



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

CENTRO UNIVERSITARIO UAEM VALLE DE MÉXICO

**Modelo Computacional para la Estimación de Oxígeno
Disuelto en Estanques de Producción Acuícola Empleando
Redes Neuronales Artificiales**

TESIS

Que para obtener el Grado de

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

Presenta

Ing. Carlos Julián Torres González

Tutor Académico:

Dr. Víctor Manuel Landassuri Moreno

Tutores Adjuntos:

Dr. José Martín Flores Albino

Dr. José Juan Carbajal Hernández



Atizapán de Zaragoza, Edo. de México, enero de 2017

Resumen

Actualmente el cultivo de especies acuícolas para consumo humano es una práctica que se realiza en todo el mundo. De ahí, el oxígeno disuelto es uno de los parámetros más importantes de la calidad del agua en sistemas acuícolas, por lo que mantener buenos niveles de concentración de este gas es esencial para una producción exitosa.

En este trabajo de tesis se propone el desarrollo de un modelo computacional para la estimación de oxígeno disuelto utilizando Redes Neuronales Artificiales (RNAs), las cuales realizarán predicciones para determinar en periodos futuros los efectos negativos en la calidad del agua. El diseño de las RNAs está basado en Algoritmos Evolutivos (AEs), particularmente en el algoritmo llamado: Selección de Características en el Algoritmo de Programación Evolutiva de Redes Neuronales Artificiales del Inglés Feature Selection of Evolutionary Programming of Artificial Neural Networks (FS-EPNet) [15], el cual determinará la arquitectura de la red, donde la función objetivo será la predicción en un lapso determinado de tiempo.

Aunque el análisis de la calidad del agua se ve afectado por varios parámetros, este trabajo considera únicamente la predicción del oxígeno, utilizando dos formas de predicción: predicción a un paso adelante y predicción iterada. Esto permitirá sentar las bases para futuras investigaciones sobre predicciones multiparamétricas, análisis del estado de la calidad del agua y control predictivo de la calidad del agua.

1. Introducción

Uno de los factores que más influyen en la producción de los sistemas acuícolas es la calidad del agua del cultivo en cuestión, de modo que su manejo apropiado juega un papel fundamental en el éxito de las operaciones acuícolas [5, 17]. De esta forma, uno de los parámetros de mayor importancia en la calidad del agua es el oxígeno disuelto (OD), mismo que determinará el estado general de la producción acuícola ya que tiene una influencia directa en la ingesta de alimento, la resistencia a enfermedades y el metabolismo, en consecuencia se vuelve crítico mantener valores adecuados de este gas disuelto en el agua [1, 16, 17]. Existen varios factores que modifican las concentraciones de OD, un ejemplo de ello es el ciclo dinámico del oxígeno en estanques, el cual fluctúa en el transcurso del día principalmente debido a la fotosíntesis y respiración del fitoplancton y camarones [1]. En contraparte, la solubilidad del oxígeno se ve afectada en gran medida por la temperatura y la salinidad del agua [1, 3]. Factores secundarios como el pH, amonio, demanda bioquímica y biológica de oxígeno, entre otros, también tienen relación en las concentraciones de este gas [1, 2, 7]. Cabe destacar que durante la noche y en la madrugada los niveles de oxígeno disminuyen de forma natural, pero dependiendo de que tan bajos estén los niveles de éste gas, ya sea durante el día o la noche, puede provocar que los organismos se alimenten menos, presenten una pobre asimilación de alimento, crecimiento más lento, ser más susceptibles a enfermedades infecciosas, estresarse e incluso presentar muertes masivas [1, 5, 10]. Así, existen normas y criterios de cultivo de especies acuícolas las cuales son publicadas por diversos organismos de sanidad: la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación [21], la National Sanitation Foundation [20] entre otros.

