



# Selección óptima de observadores de Koopman aplicados a DMDC en la obtención de gemelos digitales

*Amanda Martí Coll, Adrián Rodríguez Ramos, Orestes Llanes-Santiago*

## RESUMEN / ABSTRACT

El Internet Industrial de las Cosas, los Datos de Gran Envergadura, la Computación en la Nube y otras tecnologías permitieron el surgimiento y desarrollo de la iniciativa Industria 4.0. En este marco, los gemelos digitales se han vuelto muy populares y se pueden encontrar en todo tipo de industrias y procesos. Una parte importante en el desarrollo de un gemelo digital es la obtención del modelo del sistema, tarea para la cual existe gran cantidad de métodos. El operador de Koopman y la Descomposición en Modos Dinámicos con Control han demostrado sus capacidades en este sentido y se va extendiendo su uso dentro de la comunidad científica. En este trabajo se presenta la propuesta de un procedimiento para automatizar la selección del conjunto de observadores de Koopman resolviendo un problema de optimización. Como resultado de la aplicación del procedimiento se obtiene la cantidad mínima de observadores que minimizan el error entre las observaciones reales de un proceso y las obtenidas a partir del modelo lineal estimado para el desarrollo de su gemelo digital. Para resolver el problema de optimización se puede usar cualquiera de los muchos algoritmos de optimización que aparecen en la literatura. En este trabajo se utilizó un algoritmo genético.

Palabras claves: Descomposición en Modos Dinámicos, teoría de Koopman, gemelos digitales.

*The Industrial Internet of Things, Big Data, Cloud Computing, and other technologies have enabled the rise and development of the Industry 4.0 initiative. In this context, digital twins have become very popular and can be found in all kinds of industries and processes. An important part of developing a digital twin is obtaining the system model, a task for which there are many methods. The Koopman operator and Dynamic Mode Decomposition with Control have demonstrated their capabilities in this context and its use is spreading within the scientific community. This work presents a proposal for a procedure to automatize the selection of the set of Koopman observers by solving an optimization problem. As a result of applying the procedure, the minimum number of observers that minimize the error between the real observations of a process and those obtained from the estimated linear model for the development of its digital twin is obtained. To solve the optimization problem, any of the many optimization algorithms found in the literature can be used. In this work, a genetic algorithm was applied.*

**Keywords:** *Dynamic Mode Decomposition, Koopman theory, digital twins.*

*Optimal selection of Koopman observers applied to DMDC in the development of digital twins*

## 1. -INTRODUCCIÓN

La Industria 4.0 es un paradigma que ha revolucionado la forma en que las empresas fabrican, mejoran y distribuyen sus productos. Algunas de las tecnologías que sirven como pilar para el desarrollo de la Industria 4.0 son el Internet Industrial de las Cosas (IIoT, por sus siglas en inglés), la Computación en la Nube, el Big Data, la Inteligencia Artificial y el Aprendizaje Automático. El objetivo en este caso es mejorar la productividad y la eficiencia mediante el uso de herramientas de monitoreo

de condición [1-3]; la flexibilidad y la seguridad [4]; a la vez que se brinda la posibilidad de tomar decisiones y una personalización más inteligente en las operaciones de fabricación y de la cadena de suministro.

Los gemelos digitales son una tecnología clave en la Industria 4.0, por lo que ha habido cierto auge en su estudio [5-9]. Un gemelo digital, según [10], consiste en una representación de un objeto físico o proceso en un ambiente virtual. Los mismos se utilizan para modelar y simular procesos, sistemas y productos en tiempo real. En la Industria 4.0, los gemelos digitales se utilizan en una amplia gama de aplicaciones, desde el diseño hasta la planificación de la producción o la optimización de la cadena de suministro. Estos permiten la simulación y predicción del comportamiento de un objeto o proceso físico ante distintas circunstancias, lo que facilita la monitorización y gestión remotas, apoyo a labores de soporte y mantenimiento, la prevención de tiempos de inactividad, así como la personalización de la producción a los requerimientos del cliente.

Los gemelos digitales representan procesos dinámicos con una determinada relación entre sus variables de entrada y de salida. El contar con un modelo matemático que describa sus diferentes dinámicas es crucial para poder analizar, simular, predecir o controlar su comportamiento. Sin embargo, la obtención de este último no es algo trivial. En este sentido existen dos grandes grupos de métodos: los analíticos y los basados en datos. Los primeros se apoyan en el conocimiento previo de las leyes físicas que rigen el sistema y utilizan técnicas matemáticas para obtener un modelo que se ajuste a los datos observados. Los segundos, por el contrario, requieren ninguno o poco conocimiento a priori del sistema, y se apoyan en técnicas de aprendizaje automático o inteligencia artificial para extraer la información relevante de los datos y construir un modelo empírico. La identificación de sistemas, por su parte, puede verse en un punto medio, puesto que se puede usar el método analítico para obtener una estructura inicial del modelo y luego ajustar esta estructura con la información que se recoja a partir de datos experimentales.

En el contexto de la industria 4.0, donde los sistemas son cada vez más complejos y no lineales; la obtención de modelos a partir de los datos del proceso es una opción a tener en cuenta [11-13]. Por un lado, es flexible a los cambios del sistema; pues se puede actualizar el modelo automáticamente a partir de nuevos datos. Por otro lado, es eficiente, al permitir manejar grandes volúmenes de datos provenientes de múltiples sensores. Además de que son más rápidas y económicas, en el sentido de que la identificación ocurre al aplicar un algoritmo y no necesita la formulación de hipótesis o supuestos complicados sobre el comportamiento del proceso.

El operador de Koopman es una herramienta matemática que permite analizar sistemas dinámicos no lineales mediante una transformación lineal en un espacio de funciones (observadores). La teoría de Koopman plantea que es posible representar un sistema dinámico no lineal en términos de un operador lineal de dimensión infinita [14]. Algunos de sus campos de uso son: la robótica [15], el automovilismo [16] y el control de sistemas [17], entre otros.

