

Anexo B

Predicción Adaptativa de la Calidad del Agua Mediante Técnicas de Inteligencia Computacional



Informe Final de Pasantía

Ing. Iván Darío López Gómez

Director: Dr. Ing. Juan Carlos Corrales

Asesor: Dr. Miguel Peña Varón

Universidad del Cauca

Facultad de Ingeniería Electrónica y
Telecomunicaciones
Departamento de Telemática -
Grupo de Ingeniería Telemática -
GIT

Universidad del Valle

Instituto de Investigación y
Desarrollo en Abastecimiento de
Agua, Saneamiento Ambiental y
Conservación del Recurso
Hídrico -Cinara

Santiago de Cali, Octubre de 2015

1 PLANTEAMIENTO DEL PROBLEMA

La historia de la humanidad ha estado ligada estrechamente al agua. Desde sus albores existen reseñas a grandes inundaciones, sequías y a los esfuerzos del hombre por dominar este recurso vital de la naturaleza para cubrir sus propias necesidades [1]. Tradicionalmente el agua se ha tratado y gestionado como si fuese un recurso ilimitado debido a su naturaleza renovable, sin embargo, el incremento indiscriminado de su uso ha acarreado consigo un acelerado deterioro en su calidad y en ocasiones se ven cambios en su distribución temporal y espacial, con consecuencias que no son previstas por completo, pero de una importante gravedad ambiental, económica, social, entre otras.

La calidad del agua se puede definir como el conjunto de características físicas, químicas y biológicas de los cuerpos de agua superficiales y subterráneos, sin haber realizado un tratamiento previo, es decir, sin la intervención humana¹ [2]. Así mismo, estas características afectan la capacidad del agua para sustentar tanto a las comunidades humanas como la vida vegetal, animal y microbiana. Por otra parte, los expertos en el manejo del recurso hídrico, indican que no existe una única definición de la calidad del agua, pues esta depende estrictamente del uso al que esté destinado el preciado líquido [3], es así como por ejemplo, el agua que no se puede utilizar para el consumo humano, puede servir para otras actividades como el riego o la piscicultura, entre otros, ya que posee características específicas que la hacen apropiada para dicho uso [4].

A pesar de que Colombia es el sexto país con mayor oferta hídrica en el mundo², el Ministerio del Medio Ambiente calcula que aproximadamente un 50% de los recursos hídricos presentan problemas de calidad [5]. Lo anterior obedece en gran medida al crecimiento de la población y de las actividades económicas, siendo necesario un monitoreo y control constante que permita tomar las acciones necesarias para abordar esta problemática con el fin de disminuir su impacto en los procesos naturales y sociales, especialmente en la salud humana [6]. Sin embargo, en muchas ocasiones para controlar unas condiciones apropiadas de calidad del agua, no es suficiente con establecer actividades de monitoreo³ que brinden acciones correctivas ante determinado tipo de contaminación, sino que además de esto, se hace necesario contar con modelos o mecanismos que permitan anticiparse a la materialización del riesgo con el suficiente rango de tiempo para prevenir los efectos negativos que afecten la calidad del recurso hídrico.

De acuerdo con lo anterior, el concepto de predicción científica se centra en una declaración precisa de lo que ocurrirá en determinadas condiciones especificadas; de esta manera, su validez se mide por el éxito o acierto que tengan sus predicciones. Las teorías que generan muchas predicciones que resultan de gran valor (tanto por su interés científico como por sus aplicaciones) se confirman o se falsean fácilmente y, en muchos campos científicos, las más deseables son aquéllas que, con número bajo de principios básicos, predicen un gran

¹ Para este trabajo se manejará esta definición de Calidad del Agua.

² Este ranking se ha ido especializando aún más, actualmente para el análisis se tienen en cuenta aspectos como la calidad y la disponibilidad que la población tiene del recurso. Bajo estos nuevos criterios, Colombia se ubica en el puesto 24 a nivel mundial.

³ Estas actividades de monitoreo pueden ser mediante la toma de muestras de agua que se realizan de forma manual o automatizada mediante sensores electrónicos.

número de sucesos. Sin embargo existe una multitud de campos en la ciencia donde la predicción se convierte en una tarea compleja, bien sea por el gran número de variables involucradas o por la misma dinámica desconocida de los fenómenos involucrados en una situación problemática [7].

En este sentido, la predicción de la calidad del agua desempeña un papel muy importante para muchos sectores socio-económicos que dependen del uso del preciado líquido. Es así como en los últimos años, el campo de la Inteligencia Artificial (IA) [8], ha introducido algoritmos y técnicas de predicción que cuentan con la capacidad de estimar las condiciones futuras de un cuerpo de agua con base en el análisis de los datos que han sido recolectados en el pasado.

Dada la importancia que representa la gestión de la calidad del recurso hídrico para diferentes sectores socio-económicos, varias propuestas investigativas se han enfocado en la predicción como una herramienta útil para anticipar posibles eventos adversos tanto para la producción como para la salud humana. Los trabajos más destacados hacen uso de las Redes Neuronales Artificiales (RNA) como en [9-11]. Adicional a esto, otros trabajos plantean modelos de predicción híbridos [12][13] en los cuales combinan las RNA y los Modelos Auto-regresivos Integrados de Media Móvil, también conocidos como modelos ARIMA⁴[12] que trabajan con series de tiempo, con el objetivo de mejorar la precisión de las predicciones. Por otro lado, trabajos como [14-17] permiten optimizar la selección de los parámetros más representativos para tener en cuenta en la predicción (en muchas ocasiones adicionar demasiados parámetros no es conveniente debido a que genera ruido en el modelo y en vez de aumentar la precisión, esta disminuye). Actualmente nuevas estrategias se han planteado para mejorar la precisión del proceso de predicción como en [18] que utiliza Máquinas de Vectores de Soporte (MVS) y [19], donde se hace uso de esta misma técnica adicionando la Optimización por Nubes de Partículas (PSO)⁵ obteniendo mejores resultados en la precisión comparada con los trabajos mencionados anteriormente.

Con base en lo anterior, es importante destacar los avances y aportes de estos trabajos dentro del campo de la predicción de la calidad del agua, sin embargo ninguno de ellos considera la aplicación de su modelo de predicción sobre diferentes usos del recurso hídrico, estos se limitan a realizar el proceso de predicción tomando como referencia los datos pertenecientes a un determinado uso del agua, muchas veces sesgando los resultados hacia ese escenario en particular sin tener en cuenta la aplicabilidad que el mismo modelo de predicción pueda tener para otros usos y que puede ser de utilidad para diversos sectores que utilizan el agua en sus procesos productivos. Por consiguiente, en el presente trabajo de investigación se pretende que un mecanismo de predicción de calidad del agua cuente con la característica adaptativa, en el sentido de que pueda ser aplicado en diferentes usos del agua manteniendo la precisión de las predicciones; lo anterior mediante técnicas de Inteligencia Computacional (IC).

Teniendo en cuenta las consideraciones anteriores, en la presente propuesta de investigación se formula la pregunta ¿Cómo realizar una precisa predicción de la calidad del agua sobre diferentes usos, mediante un enfoque adaptativo que permita dar soporte a los procesos de toma de decisiones sobre la gestión del recurso hídrico?

⁴ Acrónimo del inglés, “Autoregressive Integrated Moving Average”

⁵ Por su sigla en inglés, “Particle Swarm Optimization”

2 OBJETIVOS

2.1 OBJETIVO GENERAL

Evaluar un mecanismo adaptativo de predicción de la calidad del agua con un conjunto de datos recolectados por el instituto de investigación CINARA.

2.2 OBJETIVOS ESPECÍFICOS

- Definir el conjunto de datos de calidad del agua.
- Evaluar experimentalmente el mecanismo propuesto con el conjunto de datos seleccionado.
- Validar el mecanismo propuesto con personal experto en la gestión del recurso hídrico.