Dada la importancia del oxígeno disuelto, en esta tesis se propone la creación de un modelo computacional para la predicción del éste gas empleando Programación Evolutivo de Redes Neuronales (EPNet [23]), a través del algoritmo evolutivo FS-EPNet [15], permitiendo así conocer el deterioro de la calidad del agua en un cultivo de camarón blanco. Mediante un conjunto de muestras obtenidas de estanques en funcionamiento, se entrena la red propuesta y se verifica su rendimiento predictivo. El resto de este trabajo está organizado de la siguiente forma: la Sección 2 muestra la relación del oxígeno disuelto y la calidad del agua, así como se describe la forma en la que se recolectaron los datos de esta investigación. La sección 3 presenta a grandes rasgos el algoritmo evolutivo FS-EPNet. La sección 4 muestra el modelo computacional desarrollado, donde se muestra primero la etapa de recolección de muestras y posteriormente la configuración experimental del trabajo. Los Resultados se pueden encontrar en la Sección 5 y por último las conclusiones en la Sección 6.

2. Oxígeno disuelto para una buena calidad del agua

La acuicultura es definida como el cultivo de organismos acuáticos como peces, moluscos, crustáceos entre otros. El hábitat es controlado con el fin de aumentar la producción, simulando las condiciones generales que dichas especies tienen en su ambiente silvestre [2]. Las concentraciones de oxígeno disuelto son un factor preponderante en la producción de camarón blanco debido a la importancia que tiene en su desarrollo y en la resistencia a

enfermedades [22]. Uno de los principales parámetros que influye en la cantidad de oxígeno contenido en el agua es la temperatura; si ésta aumenta, la tasa metabólica será mayor [10] generando una mayor demanda de oxígeno [2, 3]. Asimismo, la solubilidad del oxígeno disminuye a medida que aumentan las sales disueltas [1]. La mayor cantidad de oxígeno en el agua en estanques de cultivo proviene del proceso de fotosíntesis realizado por las plantas y algas microscópicas. Este proceso depende de la cantidad de luz que pueden disponer los vegetales. Debido a ello, los niveles de OD fluctúan durante el día, elevándose a lo largo de la mañana y alcanzando su punto máximo por la tarde [3, 17]. Los niveles aceptables de OD deben ser mayores a 6 mg/l, ya que con niveles inferiores se puede presentar lento crecimiento, llegando incluso a condiciones de hipoxia debajo de los 2 mg/l donde en periodos prolongados es letal para los organismos.

3. Evolución de Redes Neuronales Artificiales

En este trabajo, las Redes Neuronales Artificiales (RNAs) se utilizan para evaluar los niveles de riesgo de oxígeno en granjas extensivas de camarón debido a que han demostrado ser un modelo de aprendizaje altamente eficiente. Sin embargo, la construcción de una RNA no es una tarea trivial, siendo en sí un problema combinatorio el cual es considerado dentro de los problema NP-Complejos, por todas las variables que se tienen que ajustar. Por otro lado, los factores mencionados en la Sección 2 afectan directa o indirectamente las concentraciones de oxígeno disuelto ocasionando un comportamiento no lineal del sistema, donde dicha dinámica puede ser procesada por RNAs encontrando relaciones no lineales y permitiendo estimaciones útiles de éste gas. Así, los algoritmos evolutivos permiten adaptar durante evolución parámetros de las RNAs, como pueden ser el número de entradas, el número de nodos ocultos y su conectividad entre ellos. De esta forma, este trabajo utiliza el algoritmo evolutivo FS-EPNet [17], el cual es una mejora del EPNet [23] propuesto originalmente, donde este último no permite la evolución de las entradas, y éstas tienen que ser ajustadas por el experto. Así, el algoritmo FS-EPNet permite adaptar automáticamente las entradas de la red, ajustándolas automáticamente a través del proceso evolutivo, en donde se cuenta con una población de individuos de RNAs las cuales compiten y evolucionan para mejorar la predicción del OD. A través del proceso de adaptación de nodos y conexiones, es posible encontrar más de una capa oculta, dado el modelo de Generalized Multi-layered perceptron (Generalized MLP-GMLP) el cual elimina la restricción de tener capas adyacentes conectadas, a diferencia del modelo convencional (MLP).

