

Adaptación de algoritmo evolutivo mediante el método de evolución diferencial aplicado para problemas de la calidad del aire

Hernández Vega José Isidro, González Rodríguez Mario, Reyes Varela Elda, Reynoso Guajardo Luis Alejandro, Palomares Gorham Dolores Gabriela.

Tecnológico Nacional de México/Instituto Tecnológico de Nuevo León
División de Estudios de Posgrado e Investigación, Departamento de Sistemas y Computación
Ave. Eloy Cavazos 2001, Col. Tolteca, Guadalupe, N.L. C.P. 67170, México
Tel:(81)-8157-0500 Ext 140,161

jose.hv@nuevoleon.tecnm.mx, mariogzz1409@hotmail.com, elda.rv@nuevoleon.tecnm.mx,
luis.rg@nuevoleon.tecnm.mx, dolores.pg@nuevoleon.tecnm.mx

RESUMEN

Este artículo aborda la problemática de monitoreo de contaminantes criterio en una zona específica. Se implementó un algoritmo evolutivo diferencial para búsqueda óptima de concentración del contaminante Partículas Suspensas de 10 micras (PM10), facilitando el análisis de los datos recabados en la detección de las concentraciones máximas y mínimas del contaminante en un periodo de tiempo determinado. Mediante datos históricos fue posible conocer el comportamiento del contaminante de estudio. Los resultados obtenidos con el algoritmo evolutivo permitieron dar estimaciones de los niveles del contaminante en diferentes horas bajo un esquema de predicción.

Palabras clave: Algoritmos evolutivos, evolutivo diferencial, monitoreo contaminantes, partículas suspendidas.

ABSTRACT

This work has the purpose of contributing to the solution of the problem of monitoring criteria pollutants in a specific area. A differential evolutionary algorithm was implemented to search for the optimal concentration of the pollutant Suspended Particles of 10 microns (PM10), facilitating the analysis of the data collected in the detection of the maximum and minimum concentrations of the pollutant in a period of time analyzed. Using historical data, it was possible to know the behavior of the study pollutant. The results obtained with the evolutionary algorithm allowed estimating the levels of the pollutant at different times under a prediction scheme.

Keywords: Evolutionary algorithms, differential evolutionary, pollutant monitoring, suspended particles.

1. INTRODUCCIÓN

La calidad del aire y estudios sobre su efecto en la salud humana, han demostrado que la exposición a contaminantes en el aire ambiente son motivo de enfermedades respiratorias en las personas, particularmente en niños, mujeres embarazadas y adultos mayores, llegando a ser en la actualidad, una importante causa de mortalidad en el mundo [1].

La contaminación del aire afecta la calidad de vida de las personas, especialmente de las que habitan en zonas urbanas de la república mexicana, por esta razón es importante monitorear la calidad del aire y que los residentes de estas zonas eviten exponerse para reducir

riesgos en su salud. En este sentido, el monitoreo de la calidad del aire toma relevancia de tal manera que las instancias gubernamentales puedan informar a la sociedad sobre los niveles de contaminación ambiental y así puedan definirse estrategias que protejan a sus ciudadanos y se tomen medidas regulatorias al respecto.

La contaminación en el Área Metropolitana de Monterrey (AMM) alcanzó niveles de alarma durante el 2019 y 2020, declarando la alerta ambiental en múltiples ocasiones.

El análisis de la información proveniente de la red de monitoreo del AMM indica que, históricamente, en esta zona metropolitana se rebasan las normas de calidad de aire establecidas por la Secretaría de Salud (SS) principalmente para las partículas aerodinámicas con diámetro menor a 10 micras (PM10), en menor medida para el ozono (O3) y para monóxido de carbono (CO) [2].

Una de las problemáticas derivadas del monitoreo de contaminantes es que no se identifican a tiempo las fuentes fijas que están sobrepasando los límites permitidos, además de predecir cuál es el comportamiento de los contaminantes en un periodo de tiempo.