Debido a que el operador de Koopman es infinito-dimensional, se han desarrollado técnicas que logran una representación reducida con la que es viable el desarrollo de aplicaciones prácticas. El análisis a partir de la Descomposición en Modos Dinámicos (DMD), está conectado con el operador de Koopman en este sentido [14]. Esto se debe a que DMD es una técnica que utiliza datos del proceso para obtener una aproximación lineal de las dinámicas de un sistema [18]. Existe una generalización de este algoritmo, llamada Descomposición en Modos Dinámicos con Control (DMD) que permite incluir en el análisis las señales de control [19]. Existe una segunda generalización, conocida como Descomposición en Modos Dinámicos Extendida (EDMD) propuesta por [20]. En este caso se busca encontrar el conjunto de autofunciones, autovectores y autovalores del operador de Koopman a partir de una aproximación finita del mismo; permitiendo con esto hallar la respuesta en el tiempo del sistema. Esto ha permitido ampliar grandemente el área de aplicación de esta técnica sin implicar un aumento apreciable en la complejidad del algoritmo. Los algoritmos basados en DMD han sido utilizados en diferentes tipos de aplicaciones: estudio de fluidos [21], estimación de parámetros de una aeronave [18], análisis de la temperatura en edificios inteligentes [22], entre otros.

Aunque se ha avanzado significativamente en la conexión de DMD con sistemas no lineales, la adecuada elección de los observadores de Koopman sigue siendo un reto. Esto se debe a que la mayoría de los trabajos se enfocan en diferentes formas de encontrar una aproximación del operador de Koopman y de acuerdo a esto seleccionan los observadores a partir de lo que pudiéramos identificar como conocimiento de experto [23-26]. Sin embargo, esta investigación no sigue la tendencia de la mayoría de los investigadores asociada al uso de DMD y EDMD para la selección de los observadores [27-30]. Alternativamente, se propone la aplicación de la teoría de Koopman asociada directamente con DMDc. De esta forma se tendrán las ventajas de analizar sistemas dinámicos no lineales y obtener las matrices A y B del modelo lineal en el espacio de estados; permitiendo con ello el empleo de diferentes tipos de técnicas de la literatura clásica que son ampliamente usadas para el análisis de sistemas y el diseño de controladores como en los enfoques presentados en [31] y [32] por solo citar dos ejemplos.

El objetivo fundamental de este trabajo y su principal contribución científica es la propuesta de un procedimiento para la selección del conjunto de observadores de Koopman resolviendo un problema de optimización. Como resultado de la aplicación del procedimiento se obtiene la cantidad mínima de observadores que minimizan el error entre las observaciones

reales de un proceso y las obtenidas a partir del modelo lineal estimado para el desarrollo de un gemelo digital. Para resolver el problema de optimización se puede usar cualquiera de los muchos algoritmos de optimización que aparecen en la literatura. En este trabajo se utilizó un Algoritmo Genético (AG).

Para evaluar la efectividad del algoritmo propuesto se aplicó el mismo en la obtención del modelo de un Tanque de Reactor Continuamente Agitado (CSTR por sus siglas en inglés) obteniéndose resultados muy satisfactorios [33].

El trabajo está organizado de la siguiente forma: en la sección 2 se describen los elementos fundamentales de las herramientas utilizadas y el algoritmo propuesto; en la sección 3 se presenta el modelo del CSTR, el diseño del experimento para la prueba del algoritmo y el análisis de los resultados obtenidos. Finalmente se exponen las conclusiones.

## 2.- MATERIALES Y MÉTODOS

### 2.1.- DESCOMPOSICIÓN EN MODOS DINÁMICOS CON CONTROL

La descomposición en modos dinámicos es una técnica de análisis de datos que se utiliza para identificar sistemas dinámicos a partir de datos experimentales. La misma permite utilizar datos muestreados de elevada dimensionalidad de los procesos para extraer estructuras coherentes que contengan las características dominantes del comportamiento dinámico de estos últimos [34]. Estas estructuras, denominadas modos dinámicos, corresponden fuertemente a los autovectores del espacio de estados que gobiernan el sistema dinámico subyacente. Una vez obtenidos los modos dinámicos, se pueden utilizar para construir un modelo matemático lineal que describa el comportamiento dinámico del proceso bajo estudio.

La extensión de DMD que permite tener en cuenta señales de entrada (o de control) se denomina DMDc. La aplicación de este método necesita de tres elementos básicos:

- $p$  observaciones o mediciones del sistema objeto de estudio en diferentes instantes igualmente espaciados en el tiempo:  $\mathbf{x}_j$ ;
- las señales de entrada al proceso en los mismos instantes que el vector de observaciones:  $\mathbf{u}_j$ ;
- las observaciones del sistema avanzadas un instante de tiempo:  $\mathbf{x}_{j+1}$ .

La idea es lograr identificar las matrices A y B que describen la dinámica del sistema según la siguiente relación:

$$\mathbf{x}_{j+1} = A\mathbf{x}_j + B\mathbf{u}_j \quad (1.a)$$

$$\mathbf{y}_{j+1} = C\mathbf{x}_{j+1} \quad (1.b)$$

Donde  $A \in R^{n \times n}$ ,  $B \in R^{n \times m}$  y  $C \in R^{s \times n}$  siendo  $n$  la dimensión del sistema,  $m$  la cantidad de entradas y  $s$  la cantidad de salidas. Por otro lado,  $\mathbf{y}_{j+1}$  es el vector que representa las salidas del proceso.

Los datos del sistema pueden ser agrupados en las siguientes matrices:

$$X_a = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_{p-1}] \quad (2)$$

$$U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_{p-1}] \quad (3)$$

$$X_b = [\mathbf{x}_2 \ \mathbf{x}_3 \ \dots \ \mathbf{x}_p] \quad (4)$$

de esta forma la Ecuación 1.a puede reescribirse como:

$$X_b \approx AX_a + BU \quad (5)$$

En su caso más genérico DMDc se centra en encontrar la aproximación lineal que mejor ajusta a las matrices A y B. De forma resumida los pasos seguir son los siguientes [14]:

- a. Crear la matriz de las entradas:

$$\Gamma = \begin{bmatrix} X_a \\ U \end{bmatrix} \quad (6)$$

- b. Realizar la descomposición en valores singulares de la matriz  $\Gamma$ :

$$\Gamma \approx U_{svd} \Sigma_{svd} V_{svd}^* = \begin{bmatrix} U_{svd_1} \\ U_{svd_2} \end{bmatrix} \Sigma_{svd} V_{svd}^* \quad (7)$$

donde \* indica la traspuesta compleja conjugada de la matriz

- c. Obtener las matrices A y B del sistema lineal:

$$[A \ B] \approx X_b \begin{bmatrix} X_a \\ U \end{bmatrix}^\dagger = [X_b V_{svd} \Sigma_{svd}^{-1} U_{svd_1}^* \quad X_b V_{svd} \Sigma_{svd}^{-1} U_{svd_2}^*] \quad (8)$$

donde  $\dagger$  es la pseudo-inversa Moore-Penrose.