El entrenamiento es llevado a cabo con el algoritmo de Backpropagation modificado (MBP), el cual adapta la tasa de aprendizaje de generación en generación, así posterior a una mutación (eliminar o agregar nodos o conexiones), disminuye o incrementa esta variable de acuerdo al rendimiento de la red. Este algoritmo está diseñado para dar prioridad a RNAs pequeñas. Así, primero da prioridad a las mutaciones de eliminación antes de las mutaciones de incremento o agregación de elementos. Es decir, en cada generación es seleccionada una RNA para ser mutada, donde primero se entrena parcialmente para ver si el error puede ser reducido como se comentó anteriormente (MBP). De no ser el caso, se crean nuevos individuos a partir de él para eliminar *entradas*, *nodos ocultos* o *conexiones*, si alguna de estas mutaciones tiene éxito, lo cual significa que el error se

reduce, el nuevo individuo es insertado en el lugar de la red seleccionada. Si ninguna mutación de eliminación funciona, el algoritmo intenta agregar elementos de *entradas*, *nodos ocultos* y *conexiones*, quedándose con la red que mejor rendimiento otorgue (para predicción en este caso), lo que significa que remplazará a la red escogida inicialmente. Por último, el “bias” es un parámetro que puede ser evolucionado. Por cuestiones de espacio, no se da una descripción más detallada del algoritmo ni se presenta su diagrama de flujo, sin embargo para mayor detalle se puede referir a [17].

4. Diseño de modelo de predicción

2.1 Recolección de muestras

Con el fin de crear una base de datos para probar el modelo el oxígeno disuelto se midió éste mediante un sensor en un estanque de pruebas de la finca Rancho Chapo ubicada en Sonora, México.

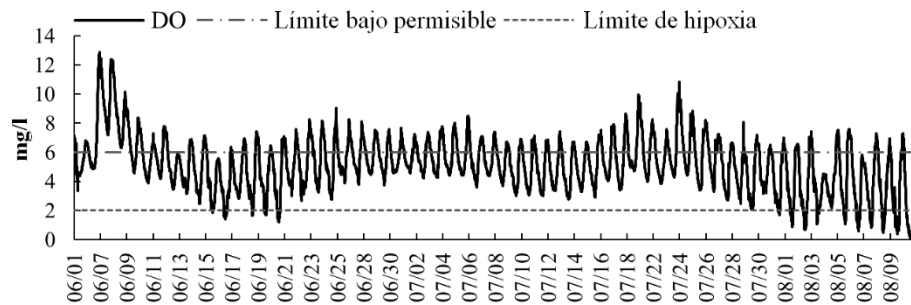


Figura 1. Comportamiento diario del oxígeno durante un periodo de cultivo

El período de seguimiento de las muestras fue de 15 minutos a lo largo de tres meses de mediciones; esto es, el conjunto de datos está formado por 8736 valores (junio, julio y agosto de 2007). Sin embargo, las fallas en el sensor generaron registros falsos, por lo que fueron borrados de la base de datos de prueba, quedando 5952 registros para este trabajo. El comportamiento de estos registros se puede observar en la figura 1, donde se aprecia que los límites permisibles son sobrepasados, generando malas condiciones de calidad del agua. Es importante remarcar que el oxígeno disuelto presenta concentraciones de hipoxia, lo que es un problema extremadamente peligroso en cualquier tipo de sistema de cultivo acuícola.