3 CRONOGRAMA

En la Tabla 1 se presentan las actividades que se desarrollaron durante el tiempo establecido en la pasantía de investigación.

NOMBRE ACTIVIDAD	SEMANAS			
	1	2	3	4
Revisión documental				
Actualización del reporte de revisión en el estado actual del conocimiento sobre los parámetros utilizados para determinar la calidad del agua en diferentes contextos (usos).	■			
Validación del reporte de revisión.	■			
Desarrollo de la experimentación				
Definición de las variables más representativas utilizadas en los usos del agua que se vayan a abordar.	■			
Definición, diseño y modelado de los datos de entrada.		■		
Preprocesamiento de datos.		■		
Aplicación del mecanismo de predicción adaptativo sobre los datos.		■	■	
Análisis de resultados (comparación de los resultados con los usos del agua en los cuales ya se ha aplicado el mecanismo de predicción).			■	■
Desarrollo de informe del proceso de experimentación.			■	■

Tabla 1. Cronograma de actividades

4 REVISIÓN DOCUMENTAL

4.1 Parámetros de calidad del agua para diferentes usos

Una política general de gestión adecuada de los recursos naturales incluye el uso racional o la conservación del agua, la cual debe estar enfocada primordialmente al desarrollo sostenible con miras al aprovechamiento de este recurso sin degradarlo y comprometer su futura disponibilidad. De esta manera se debe considerar la calidad del agua como el resultado de comparar las características físicas, químicas y microbiológicas encontradas en el agua, con el contenido de las normas que regulan la materia [20]; así mismo se denomina

Objetivo de Calidad al conjunto de parámetros que se utilizan para definir la idoneidad del recurso hídrico para un determinado uso.

De acuerdo con lo anterior, los estándares y objetivos de calidad de agua varían dependiendo de si se trata de agua para consumo humano, para piscicultura, para recreación, entre otros. Para efectos del cumplimiento de estos lineamientos, en Colombia se ha reglamentado el manejo y la administración del recurso hídrico en diferentes decretos que están basados en estándares internacionales. En este sentido el decreto 3930 de 2010 establece los siguientes usos del agua.

- **Consumo humano y doméstico.** Agua entregada a través de redes de agua potable para abastecer a los usuarios domésticos, bebida directa y preparación de alimentos para consumo inmediato.
- **Preservación de flora y fauna.** Agua utilizada en actividades destinadas a mantener la vida natural de los ecosistemas acuáticos y terrestres y de sus ecosistemas asociados, sin causar alteraciones sensibles en ellos.
- **Agrícola.** Utilización del agua para irrigación de cultivos y otras actividades conexas o complementarias.
- **Pecuario.** Uso para el consumo del ganado en sus diferentes especies y demás animales.
- **Recreativo.** Agua que se utiliza cuando se produce: a) Contacto primario, como en la natación, buceo y baños medicinales; y b) Contacto secundario, como en los deportes náuticos y la pesca.
- **Industrial.** Procesos manufactureros de transformación o explotación, generación de energía, minería, hidrocarburos, fabricación o procesamiento de drogas, medicamentos, cosméticos, aditivos y la elaboración de alimentos.
- **Estético.** Uso del agua para la armonización y embellecimiento del paisaje.
- **Pesca, maricultura y acuicultura.** Uso en actividades de reproducción, supervivencia, crecimiento, extracción y aprovechamiento de especies hidrobiológicas en cualquiera de sus formas, sin causar alteraciones en los ecosistemas en los que se desarrollan estas actividades.
- **Navegación y transporte.** Navegación de cualquier tipo de embarcación o para la movilización de materiales por contacto directo.

Así mismo se consideran las condiciones mínimas aplicables a todas las aguas en todos los lugares y en todos los tiempos; las cuales comprenden los cuerpos de agua que cumplan las siguientes condiciones:

- Libre de sustancias que puedan sedimentar y demande Oxígeno Disuelto.
- Libre de residuos flotantes: aceites, espumas, material flotante, etc.

- Libre de materiales que produzcan color, olor, u otras condiciones molestas.
- Libre de sustancias que sean tóxicas o peligrosas para la vida humana, animal, vegetal o acuática.

Lo anterior se resume en la Tabla 2, donde se presentan las variables estándar presentes en cualquier cuerpo de agua, al igual que el método y el equipo para su medición ya sea *in-situ* o en condiciones de laboratorio.

	Variable	Unidades	Standard methods	Equipo
Variables in situ	Sólidos disueltos totales	(mg/l)	Electrometrico 2540 c	Sonda multiparametros
	Oxigeno disuelto	(mg/l o2)	Electrometrico 4500-og	Sonda multiparametros
	Temperatura del agua	("c)	Thermistor 2550 b	Sonda multiparametros
	Ph	(unid.)	Electrometrico 4500-h b	Sonda multiparametros
	Conductividad	(ms/cm)	Electrometrico 2510 b	Sonda multiparametro
	Turbiedad	(ntu)	Espectrofotometrico 2130 b	Turbidimetro
Variables determinadas en laboratorio	Dbo5	(mg/l o2)	Dilución wincler 5210 b	Botellas wincler, incubadora
	Dqo	(mg/l)	Reflujo cerrado 5220 d	Espectofotometro
	Nitritos	(mg/l no2)	Diazotizacion - copulación 4500-no2 b	Espectofotometro
	Nitratos	(mg/l no3)	Reducción cadmio 4500-no3 e	Espectofotometro
	Sulfatos	(mg/l so4)	Cloruro de bario 4500-so4 c	Espectofotómetro
	Fósforo soluble	(mg/l po4)	Acido ascórbico 4500-p e	Espectofotómetro
	Nitrógeno amoniacal	(mg/l nh3)	Titulométrico	Montaje para titulación
	Cloruros	(mg/l cl)	Cloruro de mercurio 4500	Montaje para titulación
	Color aparente	(un. Pt.co)	Espectrofotométrico 2120 c	Espectofotómetro
	Color real	(un. Pt.co)	Espectrofotométrico 2120 c	Espectofotómetro
	Sólidos suspendidos totales	(mg/l)	Gravimétrico 2540 d	Estufa, mufla, crisoles, balanza analítica
	Coliformes totales	(ufc/100 cm3)	Filtración en membrana 9222 b	Incubadora, agar cromocult

Tabla 2. Variables básicas de calidad del agua medidas in-situ y en laboratorio. Tomado de [21].

Finalmente las variables de calidad del agua especializadas que son tenidas en cuenta por cada uso descrito, además de las anteriores, son presentadas en la Tabla 3.

Usos		Parámetros	Expresión	Valor	Mediciones Complementarias
Consumo humano y doméstico	Para potabilización se requiere solamente tratamiento convencional (Art. 38)	Amoniaco	N	1.0	Ninguna
		Arsénico	As	0.05	Ninguna
		Bario	Ba	1.0	Ninguna
		Cadmio	Cd	0.01	Ninguna
		Cianuro	CN-	0.2	Ninguna
		Cinc	Zn	15.0	Ninguna
		Cloruros	Cl-	250.0	Ninguna
		Cobre	Cu	1.0	Ninguna
		Color	Color verdadero	Escala Platino-cobalto	Ninguna
		Compuestos Fenólicos	Fenol	0.002	Ninguna
		Cromo	Cr+6	0.05	Ninguna
		Difenil policlorados	Concentración de agente activo	No. Detectable	Ninguna
		Mercurio	Hg	0.002	Ninguna
		Nitratos	N	10.0	Ninguna
		Nitritos	N	1.0	Ninguna
		pH	unidades	5.0 - 9.0 unidades	Ninguna
		Plata	Ag	0.05	Ninguna
		Plomo	Pb	0.05	Ninguna
		Selenio	Se	0.01	Ninguna
		Sulfatos	SO4 =	400.0	Ninguna
		Tensoactivos	Sustancias activas al azul de metileno	0.5	Ninguna
		Coliformes totales	NMP microorganismos/100	20.000	Ninguna
	Coliformes fecales	NMP microorganismos/100	2.000	Ninguna	
	Para potabilización se requiere solo desinfección (Art. 39)	Amoniaco	NH3	1.0	Ninguna
		Arsénico	As	0.05	Ninguna
		Bario	Ba	1.0	Ninguna
		Cadmio	Cd	0.01	Ninguna
		Cianuro	CN-	0.2	Ninguna
		Cinc	Zn	15.0	Ninguna
		Cloruros	Cl-	250.0	Ninguna
		Cobre	Cu	1.0	Ninguna
		Color	Color verdadero	20 unidades, escala Platino-cobalto	Ninguna