La obtención de este modelo lineal en el espacio de estados permite aplicar diferentes tipos de técnicas de la literatura clásica que son ampliamente usadas para el análisis de sistemas y el diseño de controladores.

## 2.2.- EL OPERADOR DE KOOPMAN

El operador de Koopman es otra herramienta de análisis que se utiliza para estudiar las propiedades de un sistema dinámico. Básicamente se reduce a que todo sistema no lineal puede ser aproximado a un sistema infinito-dimensional lineal [14]. Por supuesto, en la práctica es imposible trabajar con un sistema de dimensión infinita, por tanto, es necesario encontrar una aproximación de dicho operador.

Partiendo de que la evolución dinámica de un sistema discreto está dada por:

$$\mathbf{x}_{j+1} = \mathbf{F}(\mathbf{x}_j) \quad (9)$$

la teoría de Koopman puede representarse de forma simplificada con la siguiente relación:

$$K\mathbf{g}(\mathbf{x}_j) = \mathbf{g}(\mathbf{F}(\mathbf{x}_j)) = \mathbf{g}(\mathbf{x}_{j+1}) \quad (10)$$

Donde  $K$  es la aproximación del operador y  $\mathbf{g}$  es un vector de funciones, también conocidos como observadores de los estados del sistema. La introducción de estos denominados observadores, a partir de un diccionario o biblioteca de funciones ( $D$ ), es la que va a permitir captar aquellas no linealidades presentes en las dinámicas para lograr la aproximación lineal.

Para obtener el operador de Koopman en una dimensión finita basta seguir los siguientes pasos [14]:

- a. De las matrices  $X_a$  y  $X_b$  formar  $Y_a$  y  $Y_b$  respectivamente:

$$Y_a = [\mathbf{g}(\mathbf{x}_1) \ \mathbf{g}(\mathbf{x}_2) \ \dots \ \mathbf{g}(\mathbf{x}_{p-1})] \quad (11)$$

$$Y_b = [\mathbf{g}(\mathbf{x}_2) \ \mathbf{g}(\mathbf{x}_3) \ \dots \ \mathbf{g}(\mathbf{x}_p)] \quad (12)$$

- b. Aplicar DMD para obtener la aproximación del operador:

$$A_{Koopman} = Y_b Y_a^\dagger \quad (13)$$

Si se analizan con detenimiento las Ecuaciones (8) y (13) se aprecia que existe una estrecha relación entre DMDc y el operador de Koopman. Esto se debe a que la teoría detrás de DMD puede ser considerada una aproximación numérica del análisis espectral de Koopman, lo que permite su aplicación a sistemas dinámicos no lineales [35].

## 2.3.- ALGORITMO GENÉTICO

Los Algoritmos Genéticos son algoritmos de optimización estocástica inspirados en los principios de la evolución y la selección natural [36-38]. De forma resumida, un AG se encarga de evolucionar una población compuesta por diferentes individuos hacia el individuo o solución óptima [39].

El procedimiento elemental seguido por un AG mono-objetivo se representa esquemáticamente en la Fig. 1. El mismo comienza con la denominada población inicial, que consiste en un conjunto de  $N$  individuos que suelen ser generados aleatoriamente. A cada individuo se le calcula su valor de adaptación haciendo uso de la función objetivo. Luego se investiga si alguno de los individuos cumple con el requisito de optimización buscado; por ejemplo, el error mínimo permitido. En caso de que al menos uno de los individuos cumpla con este criterio de parada, se selecciona al mejor individuo como resultado final del algoritmo. En caso de no ser así, se pasa a crear una nueva generación (también de tamaño  $N$ ), la cual constituirá la nueva generación. Para esto se aplican mecanismos de selección, cruce, mutación y, en algunos casos, elitismo. Una vez creada la nueva generación, esta pasa a ser la nueva población y el algoritmo vuelve al punto de evaluar la función objetivo. Este proceso se repetirá una y otra vez hasta que se logre el criterio de parada antes mencionado, o hasta que se cumpla una cantidad predefinida de iteraciones.

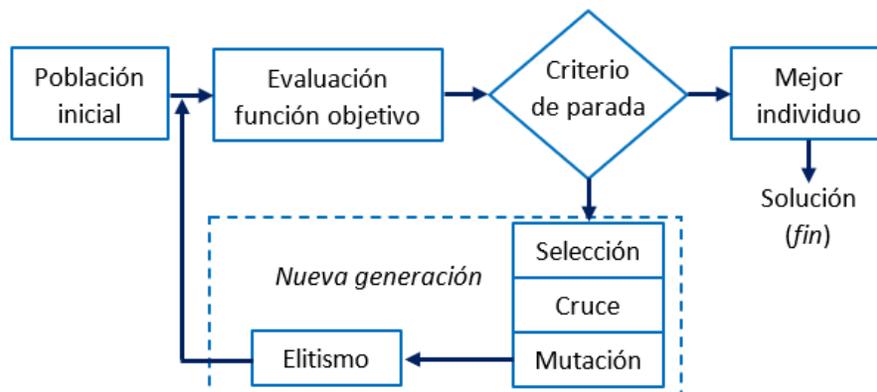


Figura 1

Procedimiento básico de un AG mono-objetivo (modificado de [40]).

## 2.4.- METODOLOGÍA PARA LA SELECCIÓN ÓPTIMA DE OBSERVADORES

### 2.4.1.- ALGUNAS CONSIDERACIONES

Como se ha mencionado anteriormente, la idea es contar con un algoritmo que, sin tener conocimiento previo del sistema y usando datos del mismo, permita la selección automática del menor conjunto de observadores de Koopman que minimizan el error entre el proceso real y el modelo lineal que se obtiene con la aplicación de DMDc. Es importante resaltar que: i) estos observadores serán escogidos a partir un diccionario de funciones de observación  $D$  previamente construido y ii) que el diccionario debe ser lo suficientemente amplio para abarcar diferentes tipos de no linealidades. Partiendo de lo anterior, se plantea un problema de optimización que utilizará como herramienta un algoritmo genético mono-objetivo cuyas características fundamentales se presentan a continuación:

- Cada población en el AG estará compuesta por  $z$  individuos, los cuales consisten en arreglos de etiquetas. Cada etiqueta será un número entero y representará la posición de un observador en el diccionario definido. Por ejemplo: dado el

diccionario  $D = \{obs_1, obs_2, \dots, obs_q\}$ , un individuo compuesto por los dos primeros observadores tendría la siguiente codificación: [1,2].

