



INSTITUTO TECNOLÓGICO DE ORIZABA

DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN

Tesis Doctoral

Presenta:

Andrés Morán Durán

Para obtener el grado de:

Doctor en Ciencias de la Ingeniería

Director de tesis:

Dr. Albino Martínez Sibaja

*Control y diagnóstico de fallas basado en datos
para sistemas PEMFC*





“2021: Año de la Independencia”

Orizaba, Veracruz, 03/12/2021
Dependencia: División de Estudios de
Posgrado e Investigación
Asunto: Autorización de Impresión
OPCION: I

C. ANDRÉS MORÁN DURÁN
Candidato a Grado de Doctor en:
CIENCIAS DE LA INGENIERÍA
P R E S E N T E.-

De acuerdo con el Reglamento de Titulación vigente de los Centros de Enseñanza Técnica Superior, dependiente de la Dirección General de Institutos Tecnológicos de la Secretaría de Educación Pública y habiendo cumplido con todas las indicaciones que la Comisión Revisora le hizo respecto a su Trabajo Profesional titulado:

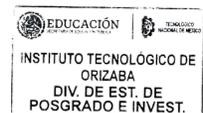
“CONTROL Y DIAGNÓSTICO DE FALLAS BASADO EN DATOS PARA SISTEMAS DE CELDAS DE COMBUSTIBLE PEM”

Comunico a Usted que este Departamento concede su autorización para que proceda a la impresión del mismo.

ATENTAMENTE

Excelencia en Educación Tecnológica®
CIENCIA - TÉCNICA - CULTURA®

DR. MARIO LEONCIO ARRIJOJA RODRÍGUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN



Avenida Oriente 9 No. 852
Col. Emiliano Zapata, C.P. 94320
Orizaba, Veracruz, México.
Teléfono: 272-110-53-60
Email: depi_orizaba@tecnm.mx
www.orizaba.tecnm.mx





Instituto Tecnológico de Orizaba
División de Estudios de Posgrado e Investigación

Orizaba, Veracruz, Mayo 05, de 2021.
Asunto: **Revisión de trabajo escrito**

C. MARIO LEONCIO ARRIJOJA RODRÍGUEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS
DE POSGRADO E INVESTIGACIÓN
P R E S E N T E . -

Los que suscriben, miembros del jurado, han realizado la revisión de la Tesis del (la) C.

ANDRÉS MORÁN DURÁN

La cual lleva el título de:

"CONTROL Y DIAGNÓSTICO DE FALLAS BASADO EN DATOS PARA SISTEMAS DE CELDAS DE COMBUSTIBLE PEM"

Y concluyen que se acepta.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
CIENCIA - TÉCNICA - CULTURA®

PRESIDENTE: DR. ALBINO MARTÍNEZ SIBAJA

FIRMA

SECRETARIO: DR. RUBÉN POSADA GÓMEZ

FIRMA

VOCAL: DR. OSCAR OSVALDO SANDOVAL GONZÁLEZ

FIRMA

VOCAL: DR. ALEJANDRO ALVARADO LASSMAN

FIRMA

VOCAL: DR. ALBERTO ALFONSO AGUILAR LASSERRE

FIRMA

VOCAL SUP.: DR. MANUEL ADAM MEDINA

FIRMA

FTI-TI-013



Avenida Oriente 9 No. 852
Col. Emiliano Zapata, C.P. 94320
Orizaba, Veracruz, México.
Teléfono: 272-110-53-60
Email: depi_orizaba@tecnm.mx
www.orizaba.tecnm.mx



Contenido

Introducción General	1
1.1 Panorama global.....	1
1.2 Consumo de energía, economía del hidrógeno y celdas de combustible	1
1.3 Contribución del trabajo.....	3
1.4 Objetivos	3
1.5 Cronograma	4
1.6 Estructura de la tesis	5
Generalidades sobre el control y diagnóstico de sistemas PEMFC.....	6
2.1 Introducción a las Celdas de Combustible	6
2.2 Principios de Operación de Celdas de Combustible	7
2.3 Sistema PEMFC	10
2.3.1 Modos de falla en PEMFC.....	11
2.3.2 Consideración sobre las fallas en PEMFC.....	11
2.4 Estado del arte sobre el control de sistemas de celda de combustible PEM basado en datos	13
2.5 Enfoques para el diagnóstico de fallas para sistemas PEMFC.....	15
2.5.1 Introducción al diagnóstico de fallas.....	15
2.5.2 Métodos de diagnóstico basados en modelos	18
2.5.3 Métodos de diagnóstico basados en datos.....	21
2.6 Desafíos del diagnóstico de fallas de sistemas PEMFC en línea	26
2.7 Algoritmos evolutivos	27
2.7.1 Contexto histórico	27
2.7.2 Programación genética	28
2.7.3 Clasificación por conjunto de reglas de programación genética	33
2.7.4 Algoritmo M3GP	36
2.7.5 Diagnostico y aislamiento de fallas utilizando Programación Genética	39
2.7.6 Software de Programación Genética	40
Control neuronal de voltaje de celda de combustible PEM basado en la regulación de la presión de hidrógeno.....	41
3.1 Materiales y métodos.....	41

3.1.1 Configuración del experimento.....	41
3.1.2 Algoritmos de selección de características y modelos basados en datos para celdas de combustible	42
3.2 Selección de características de celda de combustible	43
3.3 Modelos orientados al control basados en datos para celdas de combustible PEM...	46
3.3.1 Modelado de celdas de combustible utilizando algoritmos de regresión de aprendizaje automático	46
3.3.2. Modelado de celdas de combustible basado en Redes Neuronales.....	47
3.4 Esquema de control híbrido.....	49
3.5 Conclusiones parciales	52
Diagnóstico de fallas para sistemas de celdas de combustible PEM mediante programación genética.....	53
4.1 Modelo Dinámico PEMFC	53
4.1.1 Voltaje de salida.....	53
4.1.2 Transferencia de calor.....	54
4.2 Estructura del sistema PEMFC a diagnosticar	56
4.3 Descripción de fallas	56
4.4 Análisis de fallas.....	59
4.5 Diagnostico mediante arboles de falla utilizando el algoritmo M3GP.....	64
4.5.1 Parámetros del algoritmo M3GP.....	¡Error! Marcador no definido.
4.6 Comparación de métodos de diagnostico.....	67
4.6.1 Reporte de clasificación	68
4.6.2. Método LDA-NNA para el diagnóstico de fallas	68
4.6.3. Método PCA-NNA para el diagnóstico de fallas.....	70
4.6.3. Comparación de desempeño	71
4.6.4. Conclusiones parciales.....	73
Productos Académicos	74
Referencias	79

Introducción General

1.1 Panorama global

Transformar la producción y consumo de energía no sostenible a una economía verde que tenga como base el uso de energía renovable, es uno de los desafíos más importantes de para los gobiernos de todo el mundo.

Necesitamos reducir nuestra dependencia del petróleo y otros combustibles fósiles poniendo fin a los subsidios para las compañías petroleras y apoyando la generación de energía limpia que genera empleos y fortalece nuestra independencia energética.

Sin embargo, esta transformación no solo requiere avances tecnológicos en materia de producción de energía, también requiere cambios en la cultura de consumo de la sociedad. Algunos hábitos de consumo son irracionales, por ejemplo ¿Por qué se transportan productos alrededor del mundo cuando otros similares están disponibles localmente? Si bien este consumo genera desarrollo económico, también genera contaminación con un impacto social mayor que el generado por la venta de esos productos.

Muchos países desarrollados han avanzado en la reducción del consumo de energía, sin embargo, este avance aun no es suficiente. Se requieren avances tecnológicos para poder superar las barreras comerciales y poder utilizar las energías renovables como fuente de energía primaria. Una de las principales limitantes es que cada país cuenta con diferentes fuentes de energía renovables, así como disponibilidad limitada.

Los avances tecnológicos, por lo tanto, deben proporcionar soluciones a los siguientes 3 problemas:

1. Conversión de energía eficiente
2. Transporte seguro y a bajo costo
3. Almacenamiento de energía

“Es hora de una política energética sostenible que priorice a los consumidores, el medio ambiente, la salud humana y la paz.”

– Dennis Kucinich

1.2 Consumo de energía, economía del hidrógeno y celdas de combustible

Los datos históricos y las proyecciones del consumo de energía mundial desde 1990 hasta 2040 se muestran en [1] y se pueden observar en la **Figura 1**. Es alarmante las proyecciones del consumo de energía de los países que no pertenecen a la OCDE (organización para la cooperación y el desarrollo económico). Si este consumo no se controla, las consecuencias ambientales y económicas podrían afectar la calidad de vida de las personas. También se muestra en [1], que las fuentes de energía renovable tienen una tendencia al alza, sin embargo, a pesar de la implementación de políticas públicas para la reducción de la producción de energía a partir de combustibles fósiles, la energía producida a partir del carbón aun presenta un ligero crecimiento, vea la **Figura 2**.

Entre estas fuentes de energía renovable, el hidrógeno podría ser el combustible limpio del futuro, sus principales características son:

- Contiene el mayor contenido de energía por unidad de peso (142 kJ g^{-1}) [2]
- Es un combustible libre de carbón, ya que el producto de su combustión es agua [3]
- Puede ser usado como combustible directo o como un portador de energía para celdas de combustible [3].

Figura 1 Consumo mundial de energía, 1990–2040 (cuatrillones de Btu)

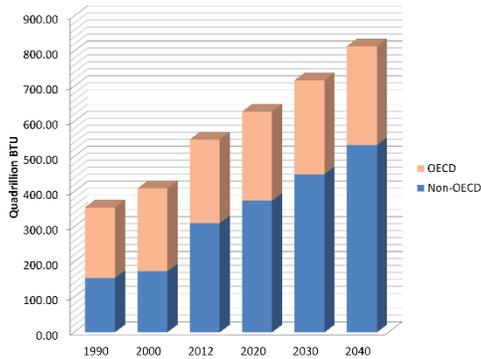
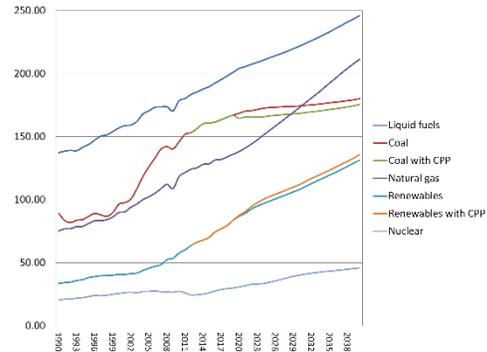


Figura 2 Consumo mundial de energía por fuente de energía.



Hoy en día, el hidrógeno compite con la red eléctrica para aplicaciones estacionarias y estaciones de servicio de gasolina para aplicaciones móviles (vehículos). Actualmente, existen diferentes métodos para producir hidrógeno, tanto procesos renovables como no renovables. Sin embargo, es necesario evaluar estos métodos no solo por su eficiencia energética o costo de producción, sino también por su impacto ambiental. De lo contrario, cometeremos los mismos errores que han causado tanto daño al medio ambiente. Es importante tener en cuenta que no existe una fuente de energía renovable que pueda sustituir, por sí sola, a los combustibles fósiles. La principal barrera en esta transición es el alto costo de producción inicial y que, en este momento, no existe la infraestructura necesaria para transportar, almacenar y dispensar el hidrógeno hasta el usuario final.

"Una de las tecnologías más prometedoras de conversión de energía del hidrógeno es la celda de combustible" [4]. Sin embargo, el costo y la confiabilidad, aun son barreras que tienen que ser superadas para que su comercialización sea posible. Por ejemplo, la vida útil actual de las celdas de combustible en aplicaciones de transporte es de 2,500 horas, la cual aún está muy lejos del objetivo de 5000 horas propuesto por U.S. Department of Energy (DOE) [5].

Para mejorar la confiabilidad y durabilidad de los sistemas PEMFC se necesitan dos estrategias operacionales. La primera enfocada en el control de celda, la cual debe tener como objetivo el uso eficiente del hidrógeno; y la segunda enfocada en la detección y aislamiento de las fallas tanto internas como de los sistemas auxiliares.