Se cuenta con una red de estaciones de monitoreo supervisadas por el Sistema de Integral de Monitoreo Ambiental (SIMA) en el Área Metropolitana de Monterrey. Se dispone de una base de datos histórica con la información captada por las estaciones, esta base de datos está ya en un formato digital para ser procesada para explotación de información y toma de decisiones. El gran volumen de datos no permite realizar un análisis rápido y flexible por algoritmos tradicionales de búsqueda y optimización. El conjunto de datos no es fácil de manipular, debido a su tamaño y complejidad.

Se han identificado trabajos relacionados con la problemática como el de Aplicación de Técnica de Inteligencia Artificial a la Predicción de Contaminantes Atmosférico [3] propone un modelo de predicción de las concentraciones de los contaminantes SO₂ y PM10 para cada caseta de monitorización de la ciudad de Salamanca, México.

Los modelos propuestos utilizan información real de las concentraciones de los contaminantes y las variables meteorológicas obtenidas de la Red de Monitorización

Atmosférica de Salamanca (REDMAS). Cada modelo propuesto utiliza técnicas de inteligencia artificial y reconocimiento de patrones, que permitan utilizar la información obtenida por la REDMAS y las variables meteorológicas para predecir las concentraciones de los contaminantes que se han reportado como preocupantes en la región.

Otro trabajo es el de Modelado de partículas PM10 y PM2.5 mediante redes neuronales artificiales sobre clima tropical de San Francisco de Campeche [4], desarrolla una metodología computacional basada en redes neuronales con 3 capas. Esta arquitectura utiliza una base de datos tomando los siguientes campos: Días de la semana, tiempo del día, temperatura ambiente, presión atmosférica, velocidad del viento, dirección del viento, humedad relativa y radiación solar. La mejor red neuronal fue compuesta por 30 neuronas en una capa oculta usando el algoritmo de optimización de Levenberg-Marquardt logrando generar predicciones con un coeficiente de determinación de 93.01% para PM2.5 y 90.10% para PM10.

Otro trabajo es el de un algoritmo basado en Evolución Diferencial para resolver problemas multiobjetivo [5], se plantea un método de Evolución Diferencial para resolver problemas Multiobjetivo, en este tipo de problemas se busca optimizar dos o más funciones las cuales se interponen entre sí. El algoritmo propuesto por el autor se plantea de dos formas diferentes: El primero utiliza una malla adaptativa la cual tiene como fin almacenar las mejores soluciones posibles que son obtenidas durante la ejecución del algoritmo. En el segundo algoritmo se utiliza la dominancia. La investigación del autor L.V. Santana brinda definiciones claras relacionadas con la evolución diferencial y su aplicación, ayuda a entender mejor el funcionamiento de esta técnica de la computación evolutiva, pero no se le dio una aplicación para el monitoreo de contaminantes.

En el trabajo de predicción de contaminantes en el aire mediante evolución diferencial con Random Forest [6], se aborda el uso de dichas técnicas para la predicción, tomando los datos recabados de la ciudad de Delhi y Patna. Este último método es un algoritmo de clasificación que consta de muchos árboles de decisiones. Utiliza la aleatoriedad de características al construir cada árbol individual para intentar crear un bosque de árboles no correlacionado cuya predicción es más precisa que la de cualquier árbol individual.

En los trabajos relacionados no se encontró alguno que fuera desarrollado mediante algoritmos evolutivos bajo la técnica diferencial haciendo un análisis con partículas PM10 para el monitoreo de contaminantes. El encontrado Predicción de contaminantes en el aire mediante evolución diferencial de [6] hace un análisis general de los niveles de contaminantes en el aire, además de abordar una técnica combinando dos modelos de programación, como lo es Random Forest y Evolución diferencial.

Los trabajos plantean aplicaciones generalizadas de técnicas de inteligencia artificial para la predicción, la propuesta de investigación derivada de este artículo tiene su aplicación centrada en la predicción de niveles de contaminantes PM10.