Usos	Parámetros	Expresión	Valor	Mediciones Complementarias
Agrícola (Art. 40)	Aluminio	Al	5.0	Conductividad (Parágrafo 2)
	Arsénico	As	0.1	Relación de absorción de sodio (RAS) (Parágrafo 2)
	Berilio	Be	0.1	Salinidad efectiva y potencial (Parágrafo)
	Boro (Parágrafo 1)	Bo	0.3 - 4.0 mg/l	Carbonato de sodio residual (Parágrafo 2)
	Cadmio	Cd	0.01	Radionucleídos (Parágrafo 2)
	Cinc	Zn	2.0	Ninguna
	Cobalto	Co	0.05	Ninguna
	Cobre	Cu	0.2	Ninguna
	Cromo	Cr+	0.1	Ninguna
	Fluor	F	1.0	Ninguna
	Hierro	Fe	5.0	Ninguna
	Litio	Li	2.5	Ninguna
	Manganeso	Mn	0.2	Ninguna
	Molibdeno	Mo	0.01	Ninguna
	Niquel	Ni	0.2	Ninguna
	pH	Unidades	4.5 - 9.0	Ninguna
	Plomo	Pb	5.0	Ninguna
	Selenio	Se	0.02	Ninguna
	Vanadio	V	0.1	Ninguna
	Coliformes totales (Parágrafo 1)	NMP microorganismos/100	5.000	Ninguna
Coliformes fecales (Parágrafo 1)	NMP microorganismos/100	1.000	Ninguna	
Pecuario (Art. 41)	Aluminio	Al	5.0	Ninguna
	Arsénico	As	0.2	Ninguna
	Boro	B	5.0	Ninguna
	Cadmio	Cd	0.05	Ninguna
	Cinc	Zn	25.0	Ninguna
	Cobre	Cu	0.5	Ninguna
	Cromo	Cr+	1.0	Ninguna
	Mercurio	Hg	0.01	Ninguna
	Nitratos + nitritos	NO3-	100.0	Ninguna
	Nitrito	NO2-	10.0	Ninguna
	Plomo	Pb	0.1	Ninguna
	Contenido de sales	Peso total	3.000	Ninguna

Usos		Parámetros	Expresión	Valor	Mediciones Complementarias
Recreativo	Contacto Primario (Art. 42)	Coliformes fecales	NMP microorganismos/100	200	Ninguna
		Coliformes totales	NMP microorganismos/100	1000	Ninguna
		Compuestos fenólicos	Fenol	0.002	Ninguna
		Oxígeno disuelto	Concentración de saturación	70%	Ninguna
		pH	Unidades	5.0 - 9.0	Ninguna
		Tensoactivos	Sustancias activas al azul de metileno	0,5	Ninguna
		No se aceptará en el recurso película visible de grasas y aceites flotantes, presencia de material flotante proveniente de actividad humana; sustancias tóxicas o irritantes cuya acción por contacto, ingestión o inhalación, produzcan reacciones adversas s			
	El nitrógeno y el fósforo deberán estar en proporción que no ocasionen eutroficación (Parágrafo 2).				Ninguna
	Contacto Secundario (Art. 43, Parágrafos 1 y 2 del Art. 42)	Coliformes totales	NMP microorganismos/100	5.000	Ninguna
		Oxígeno disuelto	Concentración de saturación	70%	Ninguna
pH		Unidades	5.0 - 9.0	Ninguna	
Tensoactivos		Sustancias activas al azul de metileno	0,5	Ninguna	
Estético (Art. 44)	Ausencia de material flotante y de espumas, provenientes de actividad humana.				Ninguna
	Ausencia de grasas y aceites que formen película visible.				Ninguna
	Ausencia de sustancias que produzcan olor.				Ninguna
Preservación de flora y fauna (Art. 45)	Sin sustancias que impartan olor o sabor a los tejidos de los organismos acuáticos, ni turbiedad o color que interfieran con la actividad fotosintética.				Ninguna
Trasporte (Art. 47)	Para los usos referentes a transporte, dilución y asimilación no se establecen criterios de calidad, sin perjuicio del control de vertimiento correspondiente.				Ninguna
Industrial (Art. 48)	Para el uso industrial, no se establecen criterios de calidad con excepción de las actividades relacionadas con explotación de cauces, playas y lechos, para las cuales se deberán tener en cuenta los criterios contemplados en el parágrafo 1 del artículo 4				Ninguna

Tabla 3. Variables de calidad del agua consideradas en cada uso. Tomado de [20].

5 DESARROLLO DE LA EXPERIMENTACIÓN

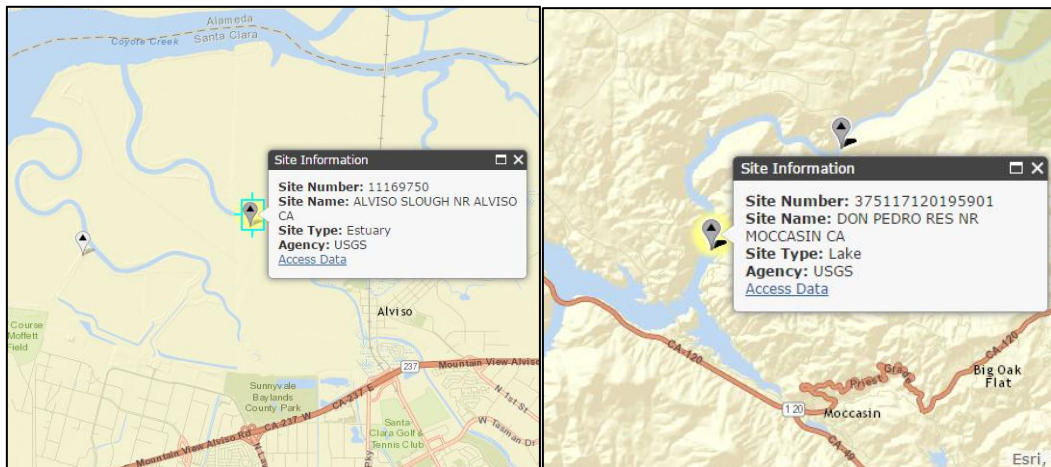
5.1 Datos y Área de Estudio.

Área de Estudio para los datos del USGS

En la primera parte del desarrollo de la tesis de maestría en la cual se enmarca este trabajo de pasantía, se establecieron dos conjuntos de datos pertenecientes al Servicio Geológico de los Estados Unidos [22] o USGS por sus siglas en inglés (United States Geological Survey). El USGS es una organización investigadora sin responsabilidades reguladoras de modo que el suministro oportuno de datos fiables de calidad del agua para el público es una de sus misiones clave, es por esto que los datos están disponibles al público desde miles de sitios a través de la nación (Estados Unidos) en los 50 estados y territorios.

