Transformada de Fourier en crecidas pluviales, de tal manera de llevar la serie de tiempo a una serie de frecuencias e identificar alguna frecuencia característica y común en las crecidas que indique la presencia de deshielo.

Referencias

- [1] ACOSTA, M. y C. ZULUAGA. 2000. "Tutorial Sobre Redes Neuronales Aplicadas a la Ingeniería Eléctrica y su implementación en un sitio Web". Memoria de Ingeniero Electricista. Universidad Tecnológica de Pereira, Facultad de Ingeniería Eléctrica.
- [2] ALTES, R. 2007. "Unconstrained Minimum Mean -Square Error Parameter Estimation with Hopfield Networks". IEEE transaction, CA 92037.
- [3] BALDI, B.; R. HORNICK. y T. KOHONEN. 1989. "Neural Networks and Principal components analysis: Learning from examples without local minima". IEEE transactions on Neural Networks, Vol 2, pp. 53-58.
- [4] BISHOP, C. 1995. "Neural Networks for Pattern Recognition". Oxford University Press, Inglaterra.
- [5] BODÉN, M. y J. WILES. 2002. "On learning context free and context sensitive languages". IEEE Transactions on Neural Networks, Vol. 13, Nº 2, pp. 491- 493.
- [6] BOWDEN, G., G. DANDY y H. MAIER. 2005. "Input determination for neural network models in water resources applications. Part 1- background and methodology". Journal of Hydrologic Engineering, pp 301 75-92.
- [7] CARRASCO, R.; M. FORCADA y L. SANTAMARÍA. 1996. "Inferring stochastic regular grammars with recurrent neuronal networks, Gramatical inference: learning syntax from sentences", Vol. 147 de Lecture Notes in Artificial Inteligence. Springer-Verlag, Berlín.
- [8] CLEEREMANS, A., D. SERVAN-SHREIBER Y J.L. MCCLELLAND. 1989. "Finite state automata and simple recurrent networks". Neural computation, Vol. 1, Nº13, pp.372 -381.

- [9] COUBALY, P. F. ANCTIL y B. Bobée. 2001. "Multivariate Reservoir Inflow Forecasting Using Temporal Neural Networks". *Journal of Hydrologic Engineering*, pp. 201.
- [10] CHOW, V. T; D. MAIDMENT y L. MAYNS. 2004. "Hidrología Aplicada". Editorial McGraw Hill Interamericana S.A. Santa Fe de Bogotá, Colombia.
- [11] DEMUTH, H. y M. BOBEÉ. 1998. "Neural Network Toolbox for Use with Matlab, User's Guide, Versión 3.0".
- [12] DOGAN, E. 2007. "Daily Streamflow Forecasting using Artificial Neural Networks". *Journal of Hydrologic Engineering*, pp 301 75-92
- [13] DÓLING, O. Y E. VARAS. 2000. "Pronóstico de caudales nivales Usando Redes Neuronales Artificiales. XIX Congreso Latinoamericano de Hidráulica. Córdoba, Argentina.
- [14] ELMAN, J. L. 1991. "Distributed representations, simple recurrent networks, and grammatical structure". *Machine Learning*, Vol.7, pp.195-225.
- [15] ESPILDORA, B; E. BROWN.; G. CABRERA y P. ISENSEE. (1975), "Elementos de Hidrología". Centro de Recursos Hídricos, Universidad de Chile.
- [16] ESTÉVEZ, P. 1999. "Clasificación de Patrones mediante Redes Neuronales Artificiales". *Anales del Instituto de Ingenieros de Chile*. Santiago, Chile.
- [17] ESTÉVEZ, P. 2000. "Teoría de Redes Neuronales. Apuntes de Redes Neuronales". Universidad de Chile, Departamento de ingeniería eléctrica.
- [18] FELDKAMP, L. A. y G.V. PUSKORIUS. 1994. "Training controllers for robustness: multi- stream DEKF". *IEEE International Conference on Neural Networks*, págs.2377-2382.

- [19] FORCADA, M. L. y R. CARRASCO. 1995. "Learning the initial state of second order recurrent neural network during regular – language inference". *Neural computation*, Vol. 7, pp. 923-930.
- [20] FORCADA, M. L.; R. CARRASCO y M. GORI. 2001. "Neural Nets, recurrent, en J.G". Webster, coordinador, *Wiley Encyclopedia of Electrical and Electronics Engineering*, John Wiley & Sons.
- [21] FREEMAN, J. y D. SKAPURA. 1993. "Redes Neuronales: Algoritmos, aplicaciones y técnicas de programación", (Versión en español de Rafael García-Bermejo Giner, con la colaboración de Luis Joyanes Aguilar). Editorial Addison-Wesley Iberoamericana S.A.
- [22] GILES, C. L., C. B. MILLER, D. CHEN. 1992. "Learning and extracting finite state automata with second order recurrent neural networks". *Neural Computation*, Vol. 4, N° 3, pp. 303-405.
- [23] GONZÁLEZ, R. 2003. "Pronóstico de Crecidas en Tiempo Real. Uso de redes neuronales". Tesis para optar al grado de Magíster en Ciencias de la Ingeniería Mención Recursos Hídricos y Medio Ambiente Hídrico, Universidad de Chile.
- [24] GOUDREAU, M. W.; C. L. GILES, S. T. CHAKRADHAR y D. CHEN. 1994. "First-order vs. second order single layer recurrent neural networks". *IEEE Transactions on Neural Networks*, Vol. 5, N°3, pp. 511- 513.
- [25] HAYKIN, S. 1999. "Neural Networks: a comprehensive foundation". Prentice Hall, New Jersey, 2ª ed.
- [26] HU, T. y K. LAM. 2001. "River for time series prediction with a range-dependent neural network". Department of Hydraulic Engineering, Wuham University, Wuham, Hubei province, China.

