

## PREDICCIÓN DE PARÁMETROS DE CALIDAD DE AGUA MEDIANTE REDES NEURONALES EN LA CUENCA DEL RÍO PILCOMAYO ARGENTINA

COPPO, RICARDO J.<sup>1</sup>; JAKOMIN, MARINA L.<sup>2</sup> Y DELRIEUX, CLAUDIO A.<sup>3</sup>

1: Departamento de Electrónica  
Universidad Tecnológica Nacional – Facultad Regional Bahía Blanca  
11 de Abril 461- 8000 Bahía Blanca  
rcoppo@frbb.utn.edu.ar

2: Subsecretaría de Recursos Hídricos de la Nación  
Ministerio del Interior, Obras Públicas y Vivienda  
Presidencia de la Nación Argentina  
Esmeralda 255 Piso 11 CABA

3: Departamento de Ingeniería Eléctrica y Computadoras  
Universidad Nacional del Sur y CONICET  
Av. Alem 1253 – 8000 Bahía Blanca

**Resumen.** *El monitoreo de parámetros de calidad de agua en ríos se realiza por medio de tomas de muestras in-situ y sus correspondientes análisis de laboratorio, complementado con información hidrométrica obtenida de lecturas directas realizadas in-situ y/o sensores automáticos. Los costos de las operaciones de muestreo (traslados de personal especializado, costos de laboratorio y reactivos, instrumental, etc.) limitan la capacidad y alcance de los programas de control de calidad del agua realizado por los organismos de contralor. Desde el punto de vista operativo se desea predecir valores de los parámetros más difíciles o costosos de observar a partir de aquellos de medición más económica o de fácil obtención. Las redes neuronales artificiales (ANN) han demostrado su capacidad para predecir series de tiempo complejas que presentan niveles importantes de ruido en sus mediciones, variaciones estacionales complejas, y características no lineales. En este trabajo se estudia el empleo de una red neuronal entrenada con datos provenientes de series de bajo costo de adquisición (caudal, conductividad eléctrica, y pH) para la predicción de la concentración de boro en agua. Se presenta los resultados obtenidos al aplicar la técnica propuesta a los datos hidrológicos y fisicoquímicos provenientes de la estación Misión La Paz del río Pilcomayo (Argentina).*

**Palabras clave:** Redes Neuronales, Calidad de Aguas, Río Pilcomayo, Boro

### 1. INTRODUCCIÓN

El monitoreo de parámetros de calidad de agua que se realiza en Misión La Paz, Cuenca del Río Pilcomayo, implica la toma de muestras *in-situ* y su correspondiente análisis en laboratorio, complementado con información hidrométrica y parámetros de campo obtenido de lecturas directas realizadas *in-situ*. El análisis químico básico puede realizarse en laboratorios dentro de la Cuenca pero otros, que implican uso de instrumental sofisticado, deberán remitirse a instituciones especializadas fuera de la misma. Estas actividades tienen un

elevado costo que requiere cubrir desde el traslado de personal técnico, preservación y traslado de muestras en forma y tiempo adecuados, preservantes, soluciones de calibración hasta la calificación de instrumental de campo. Los resultados obtenidos suelen emitirse mucho tiempo después. Los altos costos atentan contra la sistematicidad y sostenibilidad del programa de este monitoreo de la calidad del agua. Desde el punto de vista operativo se desea predecir valores de los parámetros más difíciles o costosos de observar a partir de aquellos de medición más económica o de fácil obtención y de una manera temprana, para así poder emitir una recomendación en cuanto al consumo del agua como fuente de bebida. En este trabajo proponemos el uso de redes neuronales artificiales (ANN) para dicho propósito. Las redes aquí descritas [9] han sido entrenadas y validadas con datos de medición económica obtenidos en el período 2007-2014 en la estación Misión La Paz, con el objetivo de predecir la concentración de boro en agua. La obtención del valor de concentración de boro disuelto en agua a través de una red neuronal artificial permitiría realizar una pronta recomendación en relación con valor permitido para consumo humano.