Los datos incluyen variables como la Temperatura del Agua, pH, Conductividad Específica, Turbidez, Oxígeno Disuelto y/o Nitrato dependiendo del sitio de muestreo de donde se extrae el cuerpo de agua. El tipo de sitio de muestreo determina el uso del agua ya que un tipo de sitio es un lugar generalizado en el ciclo hidrológico, o una característica hecha por el hombre que puede afectar a las condiciones hidrológicas medidas en un sitio (algunos ejemplos del tipo de sitio de muestreo son: estuario, lago, planta hidroeléctrica, planta de tratamiento de suministro de agua, entre otros). Para realizar la consulta de datos fueron tenidos en cuenta parámetros como el tipo de sitio, el estado o territorio (corresponde al nombre de un estado o territorio de los Estados Unidos), el tipo de parámetros de calidad del agua (físicoquímicos) y finalmente el rango de fechas (el más amplio hasta el desarrollo de este artículo comprende desde 2007/10/01 a 2015/03/30).

Para la obtención de los datos de prueba, en este estudio fueron seleccionadas dos zonas del estado de California, la primera comprende el territorio de Alviso, la cual es una pequeña comunidad de San José, Condado de Santa Clara. El río Guadalupe y el arroyo Coyote terminan en el humedal de Alviso por medio de un estuario que desemboca en la bahía de San Francisco; una de las principales actividades es la pesca. La segunda zona de estudio es el lago Don Pedro, ubicado en el condado de Mariposa, cubre un área de 32.56 km² donde una de los principales usos es el recreacional (natación, paseo en barco y otros deportes acuáticos). La presenta las variables de calidad del agua definidas para el uso piscícola y recreacional. La Figura 1 muestra la ubicación de estos dos sitios.



(a) Estuario de Alviso

(b) Lago Don Pedro

Figura 1. Localización geográfica del área de estudio (Datos USGS).

Área de Estudio para los datos del Río Cauca

Dentro del presente trabajo de pasantía se estableció como insumo principal para el mecanismo de predicción, un subconjunto de datos de calidad del agua pertenecientes al Proyecto de Modelación del Río Cauca (PMC Fase II) el cual comprende el valle geográfico del Río Cauca, abarcando el tramo desde la represa de La Salvajina hasta el municipio de La Virginia. Es importante resaltar que la cuenca del Río Cauca es la segunda arteria fluvial más importante de Colombia y atraviesa alrededor de 183 municipios pertenecientes a los departamentos de Cauca, Valle del Cauca, Quindío, Risaralda, Caldas, Antioquia, Córdoba, Sucre y Bolívar, en donde habitan más de 19 millones de personas que representan aproximadamente el 41 % del total de la población colombiana [23].

Los principales usos del agua que circula por este río han sido la pesca, la recreación, la generación de energía, la extracción de materiales del lecho y el agua además es captada para consumo humano, riego y la industria. La calidad del agua se ha visto afectada debido a que el Río Cauca se ha usado como fuente receptora de residuos sólidos y vertimientos de aguas residuales.

La zona de estudio de donde se obtuvieron los datos de calidad del agua es el tramo Hormiguero – Mediacanoa, específicamente la estación de Puente Juanchito, muy cercana a la ciudad de Santiago de Cali. Este es el tramo en el cual se presenta la mayor contaminación por materia orgánica asociada con las múltiples descargas de aguas residuales que son vertidas a la cuenca en este tramo. La Figura 2 muestra las estaciones de monitoreo pertenecientes la cuenca del Río Cauca en el tramo Salvajina – La Virginia, las cuales están a cargo de la Corporación Regional Autónoma del Valle del Cauca (CVC).

Dentro del PMC II fueron registrados datos de Caudal, Temperatura, OD, DBO5, DQO, pH, Sólidos Suspendidos Totales y Conductividad. El periodo en que se realizaron las muestras fue de 5 días y la frecuencia varió, dependiendo de los recursos disponibles, entre 1 muestra por día y 24 por día.

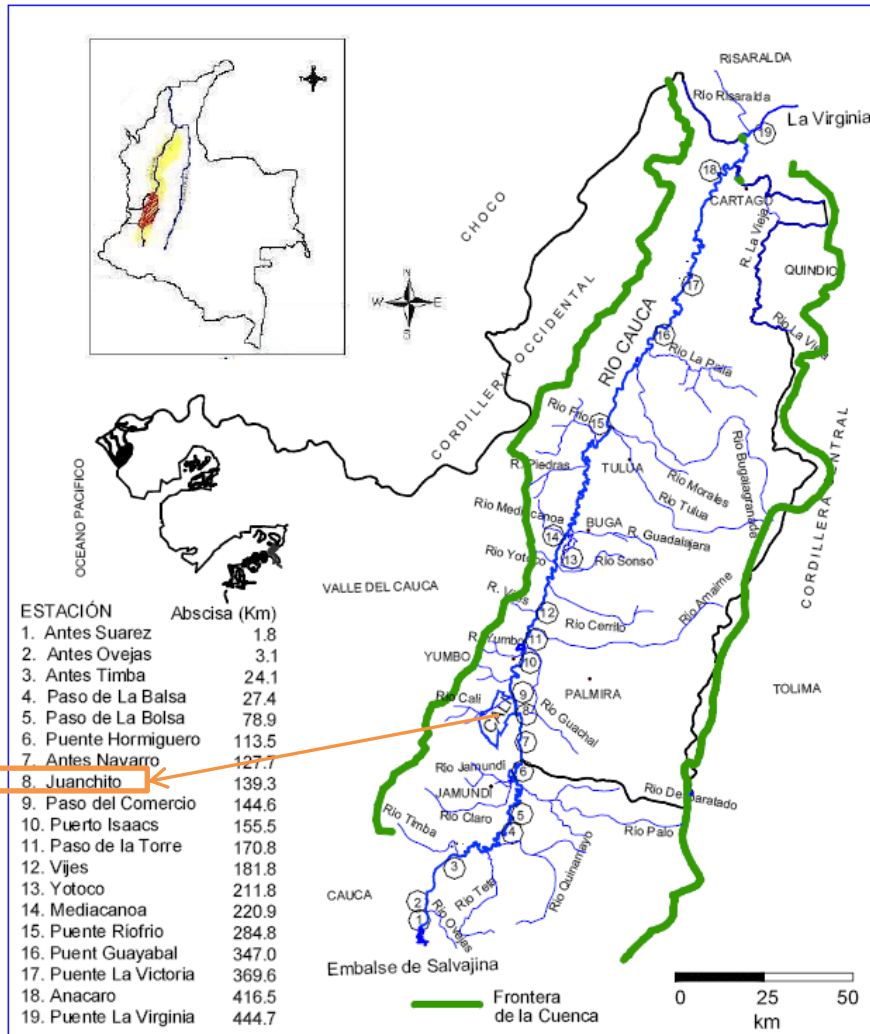


Figura 2. Estaciones de monitoreo de la calidad del agua del Río Cauca a cargo de la CVC – resaltada la estación de Puente Juanchito. Tomado de [23].

Pre-procesamiento de Datos

En la etapa de pre-procesamiento de datos se identificaron los formatos⁶ en los cuales se encontraban consignados los datos de calidad del agua y se procedió a trasladar dichos datos al formato establecido por el mecanismo de predicción, el cual utiliza los archivos de extensión .arff que maneja *Weka* para realizar el correspondiente proceso de predicción (Figura 3).

⁶ Entiéndase por “formato” al tipo de medio en el cual se encuentran almacenados los datos, por ejemplo .txt, .pdf, .xml, bases de datos relacionales, físico, etc.