1.3 Contribución del trabajo

Las estrategias de control y diagnóstico de fallas se pueden clasificar en dos categorías: basado en modelos y basado en datos. Los dos tipos de enfoques poseen ventajas y desventajas. El método de control y diagnóstico debe seleccionarse de acuerdo con las características específicas del sistema. Respecto a las celdas de combustible PEM, los modelos matemáticos generalizados que describan con precisión el comportamiento del sistema y orientados al diagnóstico suelen ser difíciles de obtener. En este caso, los métodos basados en datos parecen ser más prácticos.

El objetivo de esta tesis es desarrollar estrategias basadas en datos para el control y diagnóstico de fallas para sistemas PEMFC, que mejoren el consumo de combustible y diagnostiquen fallas en tiempo real buscando prolongar la vida útil de la celda.

1.4 Objetivos

Objetivo general

Desarrollar un Control Tolerante a Fallas para monitorear y controlar el desempeño de la celda de combustible PEM

Objetivos específicos

- Determinar las variables de estado necesarias para modelar y controlar un sistema de celdas de combustible PEM (100%)
- Desarrollar un modelo que permita simular el desempeño de la celda (100%)
- Diseñar un sistema de control que mejore el consumo de combustible (100%)
- Diseñar un árbol de falla que sirva para diagnosticar y aislar fallas (100%)
- Comparar el desempeño del árbol de falla propuesto con los métodos de diagnóstico disponibles en la literatura (100%)

1.5 Cronograma

Planificador de proyectos

Seleccione un periodo para resaltarlo a la derecha. A continuación hay una leyenda que describe el gráfico.

Periodo resaltado: 8



ACTIVIDAD	INICIO DEL PLAN	DURACIÓN DEL PLAN	INICIO REAL	DURACIÓN REAL	PORCENTAJE COMPLETADO	PERIODOS								
						1	2	3	4	5	6	7	8	
Revisión del estado del arte	1	8	1	8	100%									
Investigar los metodos para la selección y extracción de características disponibles en la literatura	2	1	2	2	100%									
Crear la base de datos para realizar la experimentacion (Control)	2	1	2	2	100%									
Ejecutar los metodos de selección para determinar las variables necerarias modelar y controlar el voltaje	2	1	2	2	100%									
Defina las entradas del sistema del subconjunto formado por el algoritmo de selección de características	3	1	3	1	100%									
Probar diferentes algoritmos de regresión para predecir la variable de salida (voltaje)	3	1	3	1	100%									
Desarrollar el modelo inverso de la celda de combustible para controlar la señal de control	4	1	4	2	100%									
Integrar el modelo inverso con un Neurocontrol-PID	4	1	4	2	100%									
Escribir un artículo de los resultados del controlador propuesto	5	1	5	1	100%									
Analizar los metodos de diagnostico de fallas basados en datos aplicados a las celdas de combustible	6	1	6	1	100%									
En base a la actividad anterior determinar el algoritmo de clasificacion a utilizar para el diagnostico del sistema PEMFC	6	1	6	1	100%									
Implementar el algoritmo seleccionado en Matlab o Python	6	1	6	1	100%									
Crear la base de datos para realizar el diagnostico de fallas	7	1	7	1	100%									
Diseñar un árbol de falla que sirva para diagnosticar y aislar fallas en un sistema PEMFC	7	1	7	1	100%									
Comparar el desempeño del árbol de falla propuesto con los métodos de diagnóstico disponibles en la literatura	7	1	7	1	100%									
Escribir un artículo de los resultados del diagnostico de fallas	8	1	7	2	100%									
Redaccion de tesis	8	1	7	2	90%									

1.6 Estructura de la tesis

En la sección “Generalidades sobre el control y diagnóstico de sistemas PEMFC” se describe la teoría sobre la cual se fundamenta este trabajo. Se describen los distintos tipos de celdas de combustible y el principio de operación de la celda de membrana de intercambio de protones (PEM). También se mencionan los modos de fallas en los sistemas de celdas de combustible de membrana de intercambio de protones (PEMFC). Por último, en esta sección, se hace una revisión del estado del arte sobre los métodos de control y diagnóstico de fallas aplicados a sistemas PEMFC, tanto basados en modelos matemáticos como basados en datos.

En la sección “Control neuronal de voltaje de celda de combustible PEM basado en la regulación de la presión de hidrógeno” se presenta una metodología para modelar y controlar celdas de combustible PEM. Se compara el desempeño de varios métodos de selección de característica con el método “Análisis de componentes principales” para la selección de las variables de estado críticas para el proceso de modelado y control. Posteriormente se compara el desempeño de varios algoritmos de regresión para determinar cual tiene la mejor precisión. Por último, se propone un esquema de control híbrido, el cual conjunta un control neuronal inverso con un control PID con el objetivo de mejorar el consumo de hidrógeno.

En la sección “Diagnóstico de fallas para sistemas de celdas de combustible PEM mediante programación genética” se presenta un método para la detección y aislamiento de fallas basado en árboles de decisión, utilizando el algoritmo de cómputo evolutivo “programación genética M3GP”. Se decidió utilizar este algoritmo ya que se pueden generar árboles basados en funciones lógicas, esto permite al investigador analizar el comportamiento de los modos de falla y su impacto en las variables del sistema. Se simuló fallas tanto en la celda de combustible como en los sistemas auxiliares (por ejemplo, en el suministro de oxígeno). Por último, se comparó la precisión del algoritmo M3GP con los métodos reportados en la literatura para el diagnóstico de sistemas PEMFC.

Generalidades sobre el control y diagnóstico de sistemas PEMFC

2.1 Introducción a las Celdas de Combustible

Las celdas de combustible son dispositivos que transforman la energía química contenida en el hidrógeno en electricidad. Una pila de combustible está compuesta por un grupo de celdas de combustible individuales colocadas en serie. Cada celda está formada por una membrana de intercambio de protones (PEM) colocada entre dos electrodos (ánodo y cátodo) que están recubiertos con una capa de catalizador, generalmente platino (ver **Figura 3**).

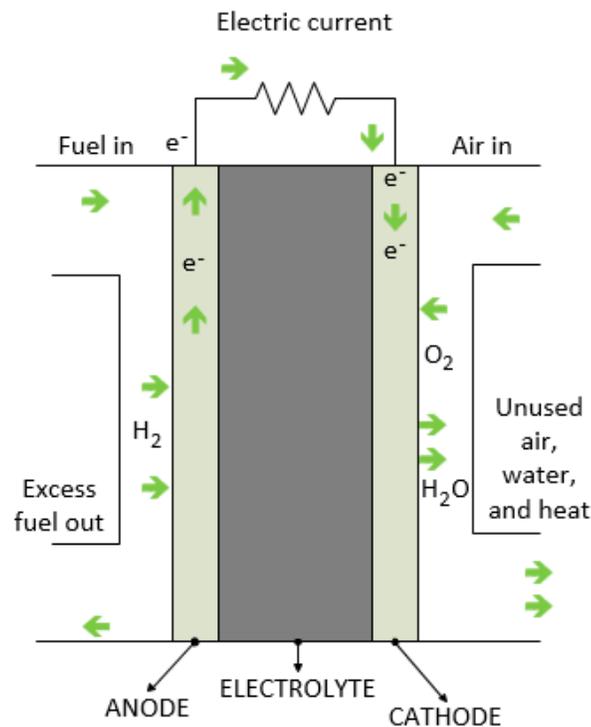


Figura 3 Diagrama de pila de combustible de intercambio de protones

El combustible (hidrógeno) se suministra en el ánodo, y el oxidante (oxígeno, generalmente tomado del aire) se suministra en el cátodo. En el ánodo, el hidrógeno en presencia de un catalizador de platino se ioniza en iones de hidrógeno cargados positivamente y electrones cargados negativamente. En el cátodo, los electrones que provienen del ánodo y los protones que han cruzado la membrana se combinan con el oxígeno del aire para formar agua que sale de la celda de combustible [6] [7].

La reacción general se describe como sigue.



Las principales ventajas de las pilas de combustible con respecto a los convertidores de energía tradicional son:

- alta tasa de conversión
- Bajo nivel de emisiones
- Bajo nivel de ruido
- Escalabilidad del sistema

Desde el siglo pasado se han propuesto una variedad de tipos de celdas de combustible, que se distinguen por el electrolito que se utiliza. La información básica sobre estos tipos de celdas de combustible se resume en la **Tabla 1** [8].

Entre los diversos tipos, debido a su simplicidad, viabilidad, arranque rápido y amplia gama de aplicaciones, la celda de combustible de membrana de intercambio de protones (PEMFC) ha atraído la mayor atención de instituciones académicas e industriales.

Tabla 1 Diferentes tipos de celdas de combustible

Tipo de celda	Ion móvil	Temperatura de operación	Combustible
PEMFC	H^+	50 - 90 °C	H_2
DMFC	H^+	50 - 90 °C	CH_3OH
AFC	OH^-	60 - 250 °C	H_2
SOFC	O^{2-}	750 - 1000 °C	H_2, CO, H_4
PAFC	H^+	160 - 250 °C	H_2

2.2 Principios de Operación de Celdas de Combustible

Como muestra la **Figura 4**, desde el medio hacia ambos lados, una celda combustible PEM típica consta de los siguientes componentes:

- Membrana

El electrolito se coloca entre las zonas activas catódica y anódica de los electrodos. El electrolito en una celda de combustible de membrana de intercambio de protones es una membrana sólida de polímero diseñada para:

- 1) Conducir protones del ánodo al cátodo con la menor cantidad de pérdidas óhmica
- 2) Aislar eléctricamente los electrodos catódicos y anódicos.
- 3) Separar los reactivos catódicos (oxígeno) y anódicos (hidrógeno) a cada lado de la membrana para evitar cualquier reacción directa entre los dos.

- Capas de catalizador (CL)

En las dos interfaces de la membrana de polímero, hay capas con partículas de catalizador. El mejor catalizador tanto para el ánodo como para el cátodo es el platino. Las partículas de catalizador están soportadas por partículas de carbono más grandes.

- Capas de difusión de gas (GDL)

La membrana de polímero, junto con las capas de catalizador asentadas en los dos lados, se encuentra entre dos láminas de capas porosas, que se denominan GDL. Estas capas generalmente están hechas de tela de carbono o papel de fibra de carbono. Una de las funciones de los GDL es permitir la difusión tanto de los gases reactivos como del agua. Además, los GDL también funcionan como electrodos conductores de electricidad y medios conductores de calor. La membrana, las CL y las GDL generalmente se combinan juntas como el conjunto de electrodo de membrana (MEA).

- Placas colectoras (CP)

Las CP se encuentran en los dos externos de una PEMFC. Los canales de flujo de gas (GFC) están ranurados en un lado o en ambos lados para el flujo de gas (los canales de un lado se muestran en la figura). Las CP no solo pueden alojar los GFC, sino que también actúan como conductores eléctricos y conductores de calor. Además, proporcionan soporte estructural para celdas de combustible.

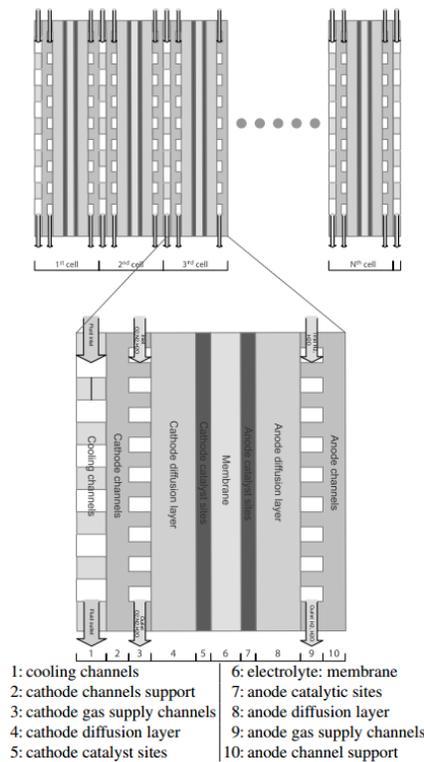
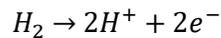
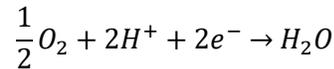


Figura 4 Estructura de la celda de combustible PEM

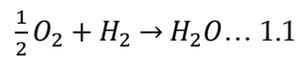
Las reacciones electroquímicas ocurren en la superficie de las CL. En el lado del ánodo, el hidrógeno se oxida:



Los protones viajan a través de la membrana, mientras que los electrones viajan a través de las CL, los GDL, y el circuito eléctrico externo donde realizan trabajos eléctricos y regresan al lado del cátodo. Con los protones transferidos a través de la membrana y los electrones a través del circuito externo, el oxígeno se reduce en el lado del cátodo:



Combinadas las reacciones en los lados del ánodo y del cátodo, la reacción global es:



Con la conversión de energía química en energía eléctrica y calor, el subproducto agua se genera y expulsa principalmente con el aire sin reaccionar del lado del cátodo. La cantidad máxima de energía eléctrica generada en un PEMFC corresponde a la energía libre de Gibbs². En la reacción (1.1), denotando la diferencia entre la energía libre de Gibbs de los productos y la de los reactivos como ΔG (expresada en (eV) para un mol de H_2O), el potencial teórico (voltaje ideal) de la celda de combustible correspondiente ΔG es

$$E = \frac{-\Delta G}{2F}$$

donde F es la constante de Faraday. Esto da un valor de aproximadamente 1,2 V para una pila de combustible operando por debajo de 100 °C [8]. Sin embargo, el voltaje de una PEMFC práctico suele ser menor que este valor (0.9 V). Las pérdidas de voltaje en una celda de combustible se deben a varios factores. La diferencia entre el voltaje de circuito abierto real y el voltaje de salida ideal es causada por varios factores como por ejemplo el cruce de los reactivos.