2. DESARROLLO

2.1. Metodología de solución

Para el desarrollo del algoritmo se siguió un modelo basado en el desarrollo de algoritmo evolutivos y sus componentes: representación de la solución, función de calidad, población de soluciones, mecanismo de selección, operadores de variación, mecanismo de reemplazo, El algoritmo evolutivo diferencial está basado en la propuesta de R. Storn, K. Price en su reporte técnico Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces [7]. El lenguaje de programación usado para implementar el algoritmo fue Java.

2.2. Propuesta general del algoritmo evolutivo

En el siguiente diagrama se muestra la propuesta del algoritmo evolutivo desde una perspectiva global.

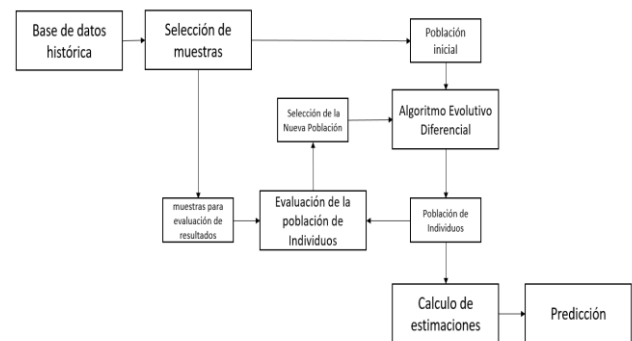


Figura 1. Diagrama general del algoritmo evolutivo propuesto

El algoritmo es alimentado por una base de datos histórica que provee el Sistema Nacional de Información de la Calidad del Aire (SINAICA) del contaminante PM10, datos recabados para la estación la Pastora, correspondiente al intervalo del 01/01/2020 al 30/07/2020 como referencia para la generación de predicciones. Dando un total de 4250 muestras del contaminante PM10. Cada una corresponde a la medición de una hora específica del día en el intervalo mencionado. Teniendo aproximadamente 177 muestras para cada hora de las 0:00-23:00.

Los datos son seleccionados de manera aleatoria siguiendo una distribución uniforme 60 de las 177 muestras mencionadas previamente, correspondientes a los niveles del contaminante de partículas suspendidas PM10, las muestras seleccionadas fueron divididas en 30 que serán la población inicial y 30 que serán utilizadas para evaluar a las posibles soluciones.

Este proceso de selección se aplicará para las 24 horas, por lo que en total fueron seleccionadas 1440 muestras, 720 usado como población inicial y 720 como evaluación de posibles soluciones.

Para representar la solución del algoritmo evolutivo mediante el método de Evolución Diferencial se maneja una población como un conjunto de vectores, donde cada vector es un "individuo" y cada individuo es una posible solución al problema. Por esto la solución se representa mediante un arreglo o ArrayList de objetos de la clase Individuo, esta clase contiene en si un arreglo correspondiente a los parámetros que definen los valores del individuo en sus distintas características, en el caso de la predicción de los contaminantes PM10 es el nivel del contaminante para dicho individuo.

La función de calidad es el medio sobre el cual se puede evaluar el desempeño de las posibles soluciones al problema, buscando normalmente minimizar o maximizar dicha función. En el caso del problema de los contaminantes, la función de calidad está dada por la comparación entre los valores del vector Individuo y valores reales registrados por la estación la pastora.

Para la evaluación del desempeño del algoritmo, se optó por calcular el error absoluto obtenido, teniendo el valor real como la medición obtenida para cierta hora en específico, y el valor aproximado como la estimación de salida que da el algoritmo. La fórmula a utilizar fue la siguiente:

$$e = \frac{|V_{Real} - V_{Aprox}|}{V_{Real}} \times 100\%$$

Donde:

e: es el error absoluto

Vreal: Es el valor registrado de la medición del contaminante PM10 en cierta hora específica.

Vaprox: Es el valor de salida del algoritmo evolutivo en cierta hora específica.

El valor absoluto de la resta entre el valor real y el valor aproximado será dividido entre el valor real y multiplicado por 100%, esto para obtener el porcentaje de error absoluto. Esta fórmula es de utilidad para evaluar la aproximación de los resultados del algoritmo con los valores reales que se esperarían obtener como salida.