### **1.1 Área de Estudio**

La cuenca del río Pilcomayo abarca una superficie del orden de 290.000 km<sup>2</sup>, constituyendo una importante región de recursos naturales de Latinoamérica compartida por Argentina, Bolivia y Paraguay. Es una cuenca transfronteriza. El río recorre más de 1000 km desde los 5500 m de altura en sus nacientes en Bolivia hasta los 250 m en los alrededores de Misión La Paz en territorio argentino. La cuenca experimenta procesos naturales y antrópicos que hacen necesaria una gestión coordinada y armónica de sus recursos hídricos, máxime en su calidad de transfronteriza, por parte de los tres países que la comparten.

Los gobiernos de los tres países han creado la Comisión Trinacional para el desarrollo de la Cuenca del Río Pilcomayo que tiene, como instancia técnica, a la Dirección Ejecutiva, encargada, entre otras funciones, de generar, recopilar y procesar información hidrológica, meteorológica y de calidad de agua y sedimentos, proveyendo información de apoyo a la toma de decisiones, así como la planificación y ejecución de los proyectos que se generen en los países.

Su Unidad de Monitoreo y Centro de datos monitorea la calidad del agua, entre otras, en la estación Misión La Paz (22° 22' 45.1"; 62° 31' 8.2"). Paralelamente el Programa de Calidad de Agua de la Subsecretaría de Recursos Hídricos de la Nación Argentina, en apoyo a la Delegación Argentina ante la Comisión Trinacional, realiza desde principios del 2007 un

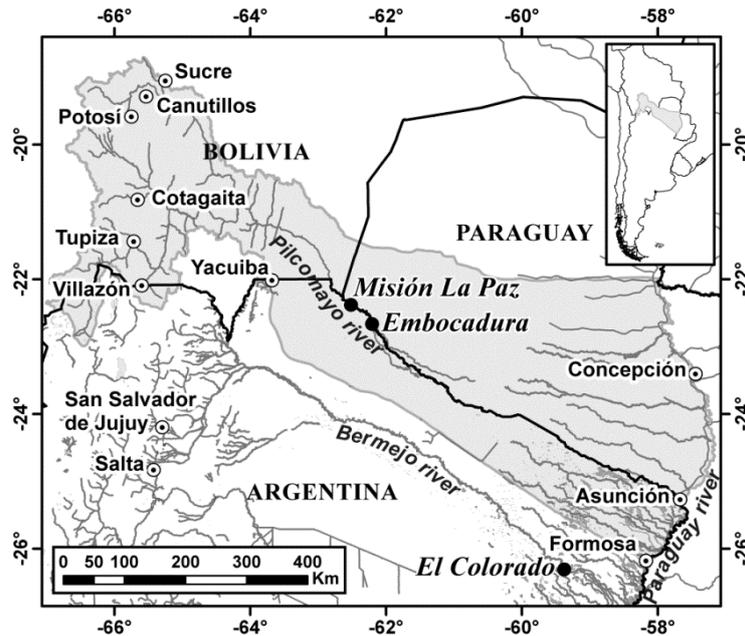


Figura 1: Estación Misión La Paz, Río Pilcomayo

programa de Monitoreo de la Calidad del Agua, Suelos y Sedimentos del Río Pilcomayo en la misma estación, sitio de ingreso de las aguas del Pilcomayo al territorio argentino–paraguayo (Fig. 1).

## 1.2 Evaluación de la presencia de boro en agua

Especialistas de los tres países definieron, en los años 2006 y 2007, la realización de monitoreos intensivos y extensivos en la cuenca, de frecuencias aproximadamente mensuales y semestrales, respectivamente, consensuando las metodologías analíticas y los puntos de monitoreo para contar con datos analíticos confiables para su interpretación. Los datos generados por los monitoreos realizados de acuerdo a tales directivas son almacenados en una Base de Datos Única [4]. Los parámetros definidos están focalizados a metales pesados. Sin embargo resaltan las elevadas concentraciones de boro disuelto en agua, de origen geológico con máximos y mínimos en Misión La Paz que fueron, entre 2007 y 2014, de 974 y 235  $\mu\text{g/l}$  respectivamente.

El nivel guía de calidad de agua para boro destinado a consumo humano, contemplando un tratamiento convencional en la fuente de provisión, referido a la muestra de agua filtrada es de 260  $\mu\text{g/l}$  [14].