La curva que caracteriza la relación entre el voltaje y la corriente es denominada curva de polarización. Los tres factores principales que afectan las pérdidas de voltaje son [9]:

1. **Pérdidas por activación.** Esta parte del voltaje se pierde al impulsar la reacción química. Estas pérdidas dependen del material catalizador, la microestructura de la MEA y la densidad de corriente.

2. **Pérdidas óhmicas.** Esta categoría de pérdidas de voltaje es causada por la resistencia asociado a los electrodos, el electrolito y otras interconexiones. Las pérdidas

óhmicas son proporcionales a la densidad de corriente. Además, la estructura de la pila de combustible y la temperatura puede afectar este tipo de pérdida.

3. **Pérdidas por concentración.** Estas pérdidas resultan de la caída de concentración de los reactivos en la superficie de los electrodos, y dependen en gran medida de la densidad de corriente, la actividad de los reactivos y la estructura del electrodo.

La **Figura 5** muestra las zonas de afectación de cada pérdida en la curva de polarización.

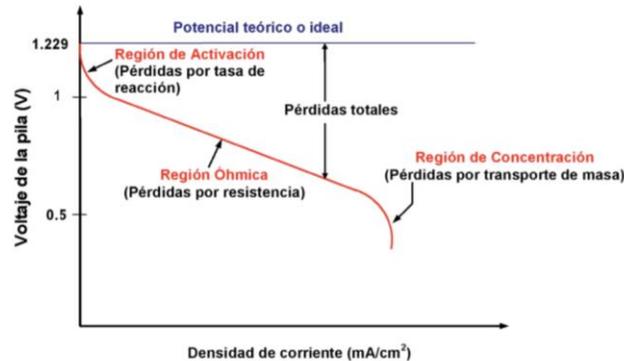


Figura 5 Estructura de la celda de combustible PEM

2.3 Sistema PEMFC

En [10] se mencionan los componentes principales que forman un sistema de celda de combustible PEM. A continuación, se describen estos cuatro subsistemas principales:

- Subsistema de flujo de reactivo

Este subsistema consiste en un circuito de suministro de hidrógeno y aire; su objetivo es mantener una estequiometría adecuada de los reactivos de acuerdo con las condiciones de funcionamiento de la celda. El circuito de suministro de aire en un sistema de celda de combustible de alta presión utiliza un compresor para alimentar el aire, mientras que, en un sistema de baja presión, se utiliza un ventilador de baja velocidad para alimentar el aire.

- Subsistema de temperatura

Una celda de combustible PEM de baja potencia solo necesita un ventilador para regular su temperatura de operación, que es de alrededor de 80 ° C. Una pila de combustible de alta potencia no puede disipar calor por convección de aire y radiación a través de la superficie de la pila; necesita ser enfriado por el flujo de agua desionizada.

- Subsistema de gestión del agua

El objetivo de este sistema es mantener una buena hidratación de la membrana mientras se equilibra el uso / consumo de agua en la celda. Las membranas secas y las celdas de combustible inundadas causan grandes pérdidas de polarización.

- Subsistema de administración de energía

Este subsistema controla la potencia extraída de la pila de pilas de combustible. La corriente de carga se considera una perturbación que tiene un impacto directo en otros subsistemas.

- Subsistema de control

Si el sistema de flujo de reactivo se controla correctamente, las principales variables de la pila, como la temperatura y la concentración de agua, se controlarán indirectamente. Este subsistema tiene un gran impacto en los otros subsistemas; debido a esto, su control es crítico para el rendimiento de la pila. Los fenómenos de polarización en la celda de combustible PEM reducen el voltaje que puede suministrar el sistema cada vez que la carga extrae más corriente, lo que puede afectar el rendimiento del equipo que requiere un voltaje fijo para funcionar correctamente. Por lo tanto, el voltaje de salida de la pila debe controlarse ajustando los caudales de hidrógeno y aire. Otra opción para controlar el voltaje de salida es utilizar medios externos como una batería o un supercondensador o ambos [11].

La pila debe operar con la máxima eficiencia la mayor parte del tiempo para lograr una operación rentable en términos financieros. La optimización del suministro de hidrógeno es un objetivo de control prioritario para lograr una operación rentable ya que, en este momento, el costo de producción de hidrógeno aún es demasiado alto [12].

2.3.1 Modos de falla en sistemas PEMFC

En la **Tabla 2** se muestran los principales modos de falla de los sistemas PEMFC, así como el componente afectado. Una descripción más detallada de las técnicas de detección utilizadas para el diagnóstico de cada una de estas fallas se puede leer en [13] en la sección V.

2.3.2 Consideración sobre las fallas en sistemas PEMFC

La degradación del rendimiento del sistema PEMFC puede ser causada por efecto del envejecimiento o por una operación incorrecta. Esta tesis se centra en el segundo tipo.

Una consecuencia común de estas fallas es la caída de voltaje. De hecho, si ocurre una falla, el voltaje de la pila es la primera indicación de un modo de trabajo degradado. Además, la gestión del agua y la temperatura son parámetros cruciales para el funcionamiento saludable de un PEMFC.

Gestión del agua en PEMFC

- Secado de membrana

La membrana seca no solo aumenta significativamente la resistencia de la membrana sino también la conductividad de los protones a la superficie del catalizador. Por lo tanto, las pérdidas de activación y las pérdidas óhmicas aumentan. Las fallas de secado se producen por la diferencia entre el arrastre osmótico en el lado del ánodo y la difusión hacia atrás en el lado del cátodo (especialmente en alta corriente en el lado del ánodo, la electro-

osmosis es mayor que en el lado del cátodo). Además, en el lado del cátodo, la producción de agua es mayor que el lado del ánodo[14].

La probabilidad de secado generalmente ocurre en el lado del ánodo. El secado causa daños irreversibles a la membrana.

El factor principal que se crea el secado es el siguiente:

- 1) Alimentación de gases de entrada sin suficiente humidificación;
- 2) aumento de la temperatura de la pila de combustible;
- 3) Electro-osmosis a alta corriente.

Tabla 2 Clasificación y modos de falla en sistemas PEMFC

Tipo	Falla	Componente	Modo de Falla
Permanente	Deterioración de la membrana	Membrana	Degradación mecánica
	Envenenamiento por CO	Catalizador/Capa del catalizador (CL)	Degradación electroquímica
	Fuga de reactantes	Junta de sellado	Falla mecánica
	Envejecimiento de la celda	Ensamble del electrodo de membrana (MEA)	Perdida de activación
Transitoria	Inundación	Capa de difusión de gas GDL	Perdida de conductividad / Disminución en el transporte de masa
	Secado	Membrana	Perdida de conductividad
Externa	Sistema de enfriamiento	Bomba de agua / ventilador	Temperatura Alta/Baja
	Sistema de suministro de aire	Compresor / válvula	Perdida de conductividad
	Sistema de suministro de hidrogeno	válvula	Disminución en la generación de poder
	Red de sensores	Sensor	Mediciones incorrectas

- Inundación

Las inundaciones ocurren en ambos lados de la celda, en el lado del cátodo y el ánodo. Las inundaciones tienen lugar en el canal de flujo o en los electrodos de la celda. Luego bloquea los canales de gas y, después de varios minutos, las gotas conducen a la caída de voltaje rápidamente.

Las inundaciones causan un aumento en las pérdidas de transporte masivo (en alta densidad de corriente). Por lo tanto, el rendimiento de FC se reduce. Sin embargo, el voltaje puede recuperarse purgando la celda. La presencia de corrosión por agua ocurre en los electrodos, los medios de difusión de gas y la membrana. Por lo tanto, las pérdidas óhmicas aumentan por este fenómeno y afectan el rendimiento de la pila de combustible [9].

Gestión de la temperatura en PEMFC

En la práctica, los voltajes PEMFC generalmente se mejoran a temperaturas más altas. Las fallas en la gestión del calor pueden conducir a una temperatura demasiado baja o demasiado alta. Por un lado, la temperatura demasiado baja reduce el voltaje de salida de la celda y, por lo tanto, disminuye la eficiencia del sistema. Además, una temperatura demasiado baja ayuda a la condensación del vapor de agua y probablemente cause inundaciones dentro de la celda. Por otro lado, una temperatura demasiado alta debido a una falla en el sistema de enfriamiento puede causar deshidratación de la membrana y, por lo tanto, disminuir su conductividad. Además, el aumento de la temperatura puede causar daños por sobrecalentamiento a la membrana [15].

2.4 Estado del arte sobre el control de sistemas de celda de combustible PEM basado en datos

Esta sección presenta trabajos relacionados con el modelado y control de celdas de combustible PEM utilizando técnicas de inteligencia artificial. En [16], Se presentó una metodología para la identificación de sistemas utilizando redes neuronales NARX (red autorregresiva no lineal con entradas exógenas) y NOE (error de salida no lineal). El modelo de caja negra orientado al control obtenido se implementó en hardware embebido con capacidad limitada para memoria y procesamiento. En [17], se comparó el rendimiento de los modelos clásicos de redes neuronales (NN) y los modelos apilados. El enfoque de apilamiento utilizando mínimos cuadrados parciales como algoritmo de combinación obtuvo la mejor predicción. En [18], los autores compararon un modelo NN con un modelo dinámico que utiliza tres índices estadísticos para validar su rendimiento: el error medio absoluto (AME), el error cuadrático medio (RMSE) y el error de desviación estándar (SDE). El valor máximo de los tres índices indica que el modelo NN es más preciso y exacto, pero tiene una mayor variación en la predicción de los resultados en comparación con un modelo dinámico. Se han probado diferentes métodos para construir modelos empíricos no lineales. En [19], Se comparó el rendimiento de una red neuronal artificial (ANN) y una máquina de vectores de soporte (SVM) para predecir el voltaje de salida de la celda de combustible. El modelo NN presentó un rendimiento excelente al predecir las curvas de polarización de la pila con $R^2 = 0.999$; El modelo SVM exhibió un rendimiento ligeramente inferior con $R^2 = 0.980$. Sin embargo, Kheirandish et al. [20] propuso un enfoque diferente para predecir el rendimiento de una bicicleta eléctrica usando SVM y ANN. Sus resultados mostraron que

SVM tiene una mejor precisión para predecir la curva de potencia, aproximadamente el 99%, mientras que ANN alcanzó una precisión del 97%. Esta diferencia se debe principalmente a la selección de los hiper-parámetros. Se utilizaron técnicas de red neuronal paramétrica (PNN) y método grupal de manejo de datos (GMDH) para predecir y controlar el voltaje de salida de una celda de combustible PEM de 25 W. Las entradas del sistema fueron presión de gas, temperatura de celda de combustible y corriente de entrada. Ambos métodos presentaron una alta precisión en la predicción del voltaje. Sin embargo, el modelo GMDH tuvo menos desviación [21]. Algunos parámetros son difíciles de medir, o es muy costoso medirlos, especialmente en pilas de pilas de combustible. Chávez-Ramírez et al. [22] desarrolló un simulador, basado en ANN, para predecir el voltaje de la pila y la temperatura de salida del cátodo. Llegaron a la conclusión de que los simuladores basados en ANN son capaces de predecir de manera confiable el comportamiento de voltaje y temperatura, ahorrando tiempo y recursos. Se utilizaron redes neuronales recurrentes para desarrollar modelos pronósticos de degradación. En [23], Se utilizó una red neuronal recurrente (RNN) de memoria a largo plazo (G-LSTM) de cuadrícula para predecir la vida útil de las celdas de combustible.