El error absoluto fue utilizado para la medición del desempeño de todos los individuos, en el cual el Valor real era el promedio de las 30 mediciones seleccionadas.

En los operadores de de variación siguiendo lo establecido con el método de evolución diferencial el método de variación se basa en la generación de un vector de prueba V. Para cada vector $X_{i,G} = 0, 1, 2, \dots, NP - 1$, un vector de prueba V es generado de acuerdo con:

$$V = \underline{x}_{r_1,G} + F \cdot (\underline{x}_{r_2,G} - \underline{x}_{r_3,G})$$

con $r_1, r_2, r_3 \in [0, NP-1]$ y son mutuamente diferentes, $F > 0$.

Dónde:

V es el vector de prueba

$X_{r_1,G}$ Es un vector parámetro elegido aleatoriamente.

r_1, r_2, r_3 son elegidos aleatoriamente en el intervalo $[0, NP-1]$ y son diferentes al índice que se está ejecutando i .

F es un real y constante factor que controla la amplificación de la variación diferencial.

En el mecanismo de reemplazo para la selección de los sobrevivientes, se eligió hacer uso de un mecanismo a la selección de los más aptos, haciendo ciertas variaciones explicadas a continuación:

Los individuos de la población son sometidos a un ordenamiento dependiendo de sus resultados en base a la función de calidad explicada anteriormente, se ordenarán los 30 individuos correspondientes a la población junto con los 30 individuos producidos por los operadores de variación, siendo este conjunto de 60 individuos ordenados del mejor al peor.

Se ha definido una ecuación basada en el cálculo de la suma de los números de 1 a n, teniendo probabilidad diferente para cada individuo dependiendo de su posición en el ordenamiento de los individuos, retirando la sumatoria correspondiente a los individuos que ya han sido seleccionados. La probabilidad de que un individuo sea seleccionado está dada por:

$$P(A) = \frac{(n - p)}{\frac{(n)(n + 1)}{2} - \sum_{i=0}^x (n - p_i)}$$

Dónde:

n es el número total de individuos.

p es la posición en la que se encuentra en la tabla de individuos (desde 0 a n-1).

x es la cantidad de individuos que ya han sido seleccionados antes que él.

2.3. Algoritmo

Selección de la población inicial

```
for (int i = 0; i < TAM_POBLACION * 2; i++) {
    int eleg = rng.nextInt(cant);
    if (elegidos[eleg] == 0) {
        if (i % 2 == 0) {
            pobs[x] = eleg;
            x++;
        } else {
            test[y] = eleg;
            y++;
        }
        elegidos[eleg] = 1;
    } else {
        i--;
    }
}
```

Operación de cruce

```
private Individuo CalcularV(Individuo i, Individuo j, Individuo k) {
    Individuo v = new Individuo(CANT_VAR);
    for (int y = 0; y < 1; y++) {
        //Aplicacion de la formula V = R1 + F(R2-R3)
        double Valor = i.getVectPos(y) + (F * (j.getVectPos(y) - k.getVectPos(y)));
        //Se agrega el resultado al individuo V
        v.setVectPos(y, Valor);
    }
    //Retornamos el Individuo v
    return v;
}
```

Operadores de mutación

```
private Individuo ElegirU(Individuo xi, Individuo v) {
    Individuo u = new Individuo(CANT_VAR);
    int j = rng.nextInt(CANT_VAR);
    for (int i = 0; i < CANT_VAR; i++) {
        double r = rng.nextDouble();
        double elegido;
        //Verificamos que r sea mayor al Cruzamiento y que i sea diferente de j
        //Recordemos que j es el valor que si o si se tomara del Vector v
        if (r > Cr && i != j) {
            elegido = xi.getVectPos(i);
        } else {
            //Verificamos si el numero no excede los limites establecidos
            elegido = v.getVectPos(i);
            if (elegido < limites[i][0] || elegido > limites[i][1]) {
                //Si excede los limites se cambia por un valor aleatorio
                //entre los limites aceptados
                elegido = rng.nextDouble() * (limites[i][1] - limites[i][0]) + limites[i][0];
            }
        }
        u.setVectPos(i, elegido);
    }
    return u;
}
```