En la Argentina, para ciertos cuerpos superficiales de agua de la Provincia de Salta han sido reportadas concentraciones altas de boro atribuibles a la geología natural. Es el caso del río Guachipas (Las Conchas, en su sección sur) y de su principal sistema tributario con

concentraciones de boro superiores a 10 mg/l [10]. Por otra parte, han sido informadas alteraciones de las condiciones basales de boro en aguas subterráneas y superficiales del Valle de Lerma, Salta (concentraciones de boro menores que 0,3 mg/l) como consecuencia de aportes antropogénicos, habiéndose detectado elevaciones de las concentraciones de boro hasta 6,2 mg/l, en agua subterránea, y hasta 50 mg/l, en el río Arenales [2].

Malin et al. [11] estudiaron la incidencia de boro en agua de bebida en mujeres embarazadas sobre su progeñe en un rango de concentraciones de 377 a 10929  $\mu\text{g/l}$ . El rango de boro en suero presentó muy buena correlación con el boro en sangre total y en orina y en menor grado con el boro en agua. El boro en suero por encima de 80  $\mu\text{g/l}$  estaba relacionado inversamente con la longitud al nacer. El impacto de boro era aún mayor si la exposición se producía en el tercer trimestre cuando se encontraron las máximas concentraciones en suero y un aumento de 100  $\mu\text{g/l}$  en suero provocaba una disminución de 0,9 cm en altura y 120 g en peso.

El boro se acumula en las hojas y a concentraciones bajas en raíces, tallos y frutas. Los síntomas se manifiestan sobre todo en las hojas más maduras como manchas, franjas o puntos cloróticos en los cuales se desarrollan lesiones necróticas [12].

Cuando los animales absorben grandes cantidades de boro, en un periodo de tiempo corto por comer o beber de agua, los órganos reproductivos masculinos son afectados [3]. También cuando son expuestos al boro durante la preñez, sus descendientes pueden sufrir defectos de nacimiento y fallos en el desarrollo [13]

## **2. MATERIALES Y METODOS**

El trabajo de investigación se desarrolló en una serie de etapas que abarcaron la obtención de los datos analíticos, un preprocesamiento consistente en la selección de las variables de entrada y su preparación para los programas de entrenamiento y resolución de las redes neuronales y finalmente un postprocesamiento para la presentación de resultados y análisis de la performance de los estimadores obtenidos. Parte de este desarrollo se realizó en planillas de cálculo Excel y el procesamiento propiamente dicho con el paquete estadístico R. A continuación se describen en mayor detalle cada etapa del trabajo.

### **2.1 Obtención de datos analíticos**

Para la obtención de muestras se utilizó el EPA Method 1669 [6]. Las muestras destinadas a la determinación de boro disuelto fueron filtradas en campo utilizando una membrana de acetato de celulosa de 0,45  $\mu\text{m}$ , almacenadas en envases de teflón y preservadas con ácido nítrico sub boiling a pH 2. Los envases se sellaron con doble bolsa de polietileno y se conservaron en hielo (Figura 2).



Figura 2. Obtención de muestras, filtración y medición en campo

Las muestras se analizaron por ICP utilizando un equipo Perkin Elmer ELAN ICP DRCII y Detector de Espectrometría de Masas [5] en el Laboratorio de Química Analítica en Medios Activos de la Comisión Nacional de Energía Atómica, laboratorio que participó satisfactoriamente del 7th Performance Evaluation Study (2010) y del 8th Performance Evaluation Study (2013) organizado por la red GEMS WATER PNUMA durante el período de obtención de las concentraciones boro aquí utilizadas.

Los datos de conductividad y pH fueron tomados utilizando un equipo WTW Multi 3430 Set utilizando soluciones de calibración certificadas.

Los datos de calidad de agua así relevados fueron tomados de la Base de Datos Unica (BDU) de la Dirección Ejecutiva de la Comisión Trinacional (CTN), base de libre acceso previa solicitud de una habilitación a la Dirección Ejecutiva.

Las mediciones de caudal en la estación Misión La Paz sobre el río Pilcomayo fueron realizadas por personal de EVARSA para la Subsecretaría de Recursos Hídricos de la Nación, dependiente del Ministerio del Interior, Obras Públicas y Vivienda.

## 2.2 Preprocesamiento de la información

Las series de tiempo obtenidas de las bases de datos mencionadas anteriormente son heterogéneas en su resolución espacial y temporal. Por ello, en primera instancia se procedió a un proceso de filtrado y fusión de la información manual para unificarlo en una tabla única para su procesamiento posterior.