3.1 Configuración experimental

Se realizaron pruebas con 20, 21, 30, 50, 100, 500, 1000 y 1500 pasos adelante en la predicción, y con diversos parámetros para de configuración como generaciones, número de individuos en la población, entre otros. Derivado de diversos experimentos previos, se determinaron los siguientes parámetros, se determinaron las redes que mejor responden para cada configuración. Así, la configuración experimental para una de ellas es la siguiente: de las 5296 mediciones disponibles, se tomaron los primeros 4200 valores, donde los primeros 3000 (56% del total) son usados para el diseño y entrenamiento de las redes con el algoritmo evolutivo, y los siguientes 1200 (22% del total) fueron usados para

probar las redes resultantes del proceso evolutivo, y seleccionar la mejor red neuronal que resuelva el problema, es decir, al término de la evolución todas las redes encontradas (población de individuos) fueron probadas con este conjunto de prueba de 1200 patrones. En cuanto a las generaciones del algoritmo evolutivo se encontró que 4000 generaciones eran adecuadas para lograr reducir considerablemente el error en comparación con la primera generación (primera población de RNAs), donde más generaciones no representaba un aumento significativo en el rendimiento, pero si en el tiempo de ejecución. Se utiliza un modelo de predicciones iteradas MSP (Multiple Step Predicción), que consiste en predecir nuevos valores sobre los valores anteriormente predichos. Por otro lado, la determinación del tipo de funciones de transferencia utilizadas en el proceso de construcción de la RNA es importante, debido a que definen las características de salida de todo el modelo neuronal. En este caso, se utilizó una función de transferencia lineal en la capa de salida, $f(x) = x$, y *tangente hiperbólica* en las capas intermedias, $f(x) = \tanh(ax)$. En cuanto al pre-procesamiento de los datos, el entrenamiento de redes neuronales puede ser más eficiente si eliminamos tendencias, por ejemplo, los parámetros de OD tienen diferentes rangos de valores como se mostró en la figura 1. Donde no es conveniente usarlos de esta forma, ya que valores más grandes tienen un mayor impacto y provocan una tendencia en el resultado final [13, 16]. Así, en primer lugar, una estandarización transforma los datos para lograr una media de cero, con la unidad de desviación estándar como sigue:

$$X^k = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

Después, los parámetros estandarizados fueron normalizados en un rango de [-0.9, 0.9] de acuerdo a la siguiente ecuación:

$$X^\omega = (N_{\max} - N_{\min}) \frac{X_i^k - \min(X^k)}{\max(X^k) - \min(X^k)} + N_{\min} \quad (2)$$

en donde N_{\max} y N_{\min} son los valores máximo y mínimo del rango deseado y X^ω es el valor normalizado con el cual se construirán los patrones de entrenamiento y prueba posteriores. Cabe resaltar que esta normalización tiene efecto a fin de satisfacer los rangos de las funciones de transferencia usadas. La métrica de errores usada para determinar la precisión en la predicción de las RNAs fue el NRMSE (*Normalized Root Mean Square Error*).

5. Resultados y discusión

La figura 2 muestra la mejor RNA encontrada por el algoritmo evolutivo para una de las configuraciones experimentales realizadas. Esta red presenta una topología (6-6-2-3-1), esto es: 6 nodos de entrada, 6 en la primera capa oculta, 2 en la segunda capa oculta, 3 en la tercera capa oculta y 1 en la capa de salida. Donde es bueno recordar que es una arquitectura obtenida con el algoritmo FS-EPNet usando un modelo GMLP, y las 6 entradas (nodos 1 al 6) de esta red corresponden a 6 valores anteriores al valor a predecir del OD, es decir los nodos del 1 al 6 toman valores pasados para predecir el siguiente. En su contraparte, los nodos del 7 al 17 son ocultos repartidos en 3 capas ocultas y el nodo 18

proporciona la salida de la RNA. Se indica con una “b” los nodos que tienen un bias asociado, para este caso, del nodo 7 al 18.