Se proporciona una descripción detallada de las técnicas de control neural aplicadas a las celdas de combustible PEM en [24]. En la **Figura 6** se muestran estos diferentes enfoques. Un sistema de control de avance, que incluye una red neuronal junto con un controlador proporcional, integral y derivado, se presentó en [25]. El objetivo de control era mantener un voltaje de pila apropiado usando un modelo inverso de la planta para calcular la señal de control (presión de aire). En [26], Se desarrolló un control adaptativo de red neuronal con linealización de retroalimentación. Las variables de control fueron los valores de presión de hidrógeno y oxígeno. El modelo presentó un excelente rechazo de perturbaciones, incluso bajo variaciones de carga.

Sin embargo, se han aplicado otras técnicas de inteligencia artificial a los sistemas de celdas de combustible para controlar la velocidad del flujo de aire, la temperatura y el flujo de masa, entre otros. En [27], Se diseñó un controlador de intervalo tipo 2 difuso-integral-derivativo (IT2FPID) para regular el flujo de aire. Los resultados se compararon con los de los controladores PID convencionales y PID difusos tipo 1. IT2FPID presentó un mejor desempeño en términos de respuesta transitoria. En [28], se usó un mapa cognitivo difuso (FCM) para modelar una bicicleta eléctrica alimentada por una celda de combustible. El algoritmo Hebbian fue propuesto para que el FCM aprendiera de sus propios datos.

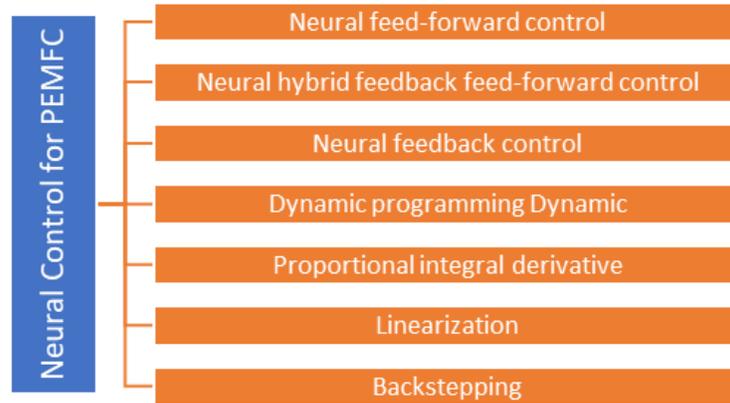


Figura 6 Técnicas de Neuro-Control aplicadas a celdas de combustible PEM

2.5 Enfoques para el diagnóstico de fallas para sistemas PEMFC

En la literatura, se han propuesto diversos enfoques para abordar los problemas de diagnóstico de sistemas PEMFC. Dependiendo de la obtención del modelo que represente la dinámica del proceso a controlar y diagnosticar, dichos enfoques se pueden clasificar en dos clases: métodos de diagnóstico basados en modelos y métodos de diagnóstico basados en datos, en la **Figura 7** se muestra la clasificación de los principales métodos publicados en la literatura. A continuación, se revisan los enfoques propuestos en estas dos categorías. Los documentos citados en este trabajo dan una visión general de cada método.

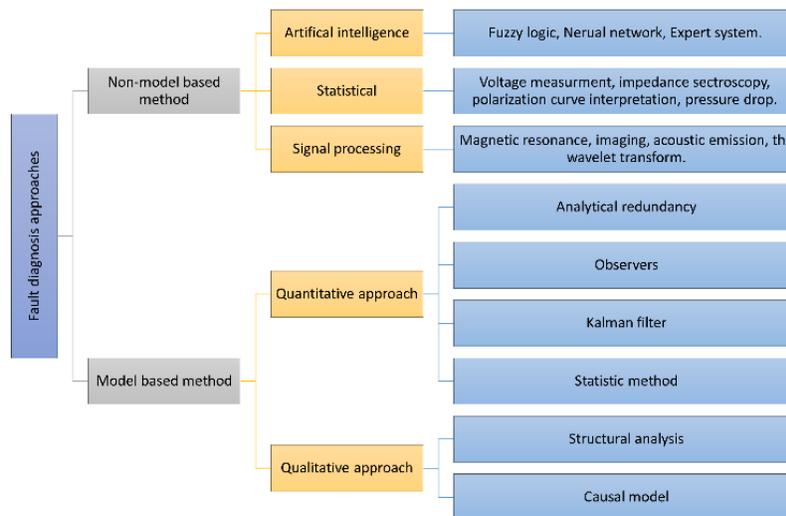


Figura 7 Métodos de diagnóstico de fallas [29]

2.5.1 Introducción al diagnóstico de fallas

El diagnóstico de fallas puede mejorar el rendimiento de una PEMFC en modo saludable y minimizar los daños permanentes provocados por mala operación. Por esta razón, en los

últimos años se ha dedicado mucha investigación al desarrollo de métodos para el diagnóstico de fallas en sistemas PEMFC [13]. En general, según [29], el diagnóstico de fallas para estos sistemas se puede clasificar en dos tipos:

- Métodos basados en modelos matemáticos
- Y métodos basados en datos

El objetivo de los métodos basados en datos es obtener información de fallas basado en conocimiento heurístico (conocimiento del problema obtenido por experimentación que conduce a reglas sin justificación teórica [30]) o en el procesamiento de señales o una combinación de ambos.

El método basado en modelos matemáticos es un enfoque analítico que requiere una comprensión integral y profunda de la célula de combustible y sus fenómenos internos. Para desarrollar este modelo son necesarias una serie de relaciones clave entre los distintos procesos que ocurren dentro de la celda, por ejemplo, proceso electroquímico, termodinámico, eléctrico y transporte fluidos. Una de sus principales desventajas es que los modelos obtenidos son difícilmente generalizables.

El diagnóstico de fallas (FDI) está compuesto de tres etapas, ver **Figura 8**. El objetivo del FDI no es solo detectar cuando ocurre una falla en el sistema, sino aislar el elemento del sistema que provocó la falla, para poder realizar una acción correctiva que regrese el sistema a un estado de operación saludable, en caso de que la falla sea de tipo transitoria o detener la operación del equipo si la falla es crítica.

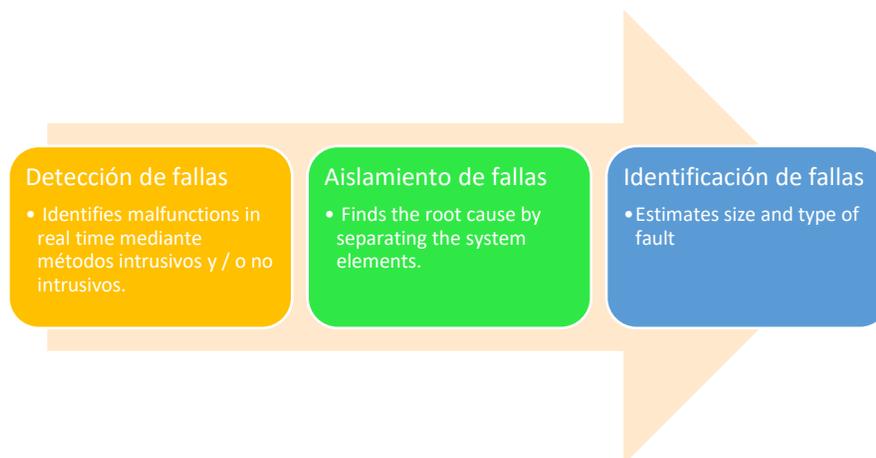


Figura 8 Proceso de diagnóstico de fallas

En ingeniería de control, los sistemas a controlar constan de 4 elementos principales.

- **Actuador**, el cual regula la entrada al sistema;

- La **planta** o proceso, la cual en base a una entrada determinada produce una respuesta deseada;
- **Sensores**, los cuales monitorean en tiempo real las condiciones del sistema;
- y **controlador**, el cual calcula la señal de control para el actuador con el objetivo de que la planta genere una respuesta deseada

Los sistemas de control deben contemplar diferentes aspectos como la disponibilidad de la planta o equipo, rentabilidad (costo de la actuación), robustez, seguridad operativa y la protección del medio ambiente. Para cumplir con este alcance, el sistema de diagnóstico de fallas debe ser capaz de detectar fallas en la planta, el actuador y el sensor. Las fallas que impactan el sistema a controlar se clasifican según se localización (ver **Figura 9**) y se enlistan a continuación.

- 1) **Fallos del sensor:** implica que se toman mediciones incorrectas del sistema real
- 2) **Fallos del actuador:** corresponde a variaciones de la entrada de control $u(t)$ aplicada al sistema controlado
- 3) **Fallas de proceso (planta):** afecta directamente los parámetros del sistema físico y, a su vez, las propiedades de entrada/salida del sistema del sistema o el proceso que deseamos diagnosticar.

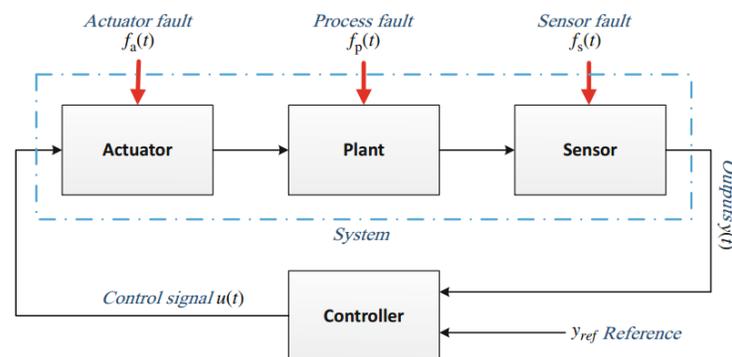


Figura 9 Sistema de circuito cerrado con fallas en actuador, proceso y sensor [31]

El proceso de diagnóstico de fallas se describe brevemente a continuación. Primero se debe generar un modelo, ya sea basado en datos o analíticamente, de cada elemento del sistema, con el cual pueda ser comparado el desempeño real del sistema. La diferencia entre el desempeño real y el modelo se conoce como residuo. Este residuo es evaluado para determinar si corresponde a una firma falla, es decir a un estado del sistema el cual ya fue previamente clasificado como falla, en la **Figura 10** se muestra este proceso. En [32] en la sección 9.5.1 se muestra el proceso de desarrollo de una matriz de firma de fallas.

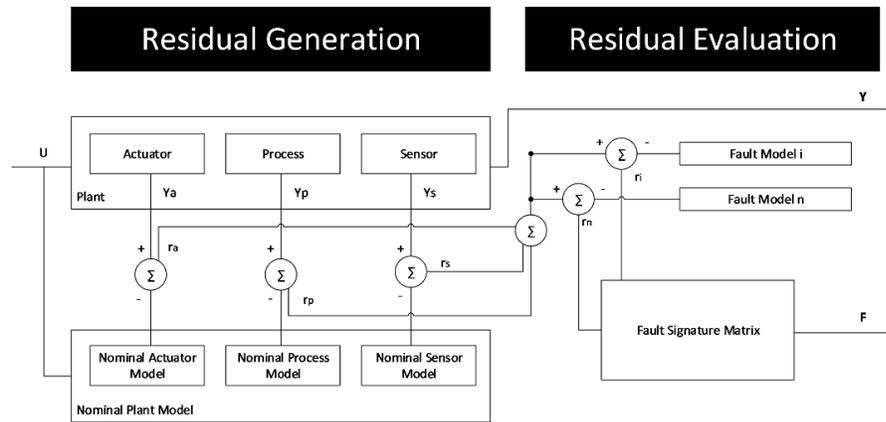


Figura 10 Proceso de generación y evaluación de residuos

2.5.2 Métodos de diagnóstico basados en modelos

Los modelos analíticos, también llamados modelos de caja blanca, explotan en ecuaciones diferenciales espaciales para simular el comportamiento del sistema. Sin embargo, los modelos nunca son perfectamente precisos y los residuos siempre se ven afectados por las incertidumbres introducidas por las mediciones. Para tener en cuenta, la sensibilidad de rendimiento de la herramienta de diagnóstico con respecto a las perturbaciones, se requiere una compensación entre precisión y robustez [33].