- Como no se tiene conocimiento previo de la cantidad de observadores que serán necesarios, el AG será aplicado para diferentes tamaños de individuo ( $l$ ). Dado un diccionario de  $q$  observadores, el intervalo de tamaños de individuo permitido queda en el rango:  $[1 \leq l \leq q]$ .
- Si bien el resultado de DMDc son las matrices A y B de un modelo con dimensión mayor a la del sistema original (debido a la aplicación de la teoría de Koopman); las salidas finales del modelo serán las estimaciones de las salidas de la Ecuación 1.b:  $\hat{\mathbf{y}}_j = [\hat{y}_{j_1}, \hat{y}_{j_2}, \dots, \hat{y}_{j_s}]$ .
- La *función objetivo* para el proceso de optimización será la media del error entre las muestras originales  $\mathbf{y}_j$  y las salidas del modelo estimado  $\hat{\mathbf{y}}_j$ . El cual se define como:

$$error = \frac{1}{h} \sum_{j=1}^h \sqrt{\sum_{i=1}^s (y_{ji} - \hat{y}_{ji})^2} \quad (11)$$

donde:

- $h$  es el número total de muestras.
- $s$  es el número total de salidas del sistema.
- $y_{ji}$  es el valor real del  $i$ -ésimo elemento de salida de la  $j$ -ésima muestra.
- $\hat{y}_{ji}$  es el valor estimado del  $i$ -ésimo elemento de salida de la  $j$ -ésima muestra.

Al hacer uso de la distancia se garantiza que, sin importar la dimensionalidad del sistema, siempre se podrá emplear un AG mono-objetivo. En este caso se usó la distancia euclídea.

- Se tendrán dos *criterios de parada*:
  - a. El primero es el error máximo permitido, que se asocia al valor de adaptación calculado por la función objetivo.
  - b. El segundo es que no existan más posibles combinaciones a evaluar por el AG.
- Los mecanismos para crear la nueva generación en el AG son: selección por torneo binario y mutación puntual. Además, también se empleará elitismo para asegurar que las mejores soluciones sean las que pasen a la siguiente población.
- El operador genético de cruce no fue empleado, pues sólo se podrían intercambiar los últimos observadores de cada combinación y, por tanto, los hijos resultantes serían iguales a sus padres.

## 2.4.2.- ALGORITMO PROPUESTO

El algoritmo empieza evaluando cada observador del diccionario de funciones para escoger al mejor individuo (aquel que produce el menor valor de la función objetivo - error). Además de este individuo, también se guardarán aquellos que devuelvan un valor de adaptación hasta un 10% mayor que el menor obtenido, constituyendo con esto el conjunto de mejores individuos. De esta forma se evita caer en un mínimo local, pues se da la oportunidad de probar otras combinaciones de funciones.

A continuación, corresponde buscar las mejores combinaciones de dos observadores. Para ello se crea un conjunto de posibles observadores que en esta etapa inicial estará formado por todos los observadores del diccionario. Con esto se permite probar cada uno de estos últimos con los mejores individuos ya seleccionados de un observador. El AG será el encargado de lograr la combinación de observadores que minimiza el error, devolviendo un nuevo conjunto de mejores individuos, donde el tamaño de cada elemento se habrá incrementado en uno respecto a sus predecesores. El elemento final de cada individuo de este último conjunto será guardado para formar la nueva lista de posibles observadores, pues son las funciones que el AG detectó que sí contribuyen a una disminución significativa del valor de adaptación.

El AG será aplicado repetidamente en cada actualización de mejores individuos y posibles observadores hasta que ya no queden combinaciones por analizar. Además, hay que resaltar que cuando se terminen de probar las combinaciones para cada individuo en mejores individuos, se analizará si algún resultado devolvió un error estrictamente menor que el mejor individuo para así actualizar el que será la solución final del método que es el conjunto menor de observadores con el que se obtiene el menor error.

La Fig. 2 muestra un diagrama con la lógica propuesta.

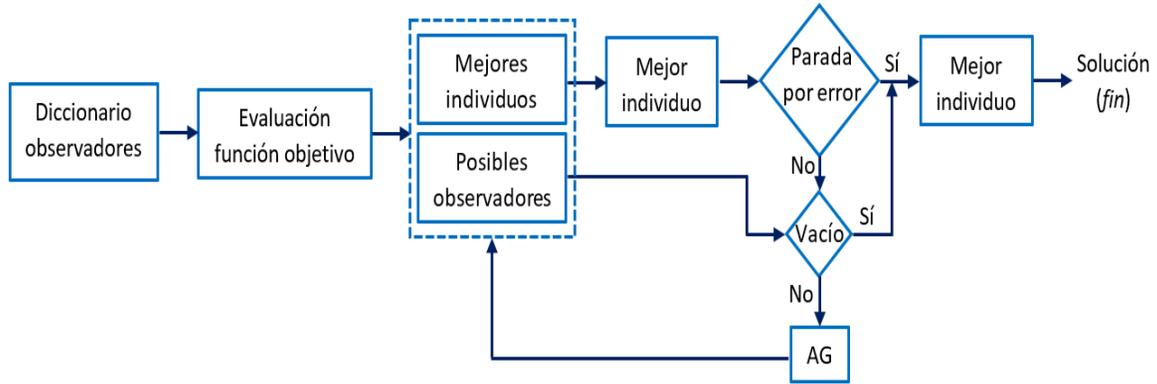


Figura 2

Diagrama en bloques del algoritmo propuesto.

### 3.- DISEÑO DE EXPERIMENTOS Y ANÁLISIS DE LOS RESULTADOS

#### 3.1.- TANQUE REACTOR CONTINUAMENTE AGITADO

Un proceso que ha sido bastante estudiado en la literatura es el caso del CSTR. El modelo, recogido en [33] se presenta a continuación:

$$\begin{aligned} \dot{C}_a(t) &= \frac{q}{v} (C_{a0} - C_a(t)) - k_0 C_a(t) e^{-\frac{E}{RT(t)}} \\ \dot{T}(t) &= \frac{q}{v} (T_0 - T(t)) - k_1 C_a(t) e^{-\frac{E}{RT(t)}} + k_2 q_c(t) \left( 1 - e^{-\frac{k_3}{q_c(t)}} \right) (T_{c0} - T(t)) \end{aligned} \quad (12)$$

Las variables de mayor interés son:

- $x_1 = C_a(t)$ : concentración del componente A (mol/l) – variable de estado 1.
- $x_2 = T(t)$ : temperatura de la mezcla (K) – variable de estado 2.
- $u = q_c(t)$ : flujo del refrigerante (l/min) – señal de control.