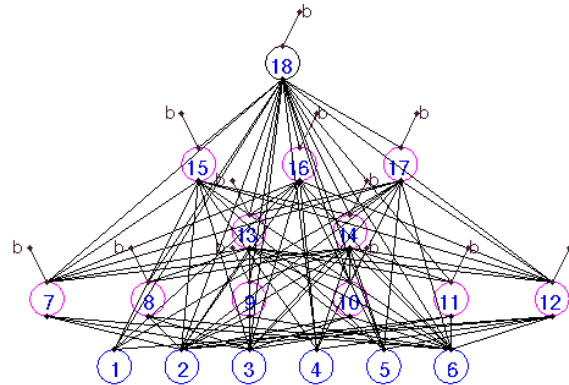


Figura 2. Mejor RNA encontrada para la predicción de 1200 pasos adelante del OD en el cultivo de camarón blanco.

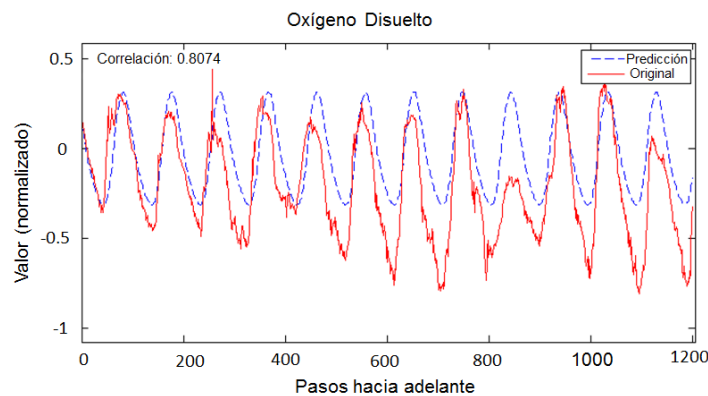


Figura 3.- Comparación de valores medidos de OD y predichos de la red neuronal

En la figura 3 se presenta la predicción realizada con 1200 pasos hacia adelante, la cual muestra un buen acercamiento a los valores reales presentando un $\text{NRMSE} = 0.7956$. Asimismo, se muestra una respuesta aceptable y una alta correlación, sin embargo, la curva de predicción parece estar marcando la tendencia del comportamiento de los datos originales, esto es, posiblemente por la gran cantidad de datos de entrenamiento y los pasos adelante a predecir (1200). En su contraparte, la figura 4a – 4c, muestra el promedio de las conexiones (4a), nodos ocultos (4b) y entradas (4c) durante las 4000 generaciones de evolución. A medida que aumentan las generaciones hay un incremento en el número de esas 3 variables, al final la figura 4c alcanza 3 nodos de entrada en promedio sobre la población entera de individuos a evolucionar; sin embargo, la mejor RNA presenta 6 neuronas de entrada como se mostró anteriormente. Observando el comportamiento del error durante evolución (figura no mostrada aquí) y las figura 4a – 4c, se nota un incremento de todas las variables mientras que el error promedio durante evolución disminuye y no aumenta, lo cual es una ventaja de usar un algoritmo evolutivo como el usado en este trabajo, en lugar de diseñar las arquitecturas de RNAs por el experto humano a mano (Hand

design Neural Networks). En la figura 4d se muestra el error de la predicción y de los datos reales. Ahí se puede ver que a medida que van avanzando los pasos a predecir se va presentando más disparidad entre las predicciones y los valores reales, en donde se puede apreciar que el error es cercano a cero al inicio de las predicciones y a medida que se incrementan el horizonte de predicción el error va aumentando, esto es debido al modelo de predicción utilizado (MSP), el cual predice los siguientes valores usando valores predichos previamente.