3. RESULTADOS

Con el fin de comprobar la efectividad del algoritmo, se realizaron pruebas en varios días elegidos teniendo como única condición contar con las mediciones de los niveles del contaminante PM10. Comparando los valores de salida del algoritmo con los valores reales capturados por la estación de monitoreo de contaminantes la Pastora a cargo del SIMA. Los días evaluados fueron: 2 de febrero del 2020, 27 de agosto del 2020, 27 de noviembre del 2020, 3 de diciembre del 2020.

Se presentan tres resultados de los experimentos realizados con el algoritmo.

Tabla 1 Comparación de los resultados obtenidos del 2 de febrero, niveles del contaminante PM10, la estimación dada por el algoritmo evolutivo y su margen de error absoluto

Comparación datos 2 de febrero 11:00 a 3 de febrero 0:00			
Hora	Estimaciones	Resultado esperado	% de error
11:00 - 12:00	75	79	5.06
12:00 - 13:00	70	86	18.60
13:00 - 14:00	63	67	5.97
14:00 - 15:00	57	59	3.39
15:00 - 16:00	60	61	1.64
16:00 - 17:00	58	42	38.10
17:00 - 18:00	53	36	47.22
18:00 - 19:00	54	59	8.47
19:00 - 20:00	56	68	17.65
20:00 - 21:00	52	77	32.47
21:00 - 22:00	52	82	36.59
22:00 - 23:00	54	72	25.00
Promedio de error			20.01

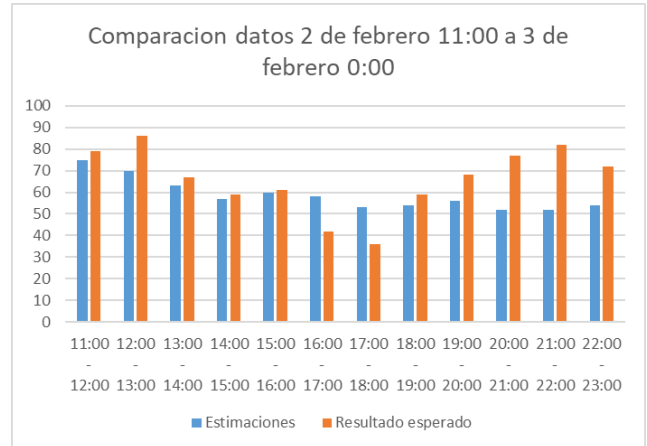


Figura 2. Comparación de los resultados del día 2 de febrero, estimaciones de los niveles de PM10 dadas por el algoritmo y valores reales captados para cada intervalo de tiempo

La estimación de un cambio más constante del algoritmo provoca que algunas veces se aleje de los valores reales, al presentar cambios más pequeños como se aprecia entre las 13:00 y las 15:00 horas, el algoritmo se acerca más a la solución correcta, las variaciones entre las 19:00 y 22:00 se presentan cambios más bruscos en el resultado esperado, el algoritmo en situaciones de horas pico hace un estimado aproximado.

Tabla 2. Resultados 2 de febrero utilizando un rango aproximado para los niveles del contaminante PM10 con un mínimo y máximo estimado.

Hora	Min	Max	Resultado esperado	Margen
1	74.00	77.00	79	2.53164557
2	68.00	74.00	86	13.95348837
3	60.00	69.00	67	0
4	53.00	65.00	59	0
5	55.00	70.00	61	0
6	52.00	70.00	42	23.80952381
7	46.00	67.00	36	27.77777778
8	46.00	70.00	59	0
9	47.00	74.00	68	0
10	42.00	72.00	77	6.493506494
11	41.00	74.00	82	9.756097561
12	42.00	78.00	72	0
Promedio de error			7.03	