La base de datos utilizada cuenta con un total de 7500 muestras, tomadas a intervalos de 6 segundos.

#### 3.2.- DISEÑO DE EXPERIMENTOS

El objetivo de este experimento es obtener, de forma automática, el menor conjunto de observadores de Koopman que aplicados a DMDC producen el mejor modelo lineal del proceso estudiado.

Teniendo esto en cuenta, se definió el diccionario de observadores formado por las 10 funciones que aparece en la Fig. 3. Este diccionario fue construido específicamente para el presente experimento, pero la variedad y cantidad de no linealidades puede variar según decidan los especialistas que usen el presente procedimiento. Por otra parte, se debe mencionar que de las 7500 muestras presentes en la base de datos se tomaron 6000 para la estimación del modelo y el resto para la validación. El criterio de parada por error fue establecido en 0.05. Para asegurar la validez del experimento desde el punto de vista estadístico el mismo se repitió 50 veces.

```
@(x1,x2) x1.^2
@(x1,x2) x1.*x2
@(x1,x2) x2.^2
@(x1,x2) exp(-1./x2)
@(x1,x2) exp(x1.*x2)
@(x1,x2) tanh(x1)./cos(x2)
@(x1,x2) cosh(x2)./tanh(x1)
@(x1,x2) x1 + rand
@(x1,x2) log(x2) + 5*rand
@(x1,x2) exp(-x2)./x2
```

Figura 3

Diccionario de funciones usado.

El experimento se realizó en una computadora con las siguientes características: procesador Intel(R) Core(TM) i7-7500U CPU @ 2.70GHz 2.90 GHz; 8GB de RAM y Sistema Operativo “Windows 10 Pro for Workstations” versión 22H2. El software utilizado fue MATLAB R2019b y no se empleó ninguna librería externa ni especializada del propio software.

### 3.3.- ANÁLISIS DE LOS RESULTADOS

Luego de realizar los experimentos se tuvo que la solución final es el individuo formado por los 4 primeros observadores del diccionario (Fig. 4). Un aspecto a señalar es que en ninguno de los casos se logró el valor de error deseado y que, por tanto, cada parada del AG fue por haberse agotado la cantidad de combinaciones de observadores.

```
El mejor individuo es:
    3    4    1    2

Con valor de adaptación:
    0.1304
```

Figura 4

Mejor individuo (conjunto óptimo de observadores).

El tiempo medio de ejecución de los experimentos fue de 0.22 segundos. Como se aprecia, este tiempo medio de ejecución no resulta elevado, pues es un experimento que no se repite continuamente y que no interviene en el funcionamiento normal del proceso bajo estudio. Ello implica que el sistema físico no se verá afectado en ninguna medida por estas demoras. No obstante, hay que tener en cuenta que aspectos como la cantidad de muestras del proceso y el tamaño del diccionario de funciones tiene una influencia directa en estos tiempos.

La Fig. 5 muestra las matrices A y B de la aproximación lineal del sistema, así como la matriz de los autovalores ( $\Sigma$ ), cuando se aplicó el algoritmo DMDC al individuo final.

$$A = \begin{bmatrix} 7.19 & -0.03 & -0.24 & -0.02 & 2.75 \times 10^{-5} & 6.38 \\ -1.23 \times 10^3 & 4.17 & 25.96 & 3.15 & -3.34 \times 10^{-3} & -7.51 \times 10^2 \\ 7.18 & 0.02 & -0.67 & -0.02 & -1.88 \times 10^{-5} & -3.98 \\ 1.93 \times 10^3 & -13.39 & 1.01 \times 10^2 & -4.29 & 0.01 & 3.22 \times 10^3 \\ -1.07 \times 10^6 & 2.68 \times 10^3 & 7.76 \times 10^3 & 2.73 \times 10^3 & -1.81 & -6.38 \times 10^5 \\ -6.71 \times 10^{-3} & 1.73 \times 10^{-5} & 3.32 \times 10^{-4} & 1.69 \times 10^{-5} & -1.83 \times 10^{-8} & 1.00 \end{bmatrix}$$

$$B = \begin{bmatrix} 2.12 \times 10^{-4} \\ -0.12 \\ 3.18 \times 10^{-5} \\ 0.08 \\ -1.06 \times 10^2 \\ -6.15 \times 10^{-7} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} 1.51 \times 10^7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5.68 \times 10^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 27.96 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.012 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.88 \times 10^{-3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 8.64 \times 10^{-5} \end{bmatrix}$$

Figura 5

Matrices resultantes para los cuatro observadores seleccionados.

Por otro lado, en la Fig. 6 se aprecia la distancia euclídea (error) para cada muestra de validación. Además, la Fig. 7 y la Fig. 8 presentan los resultados de la simulación para el conjunto de datos de validación; donde se aprecia que el modelo final resulta adecuado para el sistema bajo estudio. La primera figura presenta la comparación de los datos reales con los simulados a partir de las matrices A y B. La segunda figura muestra los errores en el tiempo de cada variable de estado.