De acuerdo con el mecanismo de generación residual, es posible clasificar los métodos de diagnóstico basados en modelos analíticos en tres categorías principales:

- identificación de parámetros basada;
- basado en el observador;
- métodos de espacio de paridad.

Modelos de identificación de parámetros

Para los métodos basados en la identificación de parámetros, los valores de algunos parámetros en los modelos están relacionados con el comportamiento de las operaciones del sistema PEMFC. Cuando la variación de estos parámetros excede un cierto límite, la falla correlacionada puede ser detectada y aislada.

Zeller et al. [34] desarrollaron un modelo cuasi estático para monitoreo y control a bordo, ver **Figura 11**. El voltaje teórico se obtiene combinando la ecuación Nernst y las pérdidas del sistema. Las pérdidas de activación y difusión se modelan como dos fuentes de voltaje diferentes opuestas a la de Nernst; mientras que una resistencia caracteriza las pérdidas óhmicas. Durante las pruebas, los datos son obtenidos mediante lecturas de corriente y se utiliza el método no lineal de mínimos cuadrados para la identificación de parámetros.

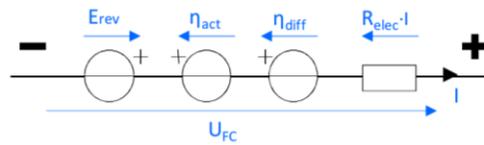


Figura 11 Modelo basado en circuitos cuasi estáticos [34]

Hernández et al. [35] proponen un modelo original destinado a reproducir el comportamiento del sistema durante las inundaciones. Los principales objetivos de la investigación son el modelado global y el diagnóstico de fallas. Los autores desarrollaron un circuito eléctrico equivalente (ver **Figura 12**) para la simulación de leyes de carga, materia y conservación de energía. La dinámica del fluido de gas se tiene en cuenta a través de las analogías entre los elementos neumáticos y los componentes eléctricos. El modelo permite estudiar las composiciones de los gases y su presión parcial. Las capas de difusión de gas, la saturación de vapor y la membrana también se simulan. Sin embargo, el modelo eléctrico no es suficiente para simular el comportamiento del sistema en condiciones extremas.

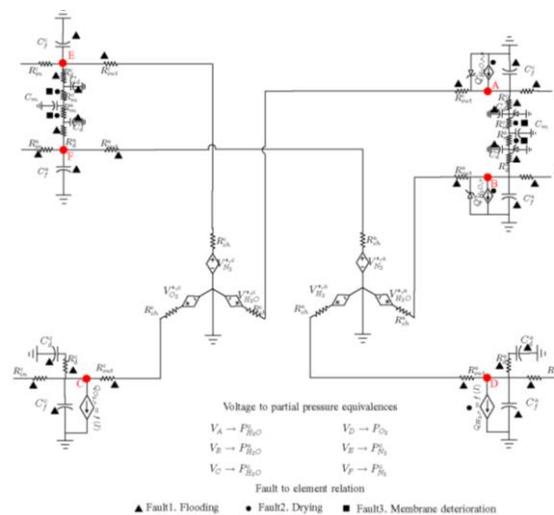


Figura 12 Sistema de circuito equivalente y fallas relacionadas con elementos [35].

Modelos basados en observadores

El modelo basado en observador es uno de los enfoques más comunes implementados para el diagnóstico basado en modelo. En este enfoque, el modelo está integrado con el sistema y se ejecuta en paralelo con él.

De Lira et al. [36][37] proponen un ejemplo de modelo basado en observador para el diagnóstico de sistemas PEMFC, adoptaron un esquema de FDI basado en un umbral adaptativo. El método ha sido probado en el sistema industrial Ballard NEXA^a simulando diferentes fallas [36]. El modelo dinámico desarrollado también tiene en cuenta los

comportamientos de los componentes auxiliares. El modelado del proceso físico se basa en la ley de conservación de masas, principios dinámicos electroquímicos, termodinámicos y de fluido de dimensión cero junto con ecuaciones empíricas. Para el cálculo residual se aplica un observador de variación lineal de parámetros (LPV) con la estructura de Luenberger. Esta metodología permite la linealización de la ecuación del sistema y la resolución del problema analítico en un espacio de estado de tiempo discreto. Para este propósito, se adopta un sistema lineal que varía en el tiempo. El diagnóstico se desarrolla comparando el comportamiento real del sistema en línea con la respuesta dinámica del modelo. El aislamiento de fallas se realiza verificando la distancia euclidiana entre los residuos relativos observados y los teóricos. El uso del umbral adaptativo garantiza la robustez del método en el diagnóstico de sistemas PEMFC. Se desarrolló un conjunto de posibles fallas para probar la robustez del algoritmo. Las salidas del sensor se analizan probando las fallas en:

- presión de suministro del sistema,
- consumo de oxígeno,
- voltaje de pila y
- velocidad del motor del compresor.

Las salidas del sensor se han evaluado con éxito detectando todos los desplazamientos. La técnica ha sido evaluada con éxito para todas las fallas consideradas.

Métodos de espacio de paridad

Basado en el modelo de espacio de estado para la caracterización de la región residual, los métodos de espacio de paridad adoptan las relaciones de paridad en lugar de un observador para la generación de residuos. El enfoque de espacio de paridad permite linealizar el sistema en un subespacio discreto para simplificar el cálculo. La ventaja de esta metodología es su marco de subespacio, que se presenta en forma de ecuaciones algebraicas lineales. Este enfoque para el diagnóstico en línea de la pila de combustible PEM es propuesto por Buchholz et al. [38]. El artículo destaca la complejidad de la implementación en línea para los enfoques basados en modelos, debido al requerimiento de una gran cantidad de mediciones y esfuerzos computacionales para la solución de ecuaciones no lineales. Como solución, los autores proponen linealizar el modelo físico en el dominio lineal del espacio de paridad. Se consideraron diferentes métodos de identificación del subespacio: los algoritmos numéricos para la identificación del sistema de espacio de estado del subespacio "N4SID", la identificación de espacio de estado de error de salida multivariable "MOESP" y el análisis de variante canónico (CVA). Los autores consideran dos enfoques diferentes para el desarrollo del FDI, uno basado en el filtro de Kalman y el otro en el modelo inverso. El propósito es demostrar la aplicabilidad de estos dos métodos. En el primer enfoque, los autores reconstruyen la secuencia de estado del filtro de Kalman

directamente en función de los datos de entrada / salida del sistema. En el segundo enfoque, utilizan el CVA para desarrollar el modelo. En la implementación del método, se utilizan todas las entradas medibles de la pila y el voltaje medio de la celda. Los enfoques de filtro de Kalman y modelo inverso muestran cómo se pueden implementar los modelos lineales de espacio de estado CVA para estimar las entradas no medibles. El enfoque del modelo inverso muestra los mejores resultados cuando en el diagnóstico de fallas.

2.5.3 Métodos de diagnóstico basados en datos

Selección de variables

El diagnóstico basado en datos se puede hacer en función de diferentes variables, y la elección de las variables para la realización de un diagnóstico es crítica. La medición de las variables puede ser obtenida por experimentos específicos o por muestreos periódicos de las variables de estado del sistema. A continuación, se listan los métodos más utilizados para el monitoreo del estado de salud y el diagnóstico de fallas en PEMFC.

Curva de polarización (PC): Es una gráfica del voltaje de celda en relación a la densidad de corriente. Como muestra la **Figura 13**, el voltaje de salida disminuye con el aumento de la corriente debido a varias pérdidas de voltaje.

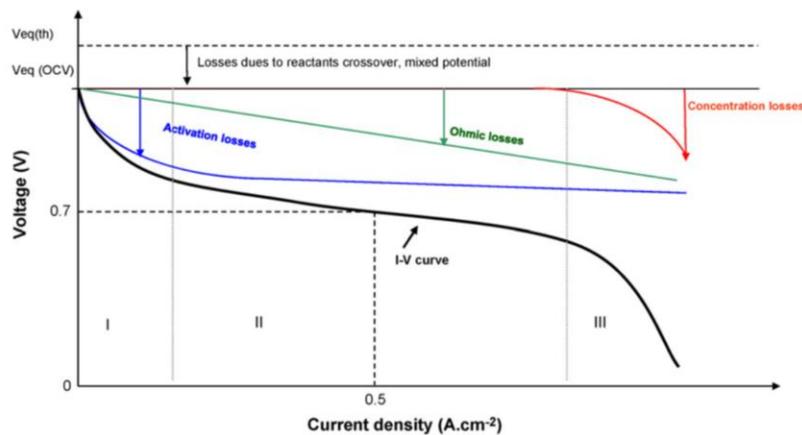


Figura 13 Aspecto de la curva de polarización con las diferentes pérdidas de tensión [9]

Espectroscopía de impedancia electroquímica (EIS): La técnica EIS se utiliza para medir la dependencia de la frecuencia de la impedancia de una celda de combustible aplicando un pequeño potencial de CA sinusoidal (o corriente) como señal de perturbación a la celda de combustible y midiendo la respuesta de corriente (o potencial). Esta técnica fue utilizada en [39] para diagnosticar las siguientes fallas inanición reactiva (tanto combustible como oxidante), inundación, secado, envenenamiento por CO y envenenamiento por H₂S. Por lo general, los puntos de cruce del EIS con eje real expresan respectivamente la resistencia interna y la resistencia a la polarización, ver **Figura 14**.

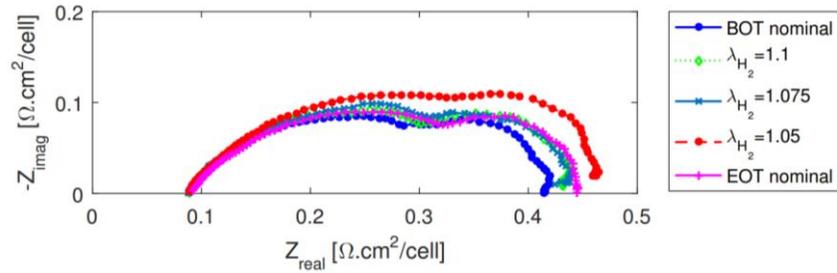


Figura 14 Espectros EIS de inanición anódica [39]

Voltamperometría de barrido lineal (LSV): El LSV es una herramienta poderosa para estimar el cruce de hidrógeno y detectar cortocircuitos electrónicos internos entre el ánodo y el cátodo. En los experimentos con LSV, el ánodo generalmente se alimenta con hidrógeno mientras que el cátodo se alimenta con nitrógeno. Un potencial entre el electrodo de trabajo y un electrodo de pseudo-referencia (en el cátodo) se barre linealmente en el tiempo, mientras que los electrones fluyen en el electrodo de trabajo se mide como corriente [40], en la **Figura 15** se muestra un ejemplo.

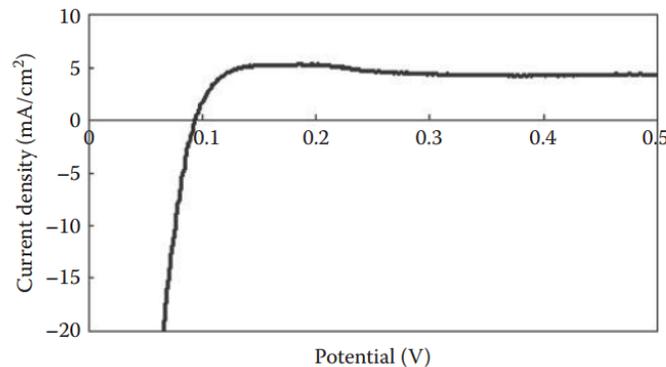


Figura 15 Una curva de LSV típica medida a 60 ° C para una membrana comercial de Nafion bajo 100% de humidificación [41]

Voltametría cíclica (CV): Esta técnica está dedicada a la evaluación de los rendimientos electrocatalíticos. La celda se alimenta con hidrógeno en el ánodo y gas inerte (nitrógeno) en el cátodo, respectivamente. En comparación con los experimentos de LSV, el voltaje del electrodo de trabajo se barre no solo de valor bajo a valor alto, sino que también se realiza una exploración de voltaje inverso adicional.