A continuación se realizó un análisis de esta tabla y, de acuerdo a los objetivos del estudio, se seleccionó como variables de entrada la conductividad eléctrica (CE), el pH (pH), y el caudal medio diario (Q). Estas variables se han elegido por ser de medición continua o por poder ser medidos in-situ con equipamiento de bajo costo sin la necesidad de realizar preservación de la muestra y transporte de la misma.

Como única variable de salida se ha elegido la concentración de boro disuelto que se

considera biodisponible y perjudicial para salud humana. En futuros pasos se podrán incorporar más variables de salida por la flexibilidad de la configuración de la red neuronal.

### 2.3 Normalización de las variables de entrada

Al observar el grado de variabilidad de los parámetros de entrada se observa una gran dispersión en los rangos numéricos de cada uno. Esta situación es peligrosa para la mayoría de las redes neuronales ya que suele producir una priorización de las variables de mayor valor numérico sobre las de menor valor durante el entrenamiento de las mismas. Para evitar este problema numérico primero se normalizan las variables para acotar su rango entre 0 y 1.

$$\hat{x}_i = \frac{x_i - x_{min_i}}{x_{max_i} - x_{min_i}} \quad (1)$$

Expresión en que  $x_i$  representa una de las variables de entrada,  $x_{min}$  y  $x_{max}$  sus valores extremos y  $\hat{x}_i$  su valor normalizado. Denotaremos al vector de todas las variables de entrada normalizadas como  $\hat{\mathbf{x}}$ .

### 2.4 Error de predicción

Para poder realizar comparaciones entre la solución ofrecida por las técnicas de resolución lineal multiparamétrica y las soluciones obtenidas de diversas redes neuronales se ha elegido como criterio de bondad de ajuste al error cuadrático medio definido sobre el error de predicción entre el modelo y lo observado sobre las  $N$  muestras utilizadas en el estudio.

$$RSME = \sum_{i=0}^N ||p_i - x_i|| = \sum_{i=0}^N \sqrt{(p_i - x_i)^2} \quad (2)$$

## 3. REDES NEURONALES PREDICTORES

Las redes neuronales del tipo “feed forward” se implementan sobre una estructura de elementos computacionales referenciados como “neuronas” por su similitud con su equivalente biológico [1, 8, 9]. Dichas neuronas se organizados en  $K$  capas, donde  $k=0$  representa las variables de entrada y la última capa identifica la(s) variables(s) de salida.

Cada neurona se caracteriza por una función de activación  $\theta(x)$  que depende de las funciones de activación de la capa anterior  $\theta^{k-1}(x)$ , y de una serie de coeficientes multiplicativos  $\omega_{ij}$  aplicados entre la función de activación de la neurona  $i$  de la capa  $k-1$  y la neurona  $j$  de la capa  $k$ . ( $i=0$  se interpreta como un valor de bias que se introduce en cada capa).

$$\theta_i^k(\hat{\mathbf{x}}) = \sigma \left[ \sum_{j=0}^{N_{k-1}} \omega_{ij}^{(k-1)} \theta_j^{(k-1)}(\hat{\mathbf{x}}) \right] \quad (3)$$

Para posibilitar la predicción no lineal de este modelo, se debe elegir una función monótona creciente  $\sigma(\lambda)$ . Para este estudio se ha elegido la función sigmoïdal de la forma:

$$\sigma(\lambda) = \frac{1}{1 + e^{-\lambda}} \quad (4)$$

La resolución de una red neuronal consiste en la determinación de los coeficientes indeterminados  $\omega_{ij}$  que mejor adaptan la predicción de la red a los datos suministrados como entrenamiento. El proceso consiste en la optimización de la ecuación de performance por medio de un algoritmo de descenso máximo (steepest descent) que itera sobre una cantidad de pasos y finaliza al alcanzar una tolerancia de variación aceptable. El algoritmo empleado se denomina “Backpropagation” por plantear una solución de los coeficientes a partir de la reducción de los errores observados en las salidas de cada neurona cuando se los presenta con entradas cuyo resultado es conocida o esperada.