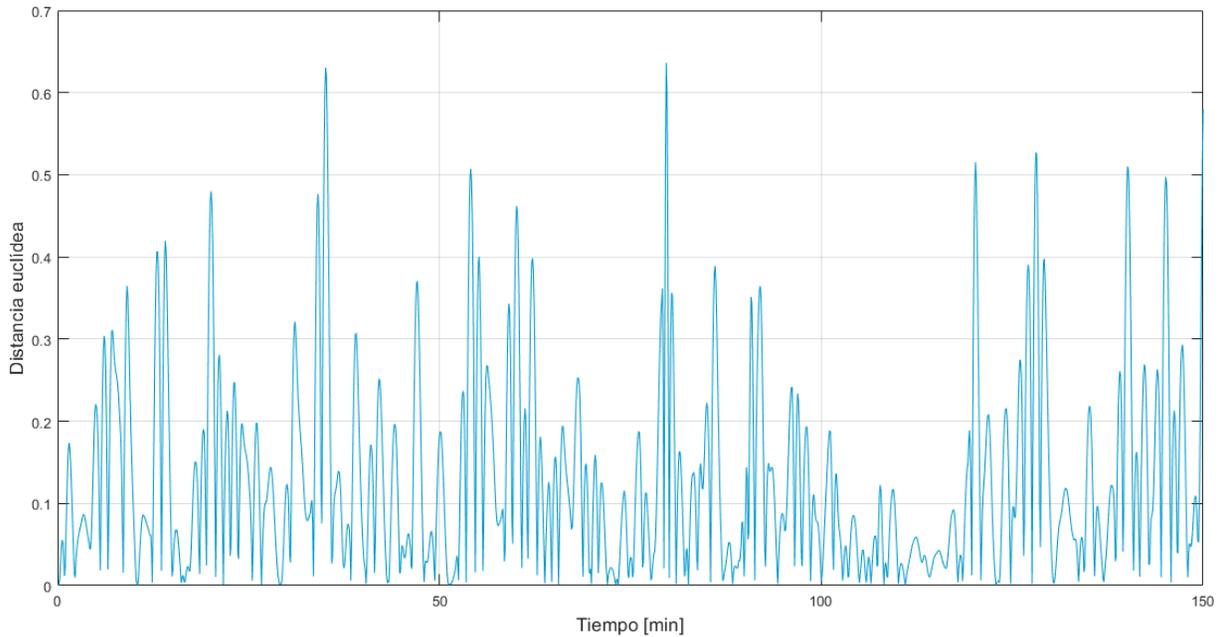
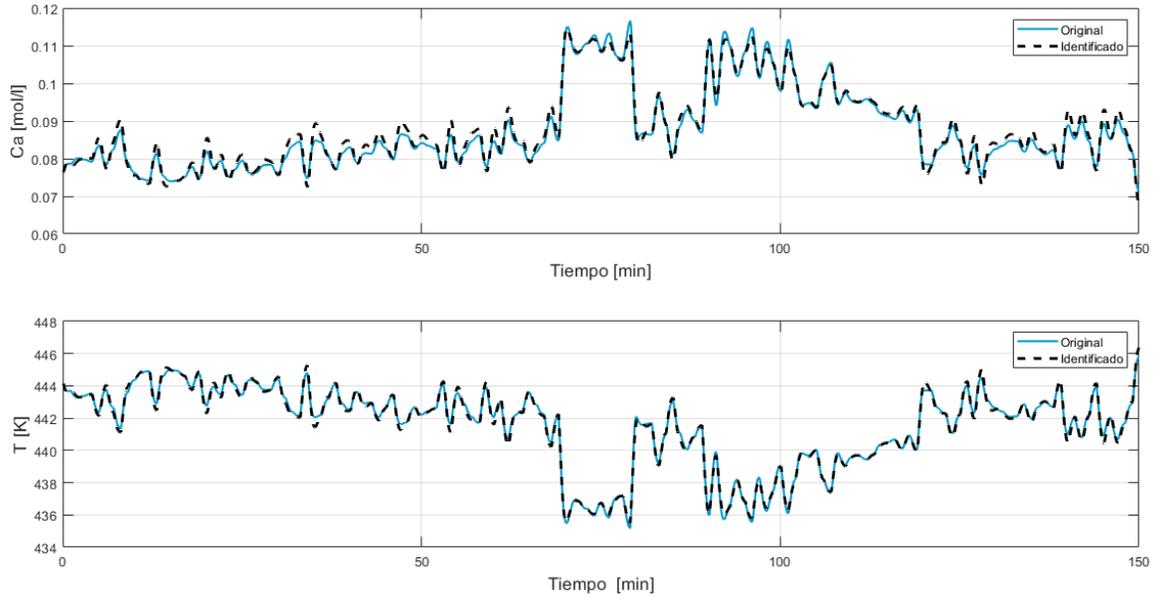


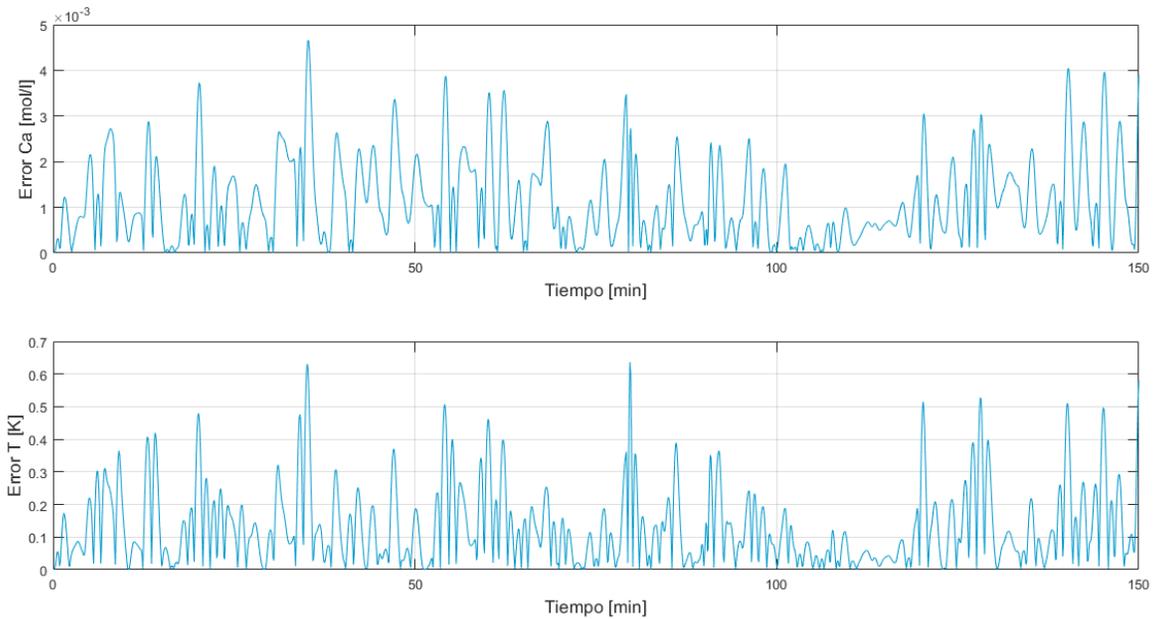
Figura 6

Error (distancia euclídea) entre los pares  $(y_1, y_2)$  y  $(\hat{y}_1, \hat{y}_2)$  para cada observación del conjunto de validación.