Método de interrupción de corriente: En este método, la corriente proporcionada por una celda o pila se corta repentinamente. El perfil de la tensión en el proceso se muestra en la **Figura 16**. Las pérdidas óhmicas se reducirán inmediatamente a cero, mientras que las pérdidas de activación tardarán un tiempo en desaparecer [8]. Por lo tanto, del perfil de voltaje, se podrían extraer las pérdidas óhmicas y las pérdidas de activación.

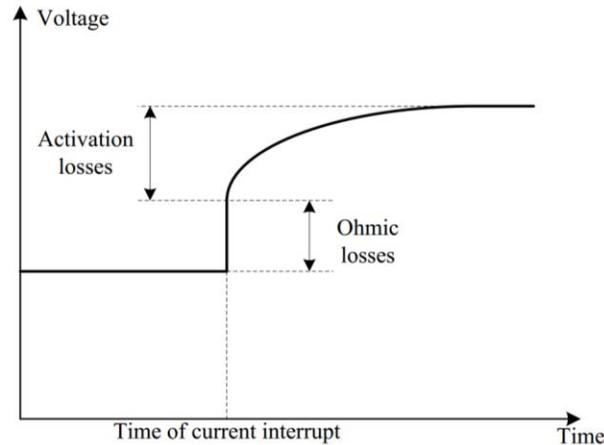


Figura 16 Voltaje contra tiempo en un proceso de interrupción de corriente

Las variables de los sistemas PEMFC y el método de medición se resume en la **Tabla 3**. Las variables que se pueden muestrear a través de sensores generalmente se usan para el diagnóstico de fallas; mientras que las variables obtenidas por mediciones especiales generalmente se usan para diagnosticar las degradaciones relacionadas con el efecto de envejecimiento dentro de las celdas de combustible. La mayoría de las mediciones especiales deben realizarse fuera de línea, lo que significa que es necesario detener el funcionamiento normal para iniciar las mediciones.

Tabla 3 Métodos de medición de variables en PEMFC

Tipo de medición	Variable	
Directa (Sensores)	Voltaje de la pila	En línea
	Voltaje individual	En línea
	Corriente	En línea
	Temperatura	En línea
	Presión	En línea
	Tasa de flujo	En línea
	Humedad	En línea
Medición especial	Curva de polarización	Fuera de línea
	EIS	Fuera de línea
	LSV	Fuera de línea
	CV	Fuera de línea

Métodos de AI para el diagnóstico de fallas en PEMFC

El diagnóstico de fallas se puede considerar como un problema de clasificación. Sin embargo, existen varias limitaciones al aplicar algoritmos de clasificación para el diagnóstico de fallas. En primer lugar, los clasificadores son técnicas supervisadas, lo que significa que los datos etiquetados como modos de falla u operación en estado saludable

deben obtenerse para la clasificación de fallos. Esto podría llevar mucho tiempo al realizar experimentos en el sistema PEMFC y puede causar daños irreversibles al sistema.

Inspirada en las redes neuronales biológicas, se demostró que la red neuronal artificial (ANN) es una herramienta poderosa para aprender y construir un mapeo no lineal cuando un conjunto dado de datos de entrada y salida está disponible. En comparación con los métodos analíticos tradicionales, la NN tiene la ventaja de una excelente capacidad de aproximación no lineal. Además, tiene una baja sensibilidad al ruido y se puede construir con una base de datos incompleta.

El procedimiento de diagnóstico basado en la clasificación generalmente se realiza en dos pasos. En primer lugar, se establece un clasificador empírico a partir de conocimientos previos y datos históricos. Esto se considera el proceso de entrenamiento. Luego, utilizando el modelo del clasificado, los datos en tiempo real se clasifican en ciertas clases que corresponden a los estados de salud.

También se ha observado que el rendimiento de la clasificación se puede mejorar combinando algunos análisis de señales y/o extracción de características, así como los métodos de selección de características. En la **Figura 17** se puede resumir un procedimiento de diagnóstico general basado en la clasificación.

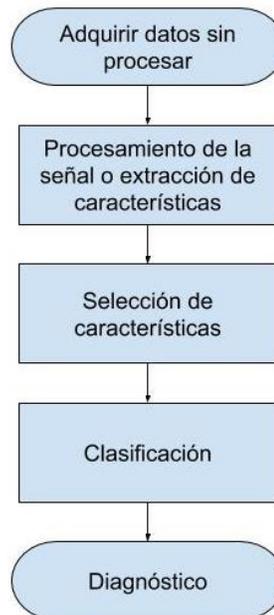


Figura 17 Diagrama de flujo de un proceso de diagnóstico basado en clasificación

Métodos estadísticos para el diagnóstico de fallas en PEMFC

Los métodos de análisis multivariante tienen como objetivo eliminar el alto grado de redundancia observado con frecuencia en los datos mediante la definición de un conjunto reducido de variables estadísticamente no correlacionadas.

En la **Figura 18** puede mostrar el principio del análisis multivariado. Suponga que se monitorean las dos variables x e y , y que el diagnóstico de fallas se basa en ellas. Las muestras en estado normal se distribuyen en la zona de color gris. Si las variables individuales se investigan de manera independiente, los límites superior e inferior de las dos variables $[x_1, x_2]$, $[y_1, y_2]$ se podrían encontrar considerando la distribución de cada variable en el estado de operación normal. Con estos límites, se puede decidir si un punto está en estado normal si está dentro de los límites variables. Sin embargo, para algunos puntos, como *point 1* y *point 2*, se diagnosticarán erróneamente al estado normal, aunque estén fuera de la zona normal. Comparativamente, mediante el uso de análisis multivariante, los datos se pueden proyectar a un nuevo espacio compuesto por x' , y' , y la zona normal obtenida en el nuevo espacio puede rechazar el *point 1* y el *point 2* con éxito.

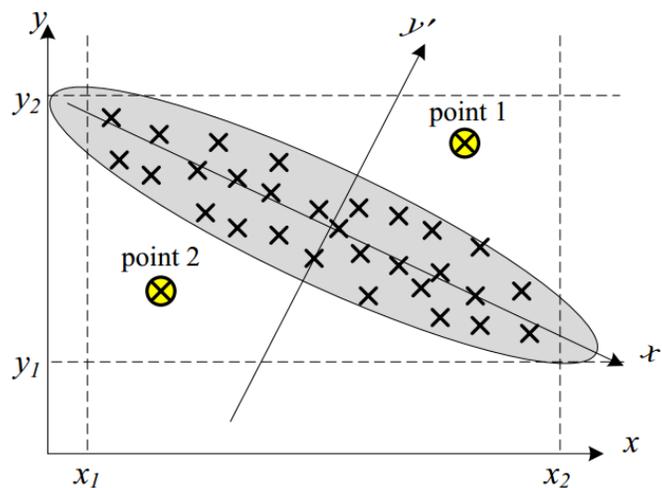


Figura 18 Esquema del análisis multivariante

En [42], la técnica de análisis multivariante, PCA (análisis de componentes principales), se utiliza para el diagnóstico de PEMFC. El PCA se utiliza aquí principalmente para la detección de fallos. La teoría del PCA es construir un modelo de componente principal en condiciones normales de trabajo utilizando la relevancia de las variables del proceso, y encontrar la falla probando la divergencia de muestras del modelo de componente principal [42]. Para lograr esto, se utilizó un conjunto de datos de 17 parámetros para construir el modelo PCA fuera de línea, y las fallas se pueden detectar a través del modelo PCA.

Aunque se cree que los métodos de análisis multivariante son simples y eficientes para realizar la detección de fallas, parece que solo son eficientes para tratar los problemas de diagnóstico en estado estable. Además, el rendimiento de aislamiento de fallas de los métodos de análisis multivariante suele ser débil. Además, se requieren muchas variables, lo que significa que es necesario instalar muchos sensores en el sistema.

2.6 Desafíos del diagnóstico de fallas de sistemas PEMFC en línea

Aunque se ha trabajado mucho para lograr un diagnóstico de sistemas PEMFC eficiente y confiable todavía existen algunos desafíos que nos motivan a realizar más esfuerzos en este tema.

1. El diagnóstico de fallas basado en modelos ofrece no solo la detección sino el aislamiento e identificación de fallas de una manera más estructurada que el diagnóstico basado en datos. Si se dispone de un modelo que describa la dinámica de la celda de combustible con precisión es deseable utilizar una estrategia de diagnóstico analítica. Sin embargo, aún es difícil encontrar, en la práctica, un modelo que describa la celda de combustible con se tiene como caso de estudio. Por lo tanto, deben realizarse más esfuerzos para modelar los sistemas de pilas de combustible.
2. Los enfoques disponibles se centran principalmente en la capacidad de detección de algunas fallas específicas. Para que el procedimiento de diagnóstico sea más completo y eficiente, se debe dar más importancia al análisis y al aislamiento de fallas.
3. Para implementar el diagnóstico, también se deben tener en cuenta los costos, especialmente en el caso de producción de alto volumen. Los costosos sensores o instrumentos deben evitarse si es posible. Por ejemplo, los sensores para medir el caudal de gas y los instrumentos especiales para las mediciones de EIS.
4. Generalmente la estrategia de diagnóstico contempla las fallas relacionadas con la operación del sistema. Sin embargo, también se deben de considerar la degradación causada por los efectos del envejecimiento. La estrategia de diagnóstico debe de distinguir las degradaciones causadas por fallas de las causadas por el envejecimiento.

Motivado por los desafíos antes mencionados, esta tesis doctoral está dedicado a avanzar en el diagnóstico de los sistemas PEMFC. Especialmente, se centra en el diagnóstico basado en datos. Cabe destacar que el estudio lanzado no solo es útil para el diagnóstico de celdas de combustible PEM, sino que también se adaptará a el diagnóstico de otros sistemas industriales.

2.7 Algoritmos evolutivos

La programación genética (GP) es un algoritmo evolutivo que optimiza tanto la estructura como los parámetros de un mapa de entrada-salida. En esta tesis, GP se usará para aprender y refinar iterativamente un árbol de falla que permita diferenciar los modos de falla de la operación en estado saludable del sistema PEMFC utilizando como entradas los sensores del sistema tratando de evitar daños mayores a la celda.

El concepto original de GP fue concebido por John Koza de la Universidad de Stanford y sus asociados [43]. GP es una de las áreas de la computación evolutiva. Cuando los conceptos de GP se aplican en IA (Inteligencia Artificial), los procesos de aprendizaje, formación de hipótesis y resolución de problemas se denominan "aprendizaje evolutivo" o Aprendizaje automático basado en genética (GBML). Este método de aprendizaje se basa en la aptitud física e implica la transformación del conocimiento y la eliminación de soluciones no aptas mediante un proceso de selección para preservar las soluciones apropiadas en la generación posterior.

2.7.1 Contexto histórico

Alan Turing, en su artículo "Computer Machinery and Intelligence", comenzó el debate sobre la capacidad de las computadoras para emular el pensamiento humano. Alan diseñó la prueba de Turing para comprobar si una computadora puede demostrar si tiene razonamiento. Durante esa década, se introdujeron más términos como: "Inteligencia artificial (IA)" [McCarthy 1955] y "Máquina de Aprendizaje (ML)" [Samuel 1959]. Otra rama de investigación se desarrolló basándose en los procesos de selección natural, esta rama diseñó algoritmos que crean iterativamente una población para alcanzar un punto final deseado. Esta rama se denominó "Computación evolutiva (EC)". Ya en su artículo de 1948 "Intelligence Machinery", Turing decía: "Existe la búsqueda genética o evolutiva mediante la cual se busca una combinación de genes, siendo el criterio el valor de supervivencia".

A principios de los noventa, el término general "Algoritmo evolutivo (EA)" se utilizó para agrupar un conjunto de cuatro "métodos" que implementaban los principios darwinianos de selección y variación natural. Ve la **Figura 19** para tener un panorama general.

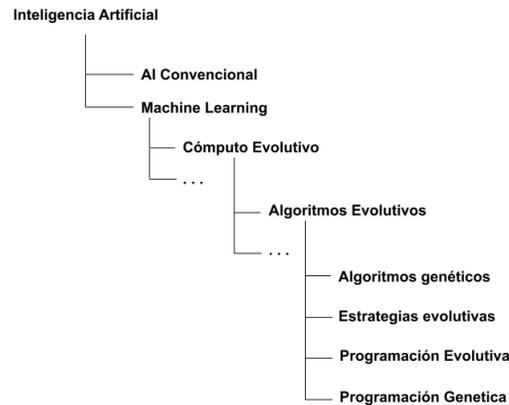


Figura 19 Clasificación de los distintos algoritmos de Inteligencia Artificial

2.7.2 Programación genética

La programación genética desarrolla una población de soluciones candidatas para resolver un problema. Para ser más específicos, una población se transforma iterativamente aplicando operaciones genéticas con el objetivo de minimizar o maximizar un criterio de optimización conocido como *función de ajuste*. Las operaciones genéticas incluyen cruzamiento, mutación y reproducción.