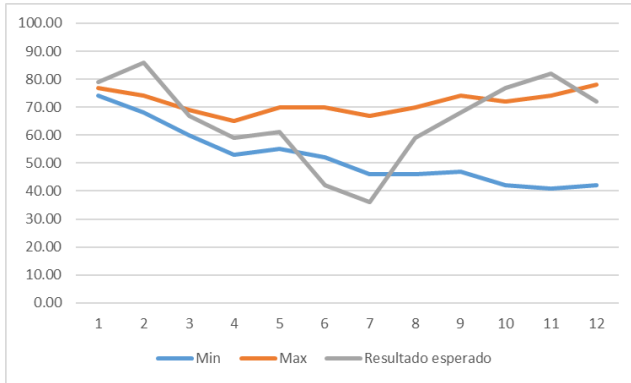


Figura 3. Comparación de los niveles del contaminante PM10 con los valores mínimo y máximo estimados, en el eje x están señaladas las horas, en el eje y la concentración del contaminante

Para este día, el cual presentó muchos cambios en los niveles del contaminante PM10, es difícil para el algoritmo abarcar estos posibles cambios en sus estimaciones por lo que aun con el margen de mínimo y máximo, algunos valores están fuera de lo que se esperaría entre los ascensos y descensos.

Para el día 28 de agosto de 2020, este fue el comportamiento del algoritmo en los valores mínimos, máximos y graficación del valor esperado.

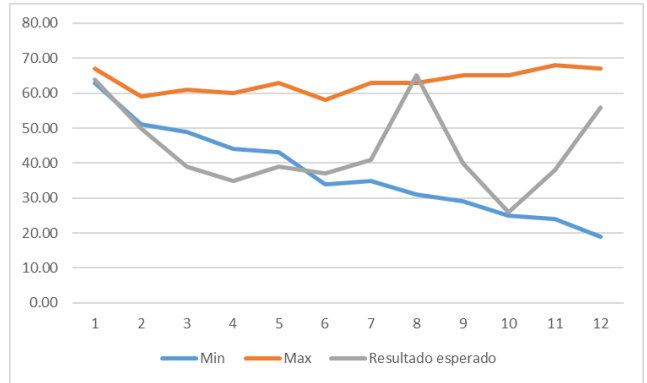
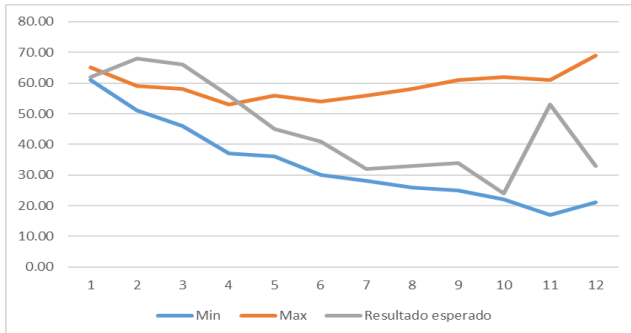


Figura 5. Comparación de los niveles del contaminante PM10 con los valores mínimo y máximo estimados, en el eje x están señaladas las horas, en el eje y la concentración del contaminante para el día 3 de diciembre de 2020

4. CONCLUSIÓN

Se cuenta con un algoritmo evolutivo codificado desde sus diferentes etapas, realizando predicciones de la concentración de los niveles del contaminante PM10 dentro de un espacio de soluciones reales tomadas por una estación de monitoreo del Sistema Integral de Monitoreo Ambiental de Nuevo León.

Analizando los resultados obtenidos por el algoritmo se puede apreciar un porcentaje de error absoluto entre 20% y 30% al compararse con el valor exacto de salida con respecto a los niveles del contaminante PM10. Esto ocurre debido a que la predicción de salida del algoritmo es un valor medio entre los valores estimados para cierta hora específica.

Al buscar un punto medio en donde podrían estar los niveles de los contaminantes, es necesario un segundo sistema el cual sea capaz de acercar este valor lo más posible al valor real estimado.