**Figura 7**

**Comparación entre la respuesta del sistema simulado y las muestras originales.**



**Figura 8**

**Residuos del sistema simulado.**

## 4.- CONCLUSIONES

En este trabajo se propuso y se demostró la validez de un algoritmo para la selección óptima de observadores de Koopman aplicados a DMDC en la obtención de gemelos digitales. Al usar DMDC se pudieron obtener, a partir de las observaciones del proceso, las matrices A y B del sistema lineal discreto en el espacio de estados, brindando así la posibilidad de aplicar

diferentes técnicas de análisis sobre el gemelo digital. Por otro lado, con base en la teoría de Koopman, se utilizaron observadores para lograr identificar las dinámicas no lineales del proceso. De este último resalta que la selección adecuada de dichos observadores es un paso crítico que afecta directamente la eficacia del método en la descripción y análisis de los sistemas dinámicos no lineales. Por tanto, con la propuesta de algoritmo a partir del planteamiento de un problema de optimización basado en AG, se garantiza que se prueben las mejores combinaciones de observadores para el sistema bajo estudio. Así, el usuario del método podrá prescindir de la necesidad de conocimiento de experto sobre las posibles dinámicas que tienen lugar en el sistema. Sin embargo, un importante aspecto a tener en cuenta, es que se debe contar con un diccionario de observadores lo suficientemente amplio para abarcar diferentes tipos de no linealidades.

Otro factor al que hay que prestarle atención es la cantidad y la calidad de las muestras utilizadas, pues las mismas deben cubrir un rango adecuado de condiciones de operación, garantizando con esto que el algoritmo pueda detectar las dinámicas fundamentales del proceso.

## REFERENCIAS

1. Camps Echevarría L., Llanes Santiago O., da Silva Neto A.J. An approach for Fault Diagnosis based on bio-inspired strategies. In: IEEE Congress on Evolutionary Computation, Barcelona, Spain; 2010:1–7.
2. Ares Milián M., Quinones Grueiro M., Verde C., Llanes Santiago O. A Leak Zone Location Approach in Water Distribution Networks Combining Data-Driven and Model-Based Methods. *Water*. 2021;13(20):2924.
3. Quinones Grueiro M., Prieto Moreno A., Llanes Santiago O. Modeling and Monitoring for Transitions Based on Local Kernel Density Estimation and Process Pattern Construction. *Industrial & Engineering Chemistry Research*. 201;55(3):692-702.
4. Quiñones Grueiro M., Prieto Moreno A., Verde C., Llanes Santiago O. Decision Support System for Cyber Attack Diagnosis in Smart Water Networks. *IFAC-PapersOnLine*. 2019;51(34):329–334. 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018.
5. Knebel F., Trevisan R., Nascimento G., Abel M., Wickboldt J. A study on cloud and edge computing for the implementation of digital twins in the Oil & Gas industries. *Computers & Industrial Engineering*. 2023;182:109363.
6. Stadtmann F., Rasheed A., Kvamsdal T., Johannessen K.A., San O., Kölle K., et al. Digital Twins in Wind Energy: Emerging Technologies and Industry-Informed Future Directions. *IEEE Access*. 2023;11:110762–110795.
7. Tepljakov A. Intelligent Control and Digital Twins for Industry 4.0. *Sensors*. 2023;23(8):4036.
8. Han B., Habibi M.A., Richerzhagen B., Schindhelm K., Zeiger F., Lamberti F., et al. Digital Twins for Industry 4.0 in the 6G Era. *IEEE Open Journal of Vehicular Technology*. 2023;4:820–835.
9. Muthuswamy, P., Shunmugesh, K. Artificial intelligence-based tool condition monitoring for digital twins and industry 4.0 applications. *International Journal on Interactive Design and Manufacturing (IJIDeM)*. 2023;17:1067–1087.
10. Semeraro C., Aljaghoub H., Abdelkareem M.A., Alami A.H., Olabi A.G. Digital twin in battery energy storage systems: Trends and gaps detection through association rule mining. *Energy*. 2023;273:127086.
11. Liu Q., Zhang C., Yang C., Chen Z., Ding J. Dynamic Inner Canonical Variate Network for Incipient Fault Monitoring. *IEEE Transactions on Industrial Informatics*. 2024;20(9):10922-10932.
12. Sui Q., Wang Y., Chen Z., Liu C., Xu D., Cai W. Deep Koopman Kalman Filter for Nonlinear Model-free Industrial Process Data Denoising and Its Soft Sensing Application. *IEEE Sensors Journal*. 2024.
13. Yan A., Jingcheng G. A Recurrent Stochastic Configuration Network for Dynamic System Modeling and Its Application. *International Journal on Artificial Intelligence Tools*. 2024.
14. Kutz J.N., Brunton S.L., Brunton B.W., Proctor J.L. *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems*. Philadelphia: SIAM; 2016.
15. Shi L., Liu Z., Karydis K. Koopman Operators for Modeling and Control of Soft Robotics. *Current Robotics Reports*. 2023;4(2):23–31.
16. Manzoor W.A., Rawashdeh S., Mohammadi A. Vehicular Applications of Koopman Operator Theory—A Survey. *IEEE Access*. 2023;11:25917–25931.
17. Rahmani M., Redkar S. Data-driven Koopman fractional order PID control of a MEMS gyroscope using bat algorithm. *Neural Computing and Applications*. 2023;35:9831–9840.
18. Swaminathan B., Manathara J.G., Vinayagam A.K. Application of Dynamic Mode Decomposition with Control (DMDC) for aircraft parameter estimation. *IFAC-PapersOnLine*. 2022;55(1):789–794. 7th International Conference on Advances in Control and Optimization of Dynamical Systems ACODS 2022.