USGS	11169750	2012-01-03 20:15	PST 11.0	A	19000	A	97.5	A	144	A
USGS	11169750	2012-01-03 20:30	PST 11.0	A	19000	A	62.1	A	87	A
USGS	11169750	2012-01-03 20:45	PST 11.0	A	19000	A	57.4	A	80	A
USGS	11169750	2012-01-03 21:00	PST 11.7	A	20300	A	52.3	A	72	A
USGS	11169750	2012-01-03 21:15	PST 11.7	A	20300	A	58.6	A	77	A
USGS	11169750	2012-01-03 21:30	PST 11.7	A	20400	A	52.7	A	79	A
USGS	11169750	2012-01-03 21:45	PST 11.7	A	20500	A	39.0	A	52	A
USGS	11169750	2012-01-03 22:00	PST 11.7	A	20600	A	35.2	A	46	A
USGS	11169750	2012-01-03 22:15	PST 11.7	A	20600	A	45.7	A	66	A
USGS	11169750	2012-01-03 22:30	PST 11.0	A	20400	A	70.2	A	100	A
USGS	11169750	2012-01-03 22:45	PST 11.0	A	20700	A	43.6	A	59	A
USGS	11169750	2012-01-03 23:00	PST 11.0	A	20800	A	43.3	A	58	A
USGS	11169750	2012-01-03 23:15	PST 11.0	A	20800	A	42.4	A	57	A
USGS	11169750	2012-01-03 23:30	PST 11.0	A	20800	A	39.6	A	53	A
USGS	11169750	2012-01-03 23:45	PST 11.0	A	20800	A	38.8	A	52	A
USGS	11169750	2012-01-04 00:00	PST 11.0	A	20800	A	36.4	A	48	A
USGS	11169750	2012-01-04 00:15	PST 11.0	A	20800	A	38.8	A	52	A
USGS	11169750	2012-01-04 00:30	PST 11.0	A	20800	A	38.8	A	52	A
USGS	11169750	2012-01-04 00:45	PST 11.0	A	20700	A	36.1	A	48	A
USGS	11169750	2012-01-04 01:00	PST 11.0	A	20700	A	36.3	A	48	A
USGS	11169750	2012-01-04 01:15	PST 11.0	A	20700	A	39.0	A	52	A
USGS	11169750	2012-01-04 01:30	PST 11.0	A	20600	A	38.7	A	52	A
USGS	11169750	2012-01-04 01:45	PST 11.0	A	20600	A	35.3	A	46	A
USGS	11169750	2012-01-04 02:00	PST 11.0	A	20800	A	38.2	A	51	A
USGS	11169750	2012-01-04 02:15	PST 11.0	A	20400	A	36.9	A	49	A
USGS	11169750	2012-01-04 02:30	PST 11.0	A	20400	A	36.0	A	48	A
USGS	11169750	2012-01-04 02:45	PST 11.0	A	20300	A	34.7	A	46	A
USGS	11169750	2012-01-04 03:00	PST 11.0	A	20300	A	38.0	A	50	A
USGS	11169750	2012-01-04 03:15	PST 11.0	A	20400	A	40.7	A	54	A
USGS	11169750	2012-01-04 03:30	PST 11.0	A	20400	A	40.2	A	54	A
USGS	11169750	2012-01-04 03:45	PST 11.0	A	20300	A	39.4	A	53	A
USGS	11169750	2012-01-04 04:00	PST 11.0	A	20200	A	41.4	A	56	A
USGS	11169750	2012-01-04 04:15	PST 11.0	A	20100	A	46.8	A	64	A
USGS	11169750	2012-01-04 04:30	PST 11.0	A	20100	A	48.5	A	66	A

Datos USGS (.txt)



```

1 @relation EstuarioCalifornia_11169750_5d
2
3 @attribute fecha date "yyyy-MM-dd-HH:mm"
4 @attribute temperatura numeric
5 @attribute conductancia numeric
6
7 @data
8
9 2010-10-15-00:00,20.0,6110
10 2010-10-15-00:15,19.9,4450
11 2010-10-15-00:30,19.8,3250
12 2010-10-15-00:45,19.7,2530
13 2010-10-15-01:00,19.6,2200
14 2010-10-15-01:15,19.6,2050
15 2010-10-15-01:30,19.5,2080
16 2010-10-15-01:45,19.6,2300
17 2010-10-15-02:00,19.6,2680
18 2010-10-15-02:15,19.6,3820
19 2010-10-15-02:30,19.7,5500

```

Datos USGS (.arff)

Segundo Muestreo de Calidad con Propósitos de Calibración del Modelo

Anexo 5, Base de Datos

PROYECTO DE MODELACION DEL RIO CAUCA PRC-FASE B
SEGUNDA CAMPAÑA DE MUESTREO CON PROPÓSITOS DE CALIBRACION DEL MODELO DE CALIDAD DEL AGUA DEL RIO CAUCA
BASE DE DATOS - RIO CAUCA - ESTACION: PUENTE JUANCHITO

Fecha-Hora	Tem Max	Cond	pH	Temp	OD	Conduc	DQO	BOD ₅	TSS	N-NH ₄	N-NO ₃
h	m ²	cm ²	°C	°C	mg/l	µmhos/cm	mg/l	mg/l	mg/l	mg/l	mg/l
2010-02-20 06:00	7.64	23.10	5.31	116.00							
2010-02-20 06:30	7.53	22.90	5.64	52.00							
2010-02-20 07:00	7.60	22.60	5.06	52.00							
2010-02-20 07:30	7.33	22.70	6.02	52.00							
2010-02-20 08:00	7.45	22.90	4.70	51.00							
2010-02-20 08:30	7.33	23.00	5.38	79.00							
2010-02-20 09:00	7.33	23.10	5.17	79.00							
2010-02-20 09:30	7.35	23.20	5.14	79.00							
2010-02-20 10:00	7.28	23.30	6.49	77.00							
2010-02-20 10:30	7.43	23.20	6.15	78.00							
2010-02-20 11:00	7.42	23.70	6.28	76.00							
2010-02-20 11:30	7.40	24.70	6.87	77.00							

Datos Río Cauca (.pdf)

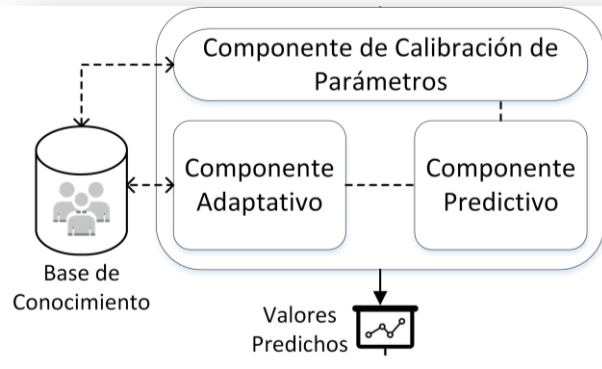
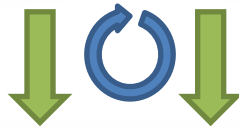


```

1 @relation RioCauca_PuenteJuanchito_5d
2
3 @attribute fecha date "yyyy-MM-dd-HH:mm"
4 @attribute pH numeric
5 @attribute temperatura numeric
6 @attribute oxigeno_disuelto numeric
7 @attribute conductancia numeric
8
9 @data
10
11 22-02-2005-06:00,7.64,23.10,5.31,116.00
12 22-02-2005-06:30,7.53,22.90,5.64,52.00
13 22-02-2005-07:00,7.60,22.60,5.06,52.00
14 22-02-2005-07:30,7.33,22.70,6.02,52.00
15 22-02-2005-08:00,7.45,22.90,4.70,51.00
16 22-02-2005-08:30,7.33,23.00,5.38,79.00
17 22-02-2005-09:00,7.33,23.10,5.17,79.00
18 22-02-2005-09:30,7.35,23.20,5.14,79.00
19 22-02-2005-10:00,7.28,23.30,6.49,77.00
20 22-02-2005-10:30,7.43,23.20,6.15,78.00
21 22-02-2005-11:00,7.42,23.70,6.28,76.00
22 22-02-2005-11:30,7.40,24.70,6.87,77.00

```

Datos Río Cauca (.arff)



Mecanismo de Predicción Adaptativo

Figura 3. Esquema de pre-procesamiento de datos. Fuente propia.