Aplicando el margen de ± 2 por cada hora transcurrida es posible abarcar los cambios drásticos en los niveles de contaminantes que ocurren durante el día, sin embargo, el margen es cada vez más grande y por lo tanto menos preciso. Por lo que sería más óptimo aplicar algún otro algoritmo extra para dicha función el cual pudiese predecir cuándo ocurrirán estos cambios drásticos, sin embargo, para esto podría ser necesario un modelo de predicción meteorológico el cual este enfocado en el movimiento y dispersión de contaminantes en el aire.

El desarrollo del algoritmo evolutivo diferencial está basado en la propuesta de R. Storn, K. Price en su reporte técnico Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces [7]. Para la implementación de este algoritmo, se llevó a cabo un análisis del comportamiento de los niveles del contaminante PM10 en un intervalo de tiempo determinado, estos datos fueron recabados por la estación La Pastora de SIMA en el municipio de Guadalupe, Nuevo León México, en un periodo de tiempo correspondiente del 1 de enero del 2020 al 31 de diciembre del 2020. Los

En este caso, con el margen propuesto es posible abarcar los incrementos dados en ocho de las doce horas estimadas, por lo que el resultado es satisfactorio. Las necesidades de agrandar el margen conforme pasan las horas ayuda a los cambios drásticos en los niveles de PM10 a los cuales, el algoritmo no es capaz de abarcar por sí solo. El espacio de solución para el algoritmo está dentro del rango esperado.

Para el día 3 de diciembre de 2020, este fue el comportamiento del algoritmo en los valores mínimos, máximos y graficación del valor esperado.

resultados obtenidos fueron utilizados para determinar el funcionamiento implementado en el algoritmo para la realización de las estimaciones finales.

El análisis y predicción de los niveles de contaminantes es muy amplio, es posible enfocar el análisis en distintos contaminantes criterio que no fueron tomados en cuenta para este algoritmo, como PM2.5, análisis de los niveles del contaminante en ciertas horas específicas, este fenómeno también puede ser investigado para encontrar los motivos que provocan dicho cambio en combinación con la parte de medición de las variables meteorológicas.

El algoritmo pudiera ser combinado con otras técnicas de inteligencia artificial como redes neuronales o lógica difusa para la selección de las entradas del algoritmo evolutivo y tener un espacio de soluciones más exacto.

Referencias

- [1] Sun Z, Zhu D (2019). Exposure to outdoor air pollution and its human health outcomes: A scoping review. *plos one* 14(5): e0216550 [en línea], disponible: <https://doi.org/10.1371/journal.pone.0216550>, sitio visitado Junio 14,2021.
- [2] INECC. Estudio de emisiones y actividad vehicular en el área metropolitana de Monterrey, N.L [en línea], disponible: https://www.gob.mx/cms/uploads/attachment/file/112367/2010_cgsa_rsd_Monterrey.pdf, sitio visitado Junio 21,2021
- [3] Cortina Januchs, M. G. Aplicación de técnicas de inteligencia artificial a la predicción de contaminantes atmosféricos (Doctoral dissertation, Telecomunicacion) [en línea], disponible: http://oa.upm.es/11206/1/maria_guadalupe_cortina_januchs.pdf, sitio visitado Mayo 7,202.
- [4] Espinosa Guzmán, A.A. May Tzuc, O. Pantí Balam, I. Reyes Trujeque, J. Pérez Quintana, I. V. & Bassam, A. Modelado de partículas pm10 y pm2.5 mediante redes neuronales artificiales sobre clima tropical de san francisco de Campeche, México. *Química Nova*, Vol. 40 No. 9, Septiembre 2017,1025-1034.
- [5] Santana, L. V. Un Algoritmo Basado en Evolución Diferencial para Resolver Problemas Multiobjetivo [en línea], disponible: <https://www.cs.cinvestav.mx/TesisGraduados/2004/tesisLuisVSantana.pdf>, sitio visitado Mayo 29,2021.
- [6] Rubal. Air pollution prediction via Differential evolution strategies with random forest method. *IRJET*, Vol. 04, Issue 1, Octubre 2017, 1083-1090.
- [7] Storn, R., & Price, K. Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, Vol. 11 No. 4, Diciembre 1997, 341-359.