- [27] JAIMES, W. 2006. “Modelado Inteligente para Soportar Decisiones de Navegación en el Río Magdalena”. Universidad de Pamplona, Facultad de Ingeniería y Arquitectura, departamento de Electrónica, Telecomunicaciones y Sistemas. Pamplona, Colombia.
- [28] KALMAN, R. E. 1960. “A New approach to linear filtering and prediction problems”. Transaction of ASME – Journal of Basic Engineering, págs.35-45.
- [29] LENKKAS, D.; C. ONFOS; M. LEE y E. BALTAS. 2005. “Application of Artificial Neural Networks for Flood Forecasting”. Department of Civil and Environmental Engineering, Imperial Collage of Science, Technology and Medicine. Londres, Inglaterra.
- [30] LJUNG, L. 1996. “The System Identification Toolbox: The Manual”. The Math Works Inc.
- [31] LJUNG, L. 1999. “System Identification Theory for the User. Prentice Hall PTR”. New Jersey, EE.UU.
- [32] LOEZ, R. 1996. “DFT Theory and FFT Techniques for System Identification”. Apuntes de Control Automático. Universidad de Wisconsin - Madison, EE. UU.
- [33] McCLUSKEY, P.G. 1993. “FeedForward and Recurrent neural networks and genetic programs for stock market and time series forecasting”. Informe técnico CS-93-36. Brown University.
- [34] MADSEN, H. y C. SKOTNER. 2005. “Adaptive state updating in real-time river flow forecasting - a combined filtering and error forecasting procedure”. Journal of Hydrologic Engineering, pp 300 75-93.
- [35] MARRA, S. 2006. “A New Technique for Solar Activity Forecasting using Recurrent Elman Networks”. Journal of Hydrologic Engineering, pp 601 65-93

- [36] MÁRQUEZ, F. 1991. “Pronóstico de Crecidas en Tiempo Real Aplicado a la cuenca del Río Mapocho”. Universidad de Chile, Departamento de Ingeniería Civil.
- [37] MUSAVI, M. 1992. “A Probabilistic Model for Evaluation of Neural Network Classifiers”. *Pattern Recognition*, N° 25, pp. 1241-1251.
- [38] NACHIMUTHU, K. y W. GRENNEY. 2005. “Neuronal Network for River Flow Prediction”. *Journal of Hydrologic Engineering*, pp 511 43-95.
- [39] OMLIN, C. W. y C. L. 1996. “Constructing deterministic finite state automata in recurrent neural Networks”. *Journal of ACM*, Vol. 43, N°6, pp. 937-972.
- [40] PÉREZ, J. 2002. “Modelos Predictivos Basados en Redes Neuronales Recurrentes en Tiempo discreto”. Tesis para optar al grado de doctor en Matemáticas, Universidad de Alicante, Departamento de Lenguajes y Sistemas Informáticos.
- [41] PÉREZ – ORTIZ, J. A. y M. L. FORCADA. 2001. “Part-of-speech tagging with recurrent neural networks”. *Proceedings of the Internacional Joint Conference on Neural Networks*, págs.1588 -1592.
- [42] PLAUT, D. C., S. J. NOWLAN y G. E. HINTON. 1986. “Experiments on learning back propagation”. Informe técnico CMU-CS-86-126, Department of Computer Science, Carnegie- Mellon University.
- [43] PUSKORIUS, G. V. y L. A. FELDKAMP. 1994. “Neurocontrol of nonlinear dynamical system with Kalman filter trained recurrent networks”. *IEEE Transactions on Neural Networks*, Vol 5, N°2, pp. 279-297.
- [44] RUMELHART, D., G. HINTON Y R. WILLIAMS. 1986. “Learning representation by back-propagation errors”. *Nature*, Vol. 323, pp. 523-536.

- [45] SERRA, S. y C. VIONNET. 2005. "Migración de Dunas durante Eventos Extremos del Río Paraná". Centro Nacional de Estudios Hidroambientales, Universidad Nacional del Litoral, Facultad de Ingeniería y Ciencias Hídricas. Santa Fé, Argentina.
- [46] SHEPHERD, A. J. 1997. "Second-order Methods for neural networks". Springer, Londres, Inglaterra.
- [47] TRONCOSO, M. 1999. "Uso de Redes Neuronales para el Pronóstico de Crecidas en Cuencas Pluvionivales". Memoria de Ingeniero Civil. Universidad de Chile, Departamento de Ingeniería Civil.
- [48] VALVERDE, M.; H. FRAGA DEL CAMPOS VELHO y N.J. FERREIRA. 2005. "Artificial Neural Network Technique for rainfall forecasting applied to the Sao Paulo region". Centro de Previsao do Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais.
- [49] WILLIAMS, R. J. y D. ZIPSER. 1989. "A learning algorithm for continually training recurrent neural networks". Neural Computation, Vol. 1, pp. 270-280.
- [50] XINYING, Y. y V. BABOVIC. 2004. "EC-SVM approach for Real-Time hydrologic forecasting". Department of Civil Engineering, National, University of Singapur.