A partir de la figura anterior puede observarse que los conjuntos de datos pre-procesados (formato .arff) contienen diferentes valores para distinto número de variables fisicoquímicas; esto se debe a que dentro del componente de calibración de parámetros se realiza un proceso de selección de parámetros que en este caso corresponden a las variables físico-químicas. Esta selección permite escoger automáticamente el subconjunto de variables que mejor representan al uso del agua de cada conjunto de datos de entrada. En la se muestra las variables seleccionadas para los dos conjuntos de datos (USGS y Río Cauca).

Conjunto de Datos	Variables Seleccionadas
USGS	Temperatura (°C)
	Conductancia (µsm/cm)
Río Cauca	Temperatura (°C)
	Conductancia (µsm/cm)
	Oxígeno Disuelto (%)
	pH

Tabla 4. Variables seleccionadas por cada conjunto de datos.

Análisis de Resultados

En la etapa de evaluación se presentan los valores predichos mediante el mecanismo de predicción adaptativo que se propone como parte del presente trabajo de maestría. Este mecanismo está compuesto por la técnica de Máquinas de Vector de Soporte para Regresión (SVR) configurada con el kernel PUK (Pearson VII Universal Kernel) y el algoritmo de Optimización por Nubes de Partículas (PSO). Las pruebas fueron realizadas utilizando los mismos conjuntos de datos correspondientes al USGS (estuario de Alviso para piscicultura y lago Don Pedro para uso recreacional) y al PMC II (estación Puente Juanchito).

Las Figuras 4 a 11 muestran el comportamiento de los valores predichos para cada variable, haciendo uso del mecanismo híbrido planteado (SVR-PUK-PSO) con los conjuntos de datos de los dos usos establecidos (piscícola, recreacional y consumo humano).

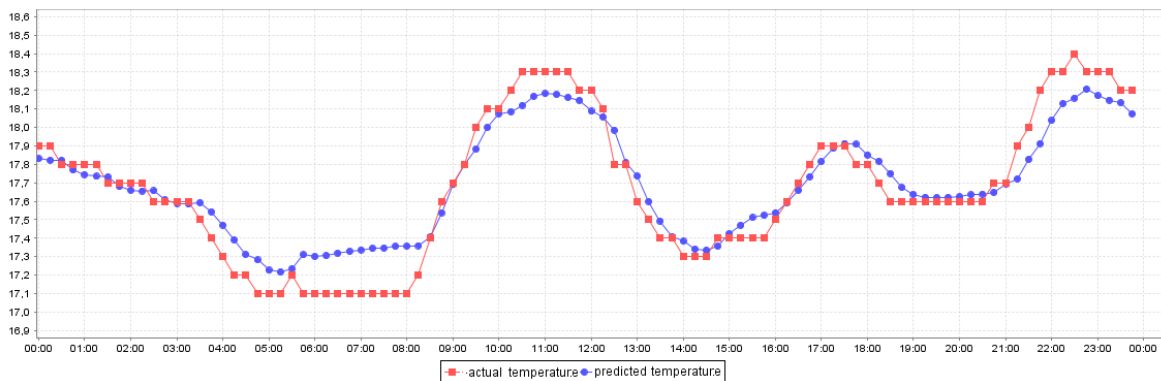


Figura 4. Valores predichos de temperatura del agua usando SVR con el kernel PUK y la técnica PSO (datos de piscicultura – estuario de Alviso).

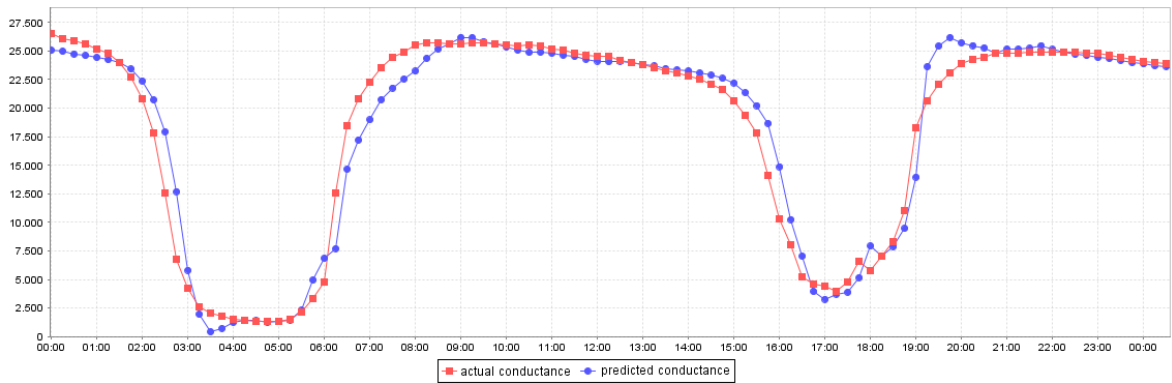


Figura 5. Valores predichos de conductividad del agua usando SVR con el kernel PUK y la técnica PSO (datos de piscicultura – estuario de Alviso).

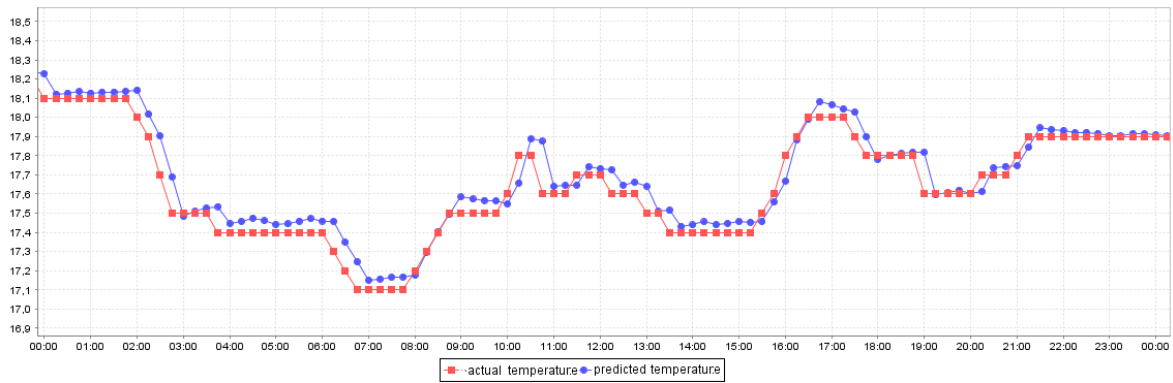


Figura 6. Valores predichos de temperatura del agua usando SVR con el kernel PUK y la técnica PSO (datos de uso recreacional – lago Don Pedro).

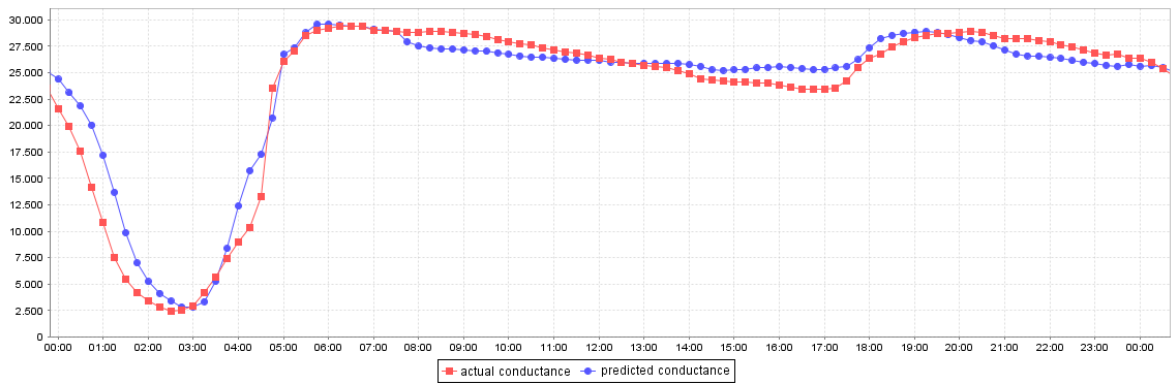


Figura 7. Valores predichos de conductividad del agua usando SVR con el kernel PUK y la técnica PSO (datos de uso recreacional – lago Don Pedro).