19. Proctor J.L., Brunton S.L., Kutz J.N. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*. 2016;15(1):142–161.
20. Williams M.O., Kevrekidis I.G., Rowley C.W. A data-driven approximation of the Koopman operator: extending dynamic mode decomposition. *Journal of Nonlinear Science*. 2015;25(6):1307–1346.
21. Bai Z., Kaiser E., Proctor J.L., Kutz J.N., Brunton S.L. Dynamic Mode Decomposition for Compressive System Identification. *AIAA Journal*. 2020;58(2):561–574.
22. Sunny K., Sheikh A., Wagh S. Application of Dynamic Mode Decomposition for Temperature Analysis in Smart Building. In: 2020 7th International Conference on Control, Decision and Information Technologies (CoDIT). 2020;1:1197–1202.
23. Colbrook M.J., Li Q., Raut R.V., Townsend A. Beyond expectations: residual dynamic mode decomposition and variance for stochastic dynamical systems. *Nonlinear Dynamics*. 2024;112(3):2037–2061.
24. Garmaev S., Fink O. Deep Koopman Operator-based degradation modelling. *Reliability Engineering & System Safety*. 2024;251:110351.
25. Nozawa I., Kamienski E., O’Neill C., Asada H.H. A Monte Carlo Approach to Koopman Direct Encoding and Its Application to the Learning of Neural-Network Observables. *IEEE Robotics and Automation Letters*. 2024;9(3):2264–2271.
26. Philipp F.M., Schaller M., Worthmann K., Peitz S., Nüske F. Error bounds for kernel-based approximations of the Koopman operator. *Applied and Computational Harmonic Analysis*. 2024;71:101657.
27. Ouala S., Chapron B., Collard F., Gaultier L., Fablet R. Extending the extended dynamic mode decomposition with latent observables: the latent EDMD framework. *Machine Learning: Science and Technology*. 2023;4(2):025018.
28. Bandtlow O.F., Just W., Slipantschuk J. EDMD for expanding circle maps and their complex perturbations. *Applied and Computational Harmonic Analysis*. 2024;73:101690.
29. Hasebe T., Uchiyama K., Masuda K. Multilayer Extended Dynamic Mode Decomposition for Coupled Van der Pol Oscillators. In: 2024 28th International Conference on Methods and Models in Automation and Robotics (MMAR); 2024: 352–357.
30. Kar A.K., Akhter M.M., Mishra A.C., Mohanty S.K. EDMD: An Entropy based Dissimilarity measure to cluster Mixed-categorical Data. *Pattern Recognition*. 2024;155:110674.
31. Sira Ramirez H., Llanes Santiago O. Sliding Mode Control of Nonlinear Mechanical Vibrations. *Journal of Dynamic Systems Measurement and Control*. 2000;122(4):674–678.
32. Sira Ramirez H., Llanes Santiago O., Fernández N. On the stabilization of nonlinear systems via input-dependent sliding surfaces. *International Journal of Robust and Nonlinear Control*. 1996;6:771–780.
33. Espinosa J.J., Vandewalle J. Predictive Control Using Fuzzy Models. In: Roy R, Furuhashi T, Chawdhry PK, editors. *Advances in Soft Computing*. London: Springer London; 1999:187–200.
34. Schmid P.J. Dynamic Mode Decomposition and Its Variants. *Annual Review of Fluid Mechanics*. 2021;54:225–254.
35. Nedzhibov G.H. Dynamic Mode Decomposition: A New Approach for Computing the DMD Modes and Eigenvalues. *Annals of the Academy of Romanian Scientists Series on Mathematics and its Applications*. 2022;14(1-2):5–16.
36. Patil R.S., Jadhav S.P., Patil M.D. Review of Intelligent and Nature-Inspired Algorithms-Based Methods for Tuning PID Controllers in Industrial Applications. *Journal of Robotics and Control (JRC)*. 2024;5(2):336–350.
37. Peña García R., Velázquez Sánchez R.D., Daza Argumedo C.G., Escobedo Alva J.O., Tapia Herrera R., Meda Campaña J.A. Physics-Based Aircraft Dynamics Identification Using Genetic Algorithms. *Aerospace*. 2024;11(2):142.
38. Guediri M., Ikhlef N., Bouchehou H., Guediri A., Guediri A. Optimization by Genetic Algorithm of a Wind Energy System applied to a Dualfeed Generator. *Engineering, Technology & Applied Science Research*. 2024;14(5):16890–16896.
39. Panska Z., Balsky M., Bayer R., Brejcha M. Unconventional Luminaire Layout Design by Genetic Algorithms. *Sustainability*. 2024;16(17):7689.
40. Guerrero R.J.A., Camacho D.J.G. Optimización de procesos de producción mediante el uso de algoritmos genéticos. *Revista Ingeniería*. 2023;7(18):316–324.

## CONFLICTO DE INTERESES

No existe conflicto de intereses entre los autores, ni con ninguna institución a la que cada uno está afiliado, ni con ninguna otra institución. Las opiniones expresadas aquí son únicamente responsabilidad de los autores y no representan la posición de la Institución o las instituciones a las que están afiliados.

## CONTRIBUCIONES DE LOS AUTORES

**Amanda Martí Coll:** Curación de datos, Análisis formal, Investigación, Metodología, Software, Validación -Verificación, Visualización y Redacción- borrador original.

**Adrián Rodríguez Ramos:** Conceptualización, Recursos, Supervisión y Redacción – revisión y edición.

**Orestes Llanes Santiago:** Conceptualización, Adquisición de fondos, Administración de proyecto, Recursos, Supervisión y Redacción – revisión y edición.

## AUTORES

**Amanda Martí Coll**, Ingeniera en Automática. Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE), La Habana, Cuba. Email: [abamanda.mc@automatica.cujae.edu.cu](mailto:abamanda.mc@automatica.cujae.edu.cu). No. ORCID: 0000-0002-3046-2447. Sus intereses de investigación se orientan hacia la minería de datos y el control de procesos industriales. Actualmente se enfoca en el estudio de los gemelos digitales aplicados a la industria farmacéutica.

**Adrián Rodríguez Ramos**, Ingeniero en Automática. Doctor en Ciencias Técnicas. Profesor Titular de la Universidad Tecnológica de La Habana José Antonio Echeverría (CUJAE). Email: [adrian.rr@automatica.cujae.edu.cu](mailto:adrian.rr@automatica.cujae.edu.cu). No ORCID: 0000-0002-0240-7491. Sus principales intereses de investigación están en el monitoreo de condición de condición y diagnóstico de fallos en procesos industriales usando herramientas de inteligencia computacional.

**Orestes Llanes-Santiago**, Ingeniero Electricista. Doctor en Ciencias Aplicadas. Profesor e Investigador Titular de la Universidad Tecnológica José Antonio Echeverría (CUJAE). Académico Titular de la Academia de Ciencias de Cuba. Email: [orestes@tesla.cujae.edu.cu](mailto:orestes@tesla.cujae.edu.cu) No. ORCID 0000-0002-6864-9629. Sus principales intereses de investigación están en el monitoreo de condición y diagnóstico de fallos en sistemas industriales, implementación del paradigma Industria 4.0, inteligencia computacional aplicada al control y el control de sistemas no lineales.



Esta revista se publica bajo una [Licencia Creative Commons Atribución-No Comercial-Sin Derivar 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/)