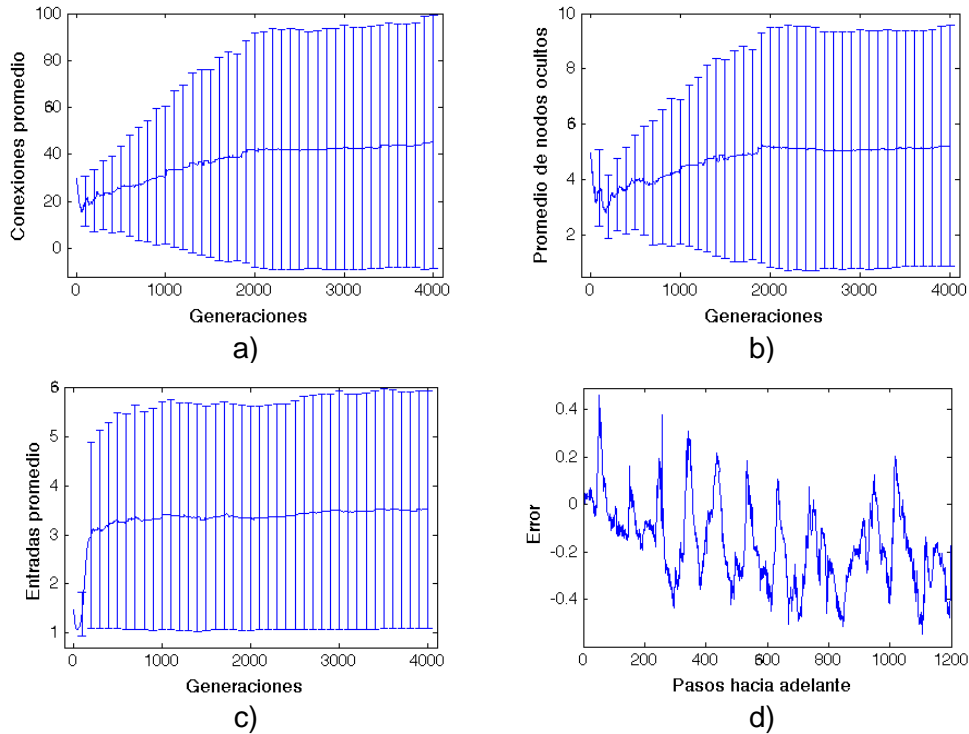


Figura 4.- Barras de Error (desviación estándar) durante evolución de: conexiones (a), nodos ocultos (b), entradas (c). La figura (d) muestra el error entre las mediciones reales y los predichos por la red neuronal.

Para eliminar la aparente tendencia en la predicción con 1200 pasos adelante se realizan pruebas con menos puntos a predecir. La mejor respuesta se obtiene con 21 pasos a predecir en dicho horizonte (ver Figura 5), donde la gráfica de datos original corresponde a los últimos 21 datos (1179-1200) de la gráfica de datos original de la figura 3. Diferentes modelos se han desarrollado y utilizado para analizar el OD.

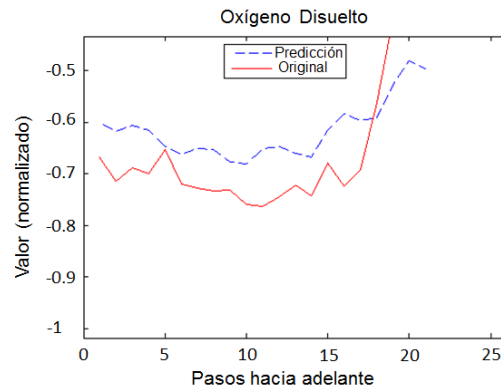


Figura 5.- Comparación de valores medidos de OD y predichos de la red neuronal, con 21 pasos adelante.

Así, el uso del algoritmo FS-EPNet en el diseño de las RNAs permite obtener una arquitectura optimizada con un rendimiento aceptable, y dado el proceso evolutivo y el entrenamiento parcial con el MBP se evita caer en mínimos locales, todo esto sin la intervención de un experto. No es posible hacer una comparación directa con los trabajos citados anteriormente dado el origen diverso de los datos, sin embargo se espera que en trabajos futuros se puedan comparar.

Para verificar los beneficios de implementar el FS-EPNet en el diseño de la RNA, se realizaron experimentos en donde la arquitectura de la red se realizaba con distintos parámetros establecidos manualmente.