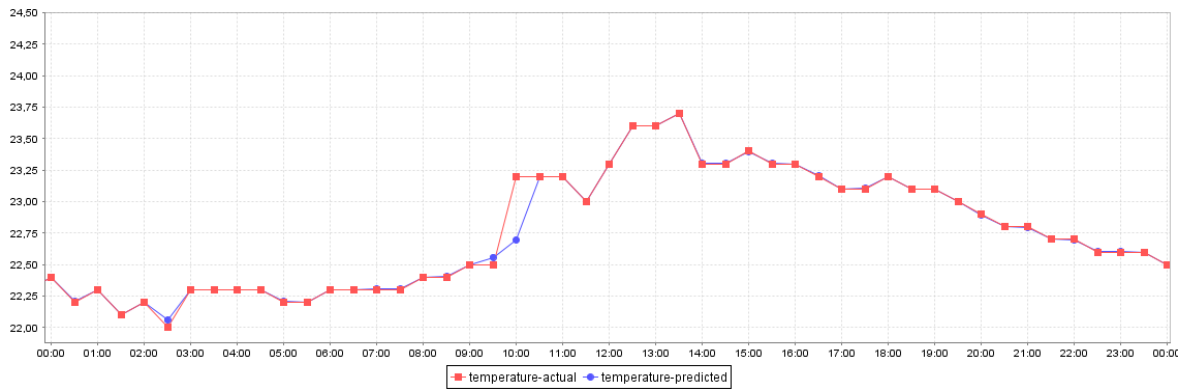


Figura 8. Valores predichos de temperatura del agua usando SVR con el kernel PUK y la técnica PSO (datos de uso para consumo humano – estación Puente Juanchito).

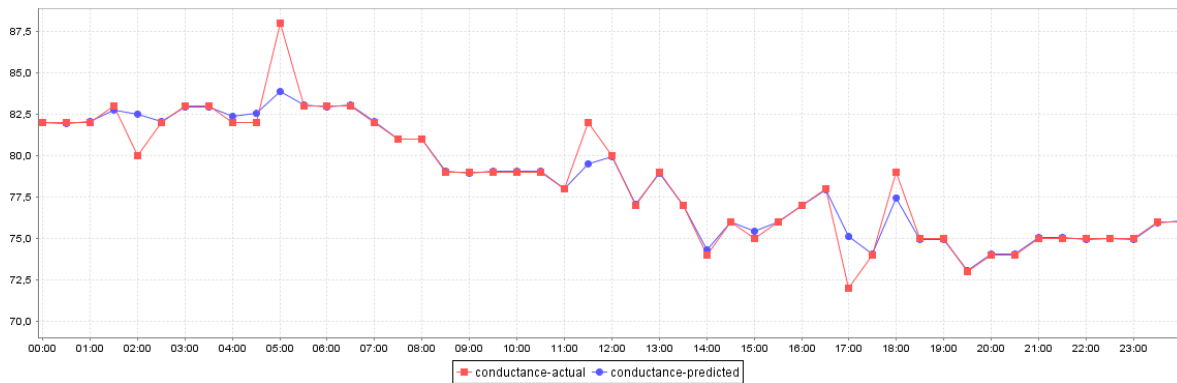


Figura 9. Valores predichos de conductividad del agua usando SVR con el kernel PUK y la técnica PSO (datos de uso para consumo humano – estación Puente Juanchito).

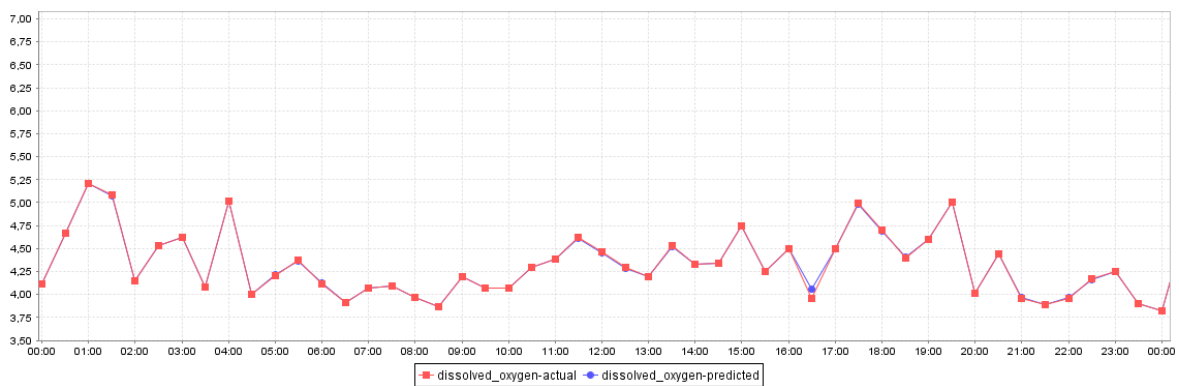


Figura 10. Valores predichos de Oxígeno Disuelto en el agua usando SVR con el kernel PUK y la técnica PSO (datos de uso para consumo humano – estación Puente Juanchito).

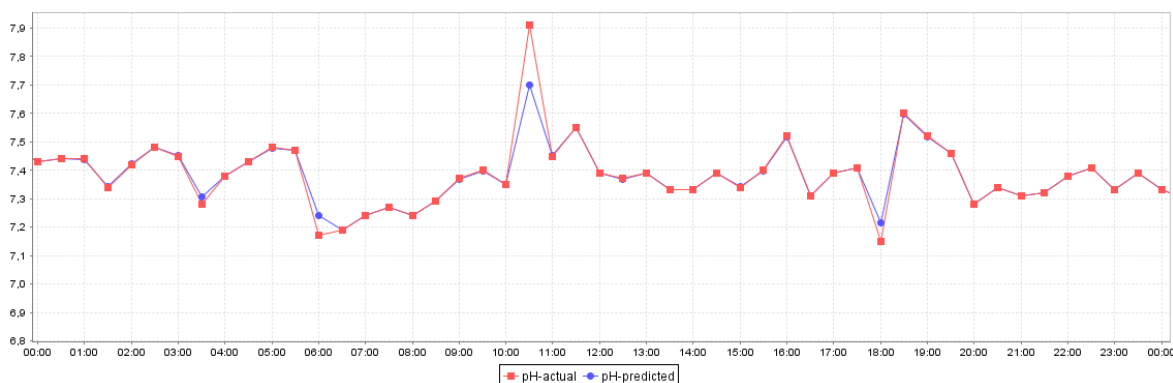


Figura 11. Valores predichos de Oxígeno Disuelto en el agua usando SVR con el kernel PUK y la técnica PSO (datos de uso para consumo humano – estación Puente Juanchito).

La Tabla 5 presenta la precisión del algoritmo SVR-PUK ajustando los valores de cada una de las variables predichas en los dos sitios definidos mediante la técnica PSO.

Conjunto de datos	Sitio de Muestreo	Variable de Calidad del Agua	MAE	RMSE	MAPE
USGS	Estuario de Alviso	Temperatura	0.1672	0.2556	0.8438
		Conductividad	1219.196	1940	8.9629
	Lago Don Pedro	Temperatura	0.0586	0.0789	0.3095
		Conductividad	1632.875	2056.435	9.0587
PMC II	Puente Juanchito	Temperatura	0.0108	0.0499	0.0473
		Conductividad	0.2805	0.8305	0.3589
		Oxígeno Disuelto	0.004	0.0079	0.0901
		pH	0.0038	0.0179	0.0514

Tabla 5. Métricas de precisión para el proceso de predicción mediante SVR-PUK-PSO. Variables de calidad del agua: Temperatura y Conductividad en el estuario de Alviso y el lago Don Pedro; Temperatura, Conductividad, Oxígeno Disuelto y pH en la estación Puente Juanchito.