#### 4. RESULTADOS

Los datos fueron procesados con el paquete “neuralnet” del entorno estadístico R [7]. La figura 3 presenta la red neuronal obtenida con un RMSE menor a 0.725 y hallada después de 341 pasos de optimización de descenso máximo empleando un algoritmo de bootstrapping para entrenar la red con distintos subconjuntos de los datos de entrada. La red ilustrada tiene como valores de entrada el caudal, pH y CE (conductividad eléctrica) normalizadas (ecuación 1) y como única salida el boro normalizado.

Desnormalizando los valores de boro obtenidos se comparan en la figura 4 los resultados de la red neuronal con las determinaciones de laboratorio (curvas de color negro y rojo). El eje x, indicado como INDEX se refiere al número de muestra.

#### 5. CONCLUSIONES

Las redes neuronales han demostrado ser excelentes predictores de variables de calidad de agua por su capacidad de adaptación a sistemas de entradas multiparamétricas y a sus características de predicción no lineal. Sin embargo para lograr una red robusta a los cambios es necesario entrenarlos con una cantidad importante de registros de monitoreo preexistentes, es decir, no pueden utilizarse en las primeras etapas de un programa de muestreo.

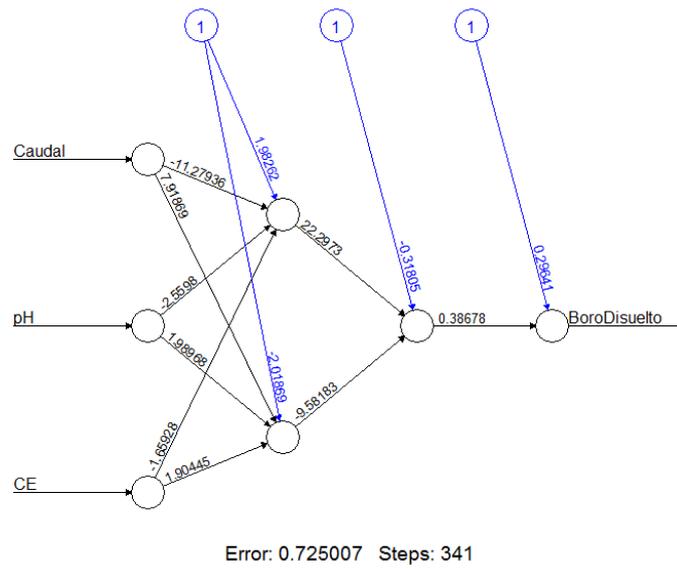


Figura 3: Red Neuronal con una capa de entrada, una de salida y 2 capas ocultas

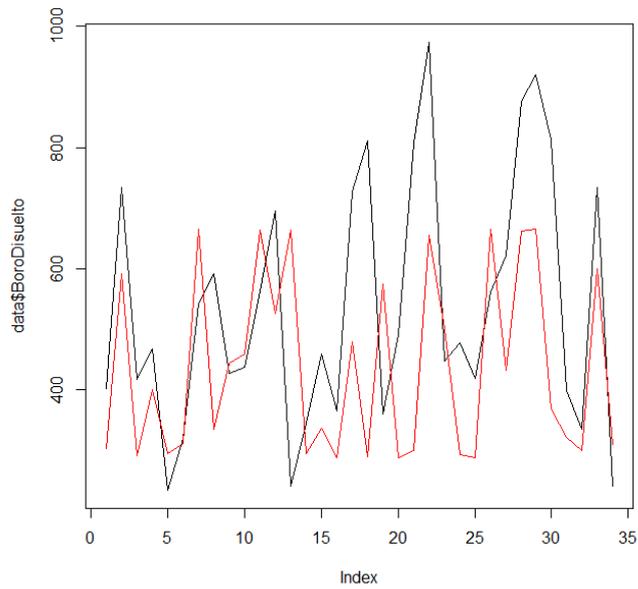


Figura 4: Datos medidos (negro) y predichos por la red neuronal (rojo)

La selección de los parámetros de entrada, la cantidad de capas de la red, y la cantidad de neuronas que conforman cada capa se determinan en forma experimental, no existiendo una

metodología única para su elección. Por ello es muy difícil afirmar que una determinada configuración de red elegida para un parámetro de salida, una estación de monitoreo, o una cuenca en particular pueda ser fácilmente trasladada a otro sin modificaciones.