Las cuatro etapas principales de la GP son:

- Definir la estructura del programa para representar una solución;
- Inicialización de la población;
- Selección mediante Operadores genéticos;
- Evolución mediante el proceso de selección basado en la *función de ajuste*.

A continuación, se describe brevemente las etapas de construcción de la GP. La última sección proporciona un diagrama de flujo del algoritmo básico.

Funciones y terminales

Una estructura de programa se construye con dos tipos de bloques de construcción primitivos (o nodos): funciones y terminales.

Las **funciones** proporcionan a la GP elementos que procesan la entrada para formar la salida. Los ejemplos típicos son: funciones matemáticas (+, -, /, ×), funciones booleanas (y, o, no) y funciones de procesamiento de señales (*integración, derivación*).

Las **terminales** proporcionan al programa un valor y por lo tanto una terminal no tiene entradas, existen varios tipos: entradas al programa, constantes y funciones sin argumentos. El valor de las constantes se elige al azar durante la creación de la estructura del programa. También se denominan *constantes efímeras aleatorias*¹. Los efímeros se utilizan para

representar los parámetros de una solución candidata, estos valores numéricos a menudo se reemplazan durante la ejecución del algoritmo GP.

La elección de la **función** y el conjunto de **terminales** es un paso preparatorio importante para que el algoritmo GP logre alcanzar la optimización.

Representación de la solución

Las funciones y terminales no deben de ser seleccionados al azar, se debe de considerar que estructura es necesaria para representar la solución al problema planteado. Debido a la gran influencia del trabajo de Koza, se asume comúnmente que esta estructura está basada en árboles. Sin embargo, existen otras estructuras, por ejemplo, estructuras lineales y gráficas.

Un árbol GP consta de un grupo de nodos interconectados. El nodo de la raíz se encuentra en la parte superior del árbol, mientras que las hojas (o terminales) aparecen en la parte inferior del árbol. Cualquier subgrupo de nodos interconectados se denomina subárbol. La **Figura 20** muestra un diagrama de un árbol GP. Un árbol se puede modificar fácilmente intercambiando ramas o nodos. La profundidad de un nodo se define como el número mínimo de nodos que deben atravesarse para llegar del nodo raíz al nodo seleccionado. La profundidad máxima del nodo corresponde a la profundidad máxima del árbol. La profundidad máxima del árbol es una medida de la complejidad del programa.

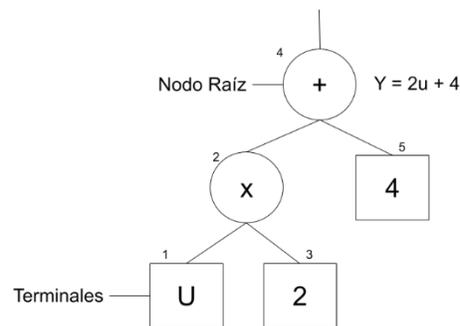


Figura 20 Un ejemplo de árbol de programa, con orden de ejecución

La ejecución del árbol ocurre en profundidad, de izquierda a derecha. El resultado es devuelto por el nodo raíz. En la **Figura 20** esto equivale a la siguiente secuencia, primero se ejecuta la terminal u y se multiplica por un valor constante 2. Finalmente, el valor constante 4 se suma al resultado de la multiplicación.

Inicialización de la población

Un algoritmo de GP comienza con la creación de la población. Una población es un grupo de individuos, donde cada individuo contiene un programa con una estructura de árbol. De hecho, la población posibilita un proceso de búsqueda paralelo. El tamaño de la población

es un parámetro clave para la GP, porque puede determinar si una ejecución es exitosa o no. La creación de la población a menudo se hace al azar, pero también se puede insertar conocimiento a priori utilizando una semilla. Esta semilla contiene una parte de la población inicial y está diseñado manualmente. Además, el resultado de una ejecución anterior se puede utilizar como semilla para continuar la evolución.

Los árboles se crean seleccionando nodos de las funciones y las terminales establecidos al azar. Hay dos métodos diferentes, *completo* y *crecimiento*. El primer método permite que la aridad del nodo determine el crecimiento del árbol. Cuando se selecciona una terminal, esa rama deja de crecer. El método *crecimiento* continúa mientras no se alcanza la profundidad máxima del árbol. Debido a que la elección entre una terminal o una función es aleatoria durante la inicialización, los árboles creados usando *crecimiento* son a menudo de forma irregular. En lugar de seleccionar los nodos al azar de la función y el conjunto de terminales, el enfoque *completo* selecciona solo funciones mientras no se haya alcanzado la profundidad máxima del árbol. En la profundidad máxima, solo se eligen terminales. El resultado es que cada rama llega a la profundidad máxima. La diversidad de arquitecturas de árboles es valiosa para GP, por lo tanto, se ha utilizado una tercera técnica llamada "mitad en rampa y mitad". Cambia entre el enfoque completo y de crecimiento durante la inicialización de una población.

Operadores Genéticos

Una vez que se crea la población inicial, la evolución procede transformando la población con operadores genéticos. Existen tres operadores principales: *cruzamiento*, *mutación* y *reproducción*. Durante la creación de una nueva generación, las operaciones genéticas se seleccionan con cierta probabilidad.

Cruzamiento

El operador de cruzamiento recombina el material genético de dos padres intercambiando una parte de uno de los padres con una parte del otro. La Figura 21 muestra un ejemplo gráfico. El cruzamiento es la base de la hipótesis de los bloques de construcción, la idea es que existen bloques de construcción útiles en la población y que el cruce permite la agregación de buenos bloques de construcción en soluciones cada vez mejores al problema. Esta idea es controvertida en la comunidad de investigadores, pero está claro que el cruzamiento tiene un gran impacto en la estructura del árbol y por lo tanto permite grandes saltos en el espacio de búsqueda.

Mutación

La mutación modifica solo a un individuo seleccionado. Hay muchas implementaciones diferentes disponibles, pero la característica común es que una parte del árbol padre se reemplaza por un subárbol o nodo elegido al azar para crear un nuevo hijo. Dos ejemplos son la mutación de la rama (una rama se reemplaza por una rama recién inicializada) y la mutación puntual (un nodo elegido se reemplaza por un nodo elegido al azar con la misma

aridad). El objetivo de la mutación es preservar la diversidad de la población y prevenir la extinción de ciertos tipos de nodos. La **Figura 22** muestra una operación de mutación de rama.

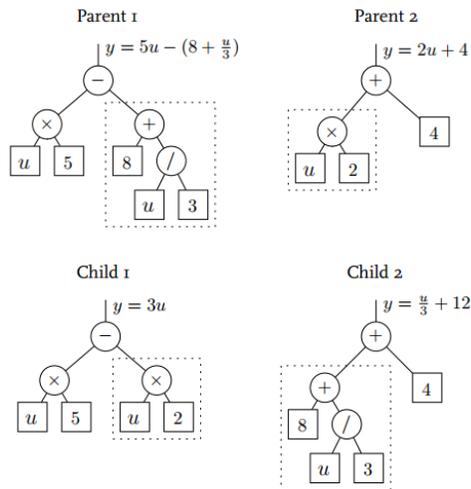


Figura 21 Ejemplo de cruzamiento

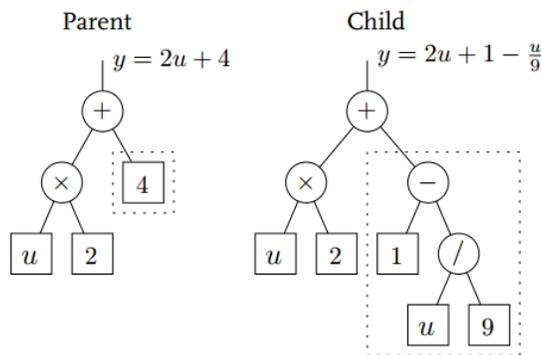


Figura 22 Un ejemplo de mutación de rama

Reproducción

La reproducción es una operación sencilla. El individuo seleccionado se copia sin ninguna modificación a la próxima generación, lo que permite que sobrevivan las buenas soluciones. Cuando la población está sembrada con diseños relativamente buenos, la reproducción evita que estos diseños creados manualmente se descarten prematuramente. Cuando la reproducción se utiliza para preservar a los mejores individuos de la generación, esto se llama *elitismo*.

Evaluación y selección

La búsqueda de la mejor solución está dirigida por un criterio de calidad, conocido como *función de ajuste*. El *ajuste* es una medida de cuán "bien" se desempeña una estructura de programa y la *función de ajuste* es responsable de ejecutar el árbol del programa, interpretar los resultados y asignar la puntuación. Es común definir cero como el resultado óptimo.

La selección de individuos para los operadores genéticos se basa en la función de *ajuste*. Un enfoque común es utilizar el método de selección de torneos, se elige un número de individuos al azar (de acuerdo con el tamaño del torneo) y el individuo con la mejor puntuación *fitness* "gana" el torneo y es seleccionado. Un torneo de gran tamaño da como resultado un algoritmo GP codicioso, es decir, durante la evolución, las operaciones genéticas solo utilizan los mejores diseños. La consecuencia es una rápida disminución de la diversidad de la población y un mayor riesgo de encontrar un óptimo local.

El algoritmo de aprendizaje básico

Ahora que se han introducido los elementos principales de la GP (funciones, terminales, árbol de programa, operadores genéticos y selección basada en la aptitud), se puede ensamblar el algoritmo básico de aprendizaje de GP. Existen dos enfoques principales, uno generacional y uno de estado estacionario. Con GP generacional, la nueva generación creada por los operadores genéticos reemplaza a la generación anterior. Con la GP de estado estacionario, el resultado de las operaciones genéticas se usa para reemplazar a los individuos dentro de la generación actual.

En la **Tabla 4** se presenta el algoritmo básico.

Tabla 4 Programación Genética – Flujo básico

Algoritmo 1.1
Crear población inicial
Evaluar a toda la población en base a la función de ajuste
Mientras condición de terminación no cumplida hacer
Seleccionar individuos más aptos para la reproducción y producir nuevos individuos (cruce y mutación)
Evaluar el ajuste de los nuevos individuos
Genere una nueva población insertando algunos nuevos individuos "buenos" y borrando a algunos viejos "malos"
terminar mientras

Pasos preparatorios

Antes de iniciar el algoritmo GP, se deben realizar estos pasos preparatorios:

- Definir el conjunto de terminales;

- Definir el conjunto de funciones;
- Definir la función de ajuste;
- Definir parámetros para la corrida (como tamaño de la población, profundidad máxima del árbol, probabilidad para cada operación genética, tamaño del torneo y criterio de terminación).

2.7.3 Clasificación por conjunto de reglas de programación genética

Esta sección presenta un panorama general sobre la clasificación de datos utilizando GP para ensamblar reglas lógicas. Primero se describe un problema básico de clasificación de datos y después se discute a detalle el problema de selección de características en este tipo de problemas.

Problema de clasificación de datos

La clasificación de datos es el problema de categorizar un artículo dada la información cuantitativa de las diversas características que describen el artículo. En este contexto, primero, un artículo se define en términos de varias características, y luego los datos sobre un conjunto de artículos de categorías conocidas se adquieren de diversas fuentes. La categoría desconocida de un elemento se determina con un clasificador que se construye utilizando este conjunto de elementos de categorías conocidas.

Formalmente, el problema se puede plantear de la siguiente manera: dado un conjunto de N elementos de entrenamiento $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, construya un clasificador hipotético $h: X \rightarrow Y$ que mapee un elemento $x \in X$ a una etiqueta de clase $y \in Y$.

Demos un ejemplo del problema de clasificación. Supongamos que queremos saber si un elemento es bueno o malo dados los valores numéricos de cuatro características. Para este propósito, primero, construimos un conjunto de datos de entrenamiento de 15 elementos; los elementos se muestran en la **Tabla 5**. Con estos datos de entrenamiento, se aprende un clasificador, que es el siguiente:

$$Clase(x) = \begin{cases} \text{Bueno SI } ((a < 15 \text{ Y } b > 75) \text{ Ó } (c < 125 \text{ Y } d < 55)); \\ \text{Malo de otra manera} \end{cases}$$

donde x es un vector de valores de las cuatro características a, b, c y d . En esta sección, usamos el término instancia para referirnos a un elemento.