Considerando los resultados que son presentados en la Tabla 5, puede observarse que mediante el ajuste del algoritmo de predicción, se obtienen resultados similares en la precisión de los valores predichos, si bien la precisión disminuye un poco con el conjunto de datos de piscicultura en el estuario de Alviso, puede considerarse que dicha disminución no es significativa si se tiene en cuenta que la precisión de los valores predichos con el conjunto de datos de uso recreacional en el lago Don Pedro, aumentó considerablemente respecto a los valores que se habían obtenido con la aplicación del algoritmo SVR-PUK, en otras palabras, el porcentaje de error disminuyó aproximadamente en un 70% para la variable temperatura del agua; y en un 50% para la variable conductancia del agua. Además de lo anterior, el mecanismo de predicción mostró un grado de precisión muy alto con el conjunto de datos PMC II – Puente Juanchito para las variables Temperatura, Oxígeno Disuelto y pH; por otro lado la predicción de valores para la variable Conductividad no fue tan precisa, sin embargo se encuentra dentro de valores aceptables de precisión teniendo en cuenta el carácter adaptativo del mecanismo propuesto.

Conclusiones

Este estudio tuvo como principal objetivo proponer, desarrollar y evaluar un mecanismo de predicción adaptativo de la calidad del agua por medio de técnicas de Inteligencia Computacional. El principal enfoque de este mecanismo es su capacidad para adaptar sus predicciones a diferentes conjuntos de datos pertenecientes a distintos usos del agua sin que la precisión se vea afectada drásticamente.

En una segunda fase de evaluación se optimizó la técnica de predicción (SVR-PUK) mediante la Optimización por Nubes de Partículas (PSO) la cual hace parte de los paradigmas de Inteligencia de Enjambres. Esta técnica permitió que los valores predichos se aproximaran a los valores reales en diferentes usos del agua (disminución del porcentaje de error en las predicciones), ya que al utilizar otro conjunto de datos (uso recreacional en el lago Don Pedro, California, USA), la precisión de las predicciones decaía ostensiblemente. La predicción realizada con los datos pertenecientes al PMC II – estación Puente Juanchito mostró una alta precisión para la mayoría de variables de calidad del agua seleccionadas lo cual permite establecer que para este nuevo conjunto de datos (uso para consumo humano) se obtiene un buen ajuste con los valores reales monitoreados.

Agradecimientos

Los autores de este trabajo agradecen al instituto CINARA, en especial al Grupo de Saneamiento Ambiental y al Grupo de Abastecimiento de Agua, al Dr. Miguel Peña, al Ing. Esp. Alberto Benavidez, al Magister Alberto Galvis y a su equipo de trabajo por los valiosos aportes sobre los datos de PMC II de calidad del agua, los cuales fueron de gran importancia en el desarrollo de esta estancia de investigación y en el complemento de la tesis de maestría que se lleva a cabo en el marco de este trabajo.

6 REFERENCIAS

- [1] P. J. and J. S. Claudia Pahl-Wostl y C. Paul Jeffrey, and Jan Sendzimir. Pahl-Wostl, *Adaptive and integrated management of water resources*. publisherNameCambridge University Press, 2011.
- [2] Consejo Económico y Social de Castilla-Mancha, «Estudio La gestión del Agua en Castilla-La Mancha». 2004.
- [3] Comunidad Autónoma de Extremadura, *Agentes Forestales de Extremadura. Legislacion Basica Ebook*. MAD-Eduforma, 2003.
- [4] C. Carbó, *Genética, patología, higiene y residuos animales*, vol. 4. Mundi-Prensa Libros, 1995.
- [5] Isis Beleño, «El 50% del agua en Colombia es de mala calidad», *UN Periódico*, Universidad Nacional de Colombia, Bogotá D.C., 12-feb-2011.
- [6] IDEAM, «Calidad del Agua Superficial en Colombia», en *Estudio Nacional del Agua*, 2010, pp. 231-277.
- [7] N. Rescher, *Predicting the Future: An Introduction to the Theory of Forecasting*. SUNY Press, 1998.
- [8] E. Kumar, *Artificial Intelligence*. I.K. International Publishing House Pvt. Limited, 2008.
- [9] C. E. Romero y J. Shan, «Development of an Artificial Neural Network-based Software for Prediction of Power Plant Canal Water Discharge Temperature», *Expert Syst Appl*, vol. 29, n.º 4, pp. 831–838, nov. 2005.
- [10] S. Palani, S.-Y. Liong, y P. Tklich, «An ANN application for water quality forecasting», *Mar. Pollut. Bull.*, vol. 56, n.º 9, pp. 1586-1597, sep. 2008.
- [11] P. A. Aguilera, A. G. Frenich, J. A. Torres, H. Castro, J. L. M. Vidal, y M. Canton, «Application of the kohonen neural network in coastal water management: methodological development for the assessment and prediction of water quality», *Water Res.*, vol. 35, n.º 17, pp. 4053-4062, dic. 2001.
- [12] D. Ömer Faruk, «A hybrid neural network and ARIMA model for water quality time series prediction», *Eng. Appl. Artif. Intell.*, vol. 23, n.º 4, pp. 586-594, jun. 2010.

- [13] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson, y J. A. Moncada-Herrera, «A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile», *Atmos. Environ.*, vol. 42, n.º 35, pp. 8331-8340, nov. 2008.
- [14] J. I. P. Di Blasi, J. Martínez Torres, P. J. García Nieto, J. R. Alonso Fernández, C. Díaz Muñiz, y J. Taboada, «Analysis and detection of outliers in water quality parameters from different automated monitoring stations in the Miño river basin (NW Spain)», *Ecol. Eng.*, vol. 60, pp. 60-66, nov. 2013.
- [15] J. Kibena, I. Nhapi, y W. Gumindoga, «Assessing the relationship between water quality parameters and changes in landuse patterns in the Upper Manyame River, Zimbabwe», *Phys. Chem. Earth Parts ABC*.
- [16] Q. Chen, W. Wu, K. Blanckaert, J. Ma, y G. Huang, «Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses», *J. Environ. Manage.*, vol. 110, pp. 116-124, nov. 2012.
- [17] M. A. T. Koçer y H. Sevgili, «Parameters selection for water quality index in the assessment of the environmental impacts of land-based trout farms», *Ecol. Indic.*, vol. 36, pp. 672-681, ene. 2014.
- [18] G. Tan, J. Yan, C. Gao, y S. Yang, «Prediction of water quality time series data based on least squares support vector machine», *Procedia Eng.*, vol. 31, pp. 1194-1199, 2012.
- [19] S. Liu, L. Xu, D. Li, Q. Li, Y. Jiang, H. Tai, y L. Zeng, «Prediction of dissolved oxygen content in river crab culture based on least squares support vector regression optimized by improved particle swarm optimization», *Comput. Electron. Agric.*, vol. 95, pp. 82-91, jul. 2013.
- [20] Ministerio de Agricultura, «Decreto 3930 de 2010». 25-oct-2010.
- [21] IDEAM, «Temas ambientales, toma, reservación de muestras», *Bogota Inst. Hidrol. Meteorol. Estud. Ambient.*, 2002.
- [22] U. S. Geological Survey, «U. S. Geological Survey», *U. S. Geological Survey*, 07-ene-2014. [En línea]. Disponible en: <http://www.usgs.gov/>. [Accedido: 14-abr-2015].
- [23] CVC, «Segunda campaña de muestreo con propositos de calibracion del modelo de calidad del agua del rio cauca», *Corporacion Auton. Reg. Val. Cauca Caracterizacion Model. Mat. Río Cauca -PMC- Fase II Conv. Interadministrativo 0168 Noviembre 27 2002*, vol. XV, 2005.