Es imposible afirmar que una configuración de red es una solución óptima global al problema de predicción de un parámetro de calidad de agua. Pueden existir configuraciones mejores para un mismo conjunto de datos de entrenamiento y dicha configuración puede variar a medida que se dispone de nuevos muestreos en la base de datos de entrenamiento.

A pesar de las dificultades mencionadas las redes neuronales se perfilan como excelentes indicadores de situaciones anómalas. Una vez elegida la configuración final son fácilmente programadas en sistemas de computación móvil y pueden ser utilizados directamente en el trabajo de campo. Sus predicciones pueden ser utilizadas para tomar la decisión de muestrear o no un parámetro de alto costo para su determinación más precisa en un laboratorio.

## 6. AGRADECIMIENTOS

Los autores desean agradecer la ayuda recibida desde la Dirección Ejecutiva de la Comisión Trinacional para el Desarrollo de la Cuenca del Río Pilcomayo y del personal de la Biblioteca de los Recursos Hídricos de Argentina, Ministerio del Interior, quienes desde los inicios de la investigación brindaron su apoyo al proyecto facilitando el acceso a la Base de Datos Única del río Pilcomayo y asistiendo en la búsqueda bibliográfica requerida.

## 7. BIBLIOGRAFIA

- [1] Bishop, C.M., *Pattern Recognition and Machine Learning*, Springer Science+Business Media LLC, 233 Spring Street, New York. (2006).
- [2] Bundschuh, J. “Boron contamination of the ground- and surface waters of Lerma Valley, Argentina”. *J. Water SRT – Aqua Vol. 41*, N° 1, pp. 13-17, (1992).
- [3] Bustos-Obregón E. Carvallo M. Hartley-belmar R. Sarabia L. y Ponce C. Histopatológico y histométrico. Evaluación de boro. Efectos de la exposición sobre la espermatogénesis del ratón. *Diario Internacional de Morfología* 25 (4), pp. 919-925, ISSN 0717-9502, (2007).
- [4] DTE-CTN, “Dirección Técnica Ejecutiva de la Comisión Trinacional para el Desarrollo de la Cuenca del Río Pilcomayo”, [www.pilcomayo.net](http://www.pilcomayo.net), visitado: 25/04/2017.
- [5] EPA Method 200.8, Determination of Trace Elements in Waters and Wastes by Inductively Coupled Plasma-Mass Spectrometry, Revision 5.4, (1994).
- [6] EPA Method 1669. Sampling Ambient Water for Trace Metals at EPA Water Quality Criteria Levels. U.S. Environmental Protection Agency. Office of Water Engineering and Analysis Division (4303). 401 M Street S.W. Washington, D.C. 20460, (1996).
- [7] Frauke F., Pritsch S., “neuralnet: Training of Neural Networks”, *The R Journal*, Vol 2., pp. 30-37, (2010).
- [8] Haykin, S., *Neural Networks – A Comprehensive Foundation*, Pearson-Prentice Hall, (2009).

- [9] Hsieh, W., *Machine Learning Methods in the Environmental Sciences-Neural Networks and Kernels*, Cambridge University Press, Cambridge, England, (2009).
- [10] Lomniczi I., Musso H. y Pereyra R., “Assessment of Boron concentration in surface and groundwaters in the Lerma and Calchaquí valleys (province of Salta, Argentina)”. *Anales de la Asociación Química Argentina*. Vol.85 N°5/6 pp. 283-293, (1997).
- [11] Malin A. I., Harari F., Ying Lu, Casimiro E., Vahter M., “Boron exposure through drinking water during pregnancy and birth size”, *Environment International* 95. pp. 54–60, (2016).
- [12] Ortega, A. E. Capítulo: Toxicidad de micronutrientes: 177-207 pp. Vázquez, Mabel (ed.). *Micronutrientes en la agricultura. Diagnóstico y fertilización en Argentina. La experiencia brasilera*. Ed. AACCS (Asociación Argentina de la Ciencia del Suelo). Buenos Aires. ISBN N° 987-21419-4-0: 207 p, (2006).
- [13] Smith, G.J. y Anders, V.P. Efectos tóxicos de boro en ánade real reproducción. *Environ Toxicol Chem*, 8: pp. 943-950, (1989).
- [14] SSRH. Nivel Guía de calidad de agua para boro. Elaboración del Programa de Calidad del Agua de la Subsecretaría de Recursos Hídricos, 2002 (pag web en construcción 3-2017)