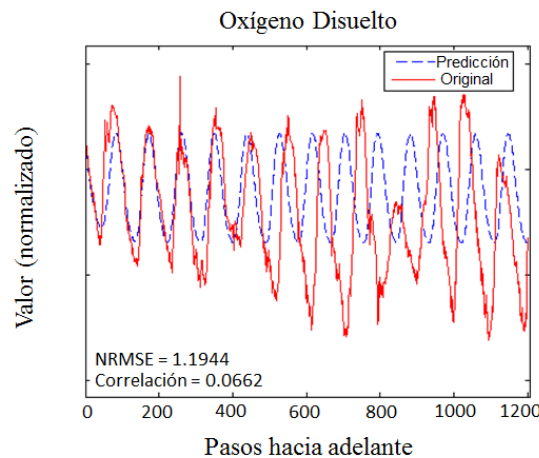


Figura 5.- Predicción con RNA hecha manualmente. Para MSP y 1200 pasos adelante.

En la figura 5 se presenta la predicción realizada, la cual muestra un ajuste de la curva predicha a la real más deficiente que en los experimentos utilizando en algoritmo FS-EPNET, con un $NRMSE = 1.1944$. El índice de correlación de forma contraria, es el menor que con el algoritmo FS-EPNET con una correlación de 0.0662. Aproximadamente a partir de los 50 pasos adelante, se comienza a marcar una tendencia cíclica de la predicción

6. Conclusiones

El estudio de la calidad del agua cobra gran relevancia para los cultivos acuícolas y por ello ha sido estudiada ampliamente.

La forma más común de monitorear el OD constantemente se realiza tomando lecturas con un sensor (i.e. oxímetro) y registrarlas de forma automática, donde se podrán identificar o esperar lecturas altas al comenzar la noche, así como las lecturas bajas muy temprano por las mañanas, donde por ejemplo, lecturas bajas en las tardes pondrán alertar sobre problemas que pudieran ocurrir a futuro [5,6]. Publicaciones como [1,5] dan métodos que son de suma utilidad para tomar decisiones en el manejo y cuidado de la calidad del agua, y a pesar de que son de utilidad, no funcionan o sirven para predecir descensos de oxígeno en todas las situaciones, especialmente las causadas por la muerte súbita de algas; viraje o circunstancias inusuales. De ahí la necesidad de métodos más robustos. Así, la importancia de la concentración de OD no puede ser subestimada, ya que es probablemente uno de los parámetros más importantes que influye en el bienestar del organismo acuático.

El modelo con RNA propuesto en este trabajo permite realizar una predicción adecuada como se mostró en el capítulo 4 de resultados, donde se puede concluir que dichas predicciones pueden ser útiles al momento de tomar decisiones en la acuicultura. En otro aspecto, al usar el algoritmo evolutivo FS-EPNet permite automatizar la creación de ellas (arquitecturas), reduciendo el trabajo y tiempo para encontrar la estructura de la red neuronal adecuada. Así las RNAs evolucionadas permiten abordar de una mejor manera el problema de no linealidad del OD y así anticipar con mayor precisión el estado de la calidad del agua del cultivo, dado que el proceso evolutivo permite ajustar de una mejor forma valores que serían difíciles de ajustar por un experto (Hand Design Neural Networks). Por lo anterior se concluye que es conveniente el uso del FS-EPNet al diseñar RNAs como alternativa para controlar el estado de la calidad del agua.

Así, el uso del algoritmo FS-EPNet en el diseño de las RNAs permite obtener una arquitectura optimizada con un rendimiento aceptable, y dado el proceso evolutivo y el entrenamiento parcial con el MBP se evita caer en mínimos locales, todo esto sin la intervención de un experto. No es posible hacer una comparación directa con los trabajos citados anteriormente dado el origen diverso de los datos, sin embargo se espera que en trabajos futuros se puedan comparar

Referencias

1. La Oxigenación en la Acuicultura. *AquaFeed Internacional*, <http://aquafeed.co/la-oxigenacion-en-la-acuicultura/> (2012). Accedido el 26 de Octubre del 2015.
2. Areerachakul S., Junsawang P., Pomsathit A.: Prediction of Dissolved Oxygen Using Artificial Neural Network. *2011 International Conference on Computer Communication and Management*. 5, 524 – 528 (2011).
3. Boyd B. C., Handson T.: Dissolved-Oxygen Concentrations In Pond Aquaculture. *Global Aquaculture Advocate*. January/February 2010, 40 – 41 (2010).
4. Clear Blue Environmental, <http://clrbu.com/industry/aquaculture-fish-farm-aeration/> . Accedido el 5 de Febrero del 2016
5. Doaa M. A, Faten H. F.: Design and control strategy of diffused air aereation system. *The Online Journal of Power and Energy Engineering* 3, 2, 284 - 290 (2013)
6. Datta S.: Management of wáter quality in intensive aquaculture. *Researchgate*. 1-17 (2012)
7. Fraser, A. S.: Simulation of genetic systems by automatic digital computers. II: Effects of linkage on rates under selection, *Austral. J. Biol.*, 10, 492–499 (1957)
8. Freeman J., Skapura J.: Redes neuronales. Algoritmos, aplicaciones y técnicas de programación. <http://unilectura.blogspot.mx/2013/08/redes-neuronales-algoritmos.html> . (1991)
9. Goyenola G.: Oxígeno Disuelto, Junio. *RED MAPSA*, 1.0, 3 (2007)
10. Haykin S.: Neural networks. A cromprehensive foundation. Pearson, 2 (2005)
11. Holland, J. H., 1975, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.
12. Ibañez. J.: Conceptos y tipos de modelos científicos. Madrid <http://www.madrimasd.org/blogs/universo/2008/05/10/91441> (2008) Accedido el 20 de Julio del 2016
13. Kriesel D.: Neural Networks. A brief introduction. *Dkdriese.com* (2005)
14. Landassuri V. M., Bustillo C. L., Carbajal J. J., Sánchez L. P.: Single-Step-Ahead and Multi-Step-Ahead Prediction with Evolutionary Artificial Neural Networks. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes in Computer Science*, 8252, Springer, 65-72 (2013)
15. Lee G. P.: Process control and artificial intelligence software for aquaculture. *Aquaculture Engineering*, 23 (2000)
16. Lekang O. I.: *Aquaculture Engineering*. Blackwell Publishing (2007)
17. Mendez P., Ibarra J.: Implementación de una red neuronal de convolución para el reconocimiento de poses en imágenes de rostros. *Avances en ciencias e ingeniería*, 6, 2, A1-P6 (2004)
18. Michalewicz Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Heidelberg, (1996)
19. NSF Internaional, <http://www.nsf.org> Accedido el 26 de Junio del 2015
20. SAGARPA. Normatividad, http://www.sagarpa.gob.mx/agricultura/Programas/proagro/Normatividad/Paginas/Normatividad_2014 Accedido el 24 de Junio del 2015
21. Tucker C.: Pond Aeration. *Southern regional aquaculture center SRAC*. Septiembre, 3700, 8. (2005)
22. Yao X., Liu Y.: A new evolutionary system for evolving artificial neural networks. *IEEE Transactions on Neural Networks* 8, 3, 694-713 (1997)
23. Stefan H. G., Fang X., Wright D., Eaton J. G., McCormick J. H.: Simulation of dissolved oxygen profiles in a transparent, dimictic lake, *Limnology and Oceanography*, 4, 105-118, January 1995.
24. Boano F., Revell R., Ridolfi L.: Stochastic modelling of DO and BOD components in a stream with random inputs, *Advances in Water Resources*, 29, 1341-1350, September 2006.
25. Faruk D. O.: A hybrid neural network and ARIMA model for wáter quality time series prediction, *Engineering Applications of Artificial Intelligence*, 586-594, January 2010.