Después de que se construye un clasificador a partir de los datos de entrenamiento, su desempeño se evalúa en los datos de prueba independientes (también llamados datos de validación) para determinar con qué precisión el clasificador clasificará las instancias desconocidas. A veces, los datos de entrenamiento y prueba están disponibles; a veces, solo están disponibles los datos de entrenamiento. Cuando solo se dispone de datos de entrenamiento, el rendimiento de un clasificador se evalúa mediante una técnica llamada validación cruzada. Además, durante la selección de características importantes a partir de

los datos, la técnica de validación cruzada se usa ampliamente para la evaluación de la puntuación de bondad de un subconjunto de características. En la validación cruzada de k veces, los datos D se dividen aleatoriamente en k subconjuntos mutuamente excluyentes, D_1, D_2, \dots, D_k de aproximadamente el mismo tamaño. El clasificador es entrenado y probado k veces; cada vez que i ($i = 1, 2, \dots, k$), se entrena con D/D_i y se prueba en D_i . La secuencia de pasos en la técnica de validación cruzada se muestra en la **Figura 23**. Después de la validación cruzada, obtenemos el resumen de las estadísticas de clasificación, es decir, el número de verdaderos positivos (NTP), verdaderos negativos (NTN), falsos positivos (NFP) y falsos negativos (NFN). Por lo general, estas estadísticas se convierten a un valor único, como la precisión o la puntuación F . La precisión de la validación cruzada es el número total de instancias clasificadas correctamente, dividido por el número de instancias en los datos. Cuando un clasificador es estable para un conjunto de datos dado con validación cruzada de k veces, la varianza de la precisión estimada sería aproximadamente igual a $\frac{Precisión(1-Precisión)}{N}$, donde N es el número de instancias en el conjunto de datos.

Tabla 5 Un ejemplo de datos de entrenamiento usados para construir un clasificador

Item	Característica				Clase
	a	b	c	d	
x1	13	87	134	59	Malo
x2	15	76	116	59	Malo
x3	13	81	131	52	Malo
x4	12	64	100	58	Bueno
x5	10	83	145	53	Malo
x6	14	50	123	56	Bueno
x7	16	81	142	54	Malo
x8	16	51	130	55	Malo
x9	19	99	102	53	Bueno
x10	13	60	103	57	Bueno
x11	17	67	138	51	Malo
x12	17	90	102	51	Bueno
x13	12	76	112	50	Bueno
x14	18	96	110	59	Malo
x15	20	94	134	57	Malo

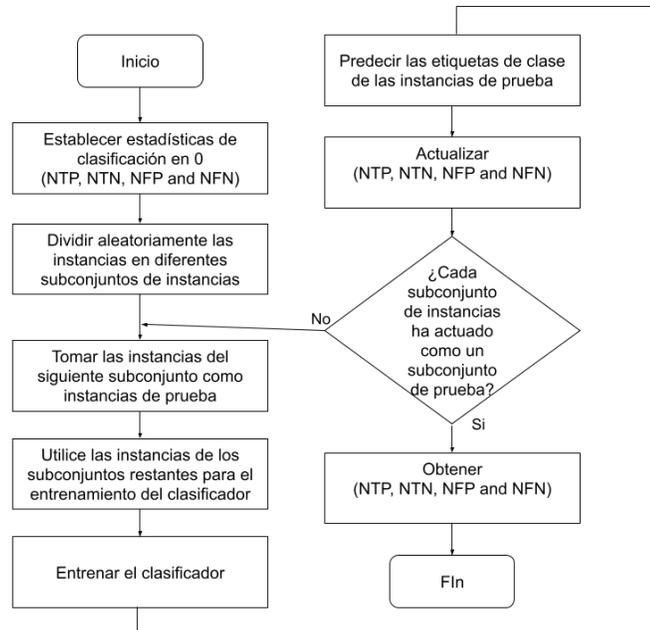


Figura 23 Flujo validación cruzada

Problema de selección de características

A veces, no todas las características de los datos son importantes para la clasificación de las instancias, y en ciertos tipos de datos, especialmente en los datos de *microarrays* de ADN, hay muchas características redundantes. Estas características irrelevantes y redundantes a veces tienen un efecto negativo en la precisión del clasificador y aumentan el costo de adquisición de datos y el tiempo de aprendizaje. Además, una gran cantidad de características ocultan la interpretabilidad del modelo de aprendizaje. La selección de características, también conocida como selección de atributos o reducción de dimensionalidad, es una técnica para seleccionar un subconjunto de características relevantes a partir de los datos, que se usa ampliamente en inteligencia artificial y aprendizaje automático para construir un clasificador rápido y robusto. Aquí, se ofrece un ejemplo de selección de características. Para los datos de entrenamiento en la Tabla 6, un clasificador que puede clasificar perfectamente todas las instancias de entrenamiento se da como:

$$\text{Clase } (x) = a \text{ XOR } b \text{ XOR } c \text{ XOR } e$$

donde la operación *XOR* se define de la siguiente manera:

$$a \text{ XOR } b = (a \text{ AND } \text{NOT}(b)) \text{ OR } (\text{NOT}(a) \text{ AND } b)$$

Con respecto a este clasificador, la característica *f* es redundante como $a = f$, y la característica *d* es irrelevante y tiene un efecto negativo en el clasificador. Al eliminar estas dos características de los datos, obtenemos un subconjunto de cuatro características importantes, *a*, *b*, *c* y *e*, que crean un clasificador perfecto.

Tabla 6 Un ejemplo de conjunto de datos con características irrelevantes y redundantes

Item	Característica						Clase
	a	b	c	d	e	f	
x1	1	1	1	1	0	1	1
x2	0	1	1	1	1	0	1
x3	1	0	0	1	0	1	1
x4	1	1	1	0	1	1	0
x5	1	1	0	1	1	1	1
x6	0	1	0	0	1	0	0
x7	0	0	0	1	0	0	0
x8	1	0	0	0	1	1	0
x9	1	1	0	0	0	1	0
x10	1	1	1	1	0	1	1
x11	1	0	0	1	0	1	1
x12	0	1	0	0	1	0	0
x13	1	0	1	0	0	1	0
x14	0	0	0	1	0	0	0
x15	0	0	0	1	0	0	0

Para un clasificador dado y un conjunto de datos de entrenamiento, se puede encontrar un subconjunto óptimo de características buscando exhaustivamente todos los posibles subconjuntos de características. Para un conjunto de datos con n características, hay 2^n subconjuntos de características. Por lo tanto, no es práctico buscar todo el espacio de forma exhaustiva, a menos que n sea pequeño.

2.7.4 Algoritmo M3GP

Esta sección presenta una extensión del algoritmo [44] de Programación Genética Multidimensional Multiclase (M₂GP, en adelante M2GP) recientemente propuesto, un clasificador GP basado en envoltorios que trata eficazmente problemas multiclase mediante la realización de una transformación multidimensional de los datos de entrada. El algoritmo M2GP utiliza un número fijo de nuevas dimensiones d , que deben elegirse y corregirse antes de que comience la ejecución. Por otro lado, el algoritmo propuesto en [45] es capaz de determinar heurísticamente un valor apropiado para d durante la ejecución.

En la **Tabla 7** se muestra el algoritmo de clasificación M2GP.

Las siguientes subsecciones describen M3GP, que significa M2GP con poblaciones multidimensionales.

Tabla 7 Algoritmo M2GP

Algoritmo 1.2 M2GP – Modulo de entrenamiento

Entrada: X, E, d

Para $g \in 1 \dots G$ **Hacer**

 Generar: $F = \{f_1, f_1, \dots, f_d\}$ Conjunto de d soluciones

 Evaluar: $Z_s = Eval(f_s(X))$ para cada $s \in 1, \dots, d$

 Agrupar: $Z^k \in Z$ para cada $k \in 1, \dots, |C|$

Para $k \in 1, \dots, |C|$ **Hacer**

$c^k = covar(Z^k)$, una $d \times d$ matriz de covarianza

$M^k = centroid(Z^k)$, una $1 \times d$ Vector centroide

$D_i^k = \sqrt{(z_i - M^k) * (c^k)^{-1} * (z_i - M^k)^T}$, $\forall_i = 1, 2, \dots, n$ %Distancia Mahalanobis%

 %donde n es el número de muestras en el conjunto de entrenamiento%

$\forall_i = 1, 2, \dots, n$: $Pred_i = h$ tal que $D_i^h = \min(D_i^1, D_i^2, \dots, D_i^{|C|})$

$\forall_i = 1, 2, \dots, n$: $Coincidencia_i = 1$ Si $Pred_i = E_i$

 Maximizar: $\frac{1}{n} \sum_{i \in n} Coincidencia_i$ %Función de ajuste%

Regresa: F, C, M

Población inicial.

M3GP inicia la evolución con una población aleatoria donde todos los individuos tienen una sola dimensión. Esto asegura que la búsqueda evolutiva comience a buscar soluciones simples y unidimensionales, antes de avanzar hacia soluciones de dimensiones superiores, que también podrían ser más complejas.

Para M2GP, se recomendó una inicialización *Rampa Mitad-Mitad* [43] sesgada a 25% *Crecimiento* y 75% *Completo* [44], lo que sugiere que una mayor proporción de árboles completos facilita la evolución inicial. Debido a que todos los individuos iniciales de M3GP son unidimensionales, tiene sentido creer que la necesidad de árboles iniciales más grandes es aún mayor. Por lo tanto, todos los individuos de la población inicial de M3GP se crean utilizando el método de inicialización completa [43]. Además de la inicialización completa, también se intentó utilizar árboles iniciales más profundos de profundidad 9 en lugar de 6. Sin embargo, los resultados preliminares no mostraron ninguna mejora, y por lo tanto, se utilizó la profundidad inicial tradicional de 6 niveles.

Mutación

Durante la fase de reproducción, siempre que la mutación es el operador genético elegido, se realiza una de tres acciones, con la misma probabilidad: (1) mutación de subárbol estándar, donde un nuevo árbol creado aleatoriamente reemplaza una rama elegida aleatoriamente (excluyendo el nodo raíz) del árbol padre; (2) agregar un nuevo árbol creado al azar como una nueva rama del nodo raíz, agregando efectivamente una dimensión al

árbol padre; y (3) eliminar al azar una rama completa del nodo raíz, eliminando efectivamente una dimensión del árbol padre.

Como se mencionó anteriormente, M3GP comienza con una población que solo contiene individuos unidimensionales. A partir de aquí, el algoritmo será capaz de explorar varias dimensiones diferentes. En M3GP la mutación es la única forma de sumar y quitar dimensiones, por lo que hemos aumentado su probabilidad de ocurrencia de 0.1 (usada en M2GP) a 0.5, para garantizar una búsqueda adecuada de la dimensión correcta. Los resultados preliminares han confirmado que una mayor tasa de mutación de hecho mejora la aptitud.

Cruzamiento

Siempre que se elige un cruce, se realiza una de dos acciones, con la misma probabilidad: (1) cruce de subárbol estándar, donde se elige un nodo aleatorio (excluyendo el nodo raíz) en cada uno de los padres y se intercambian las ramas respectivas; (2) intercambio de dimensiones, donde se elige una rama completa aleatoria del nodo raíz en cada padre, y se intercambia entre sí, intercambiando efectivamente dimensiones entre los padres. El segundo evento es solo un caso particular del primero, donde se garantiza que los nodos cruzados están conectados directamente al nodo raíz.

Poda

La mutación, como se describe anteriormente, facilita que M3GP agregue dimensiones a las soluciones. Sin embargo, muchas veces algunas de las dimensiones degradan el *ajuste* del individuo, por lo que sería mejor eliminarlas. La mutación también puede eliminar dimensiones, pero como se describió anteriormente, lo hace de forma aleatoria y ciega al *ajuste*. Para mantener la simplicidad y la estocasticidad completa de los operadores genéticos, hemos decidido no hacer que ninguno de ellos sea más "inteligente", y en su lugar eliminamos las dimensiones perjudiciales podando al mejor individuo después de la fase de reproducción.

El procedimiento de poda elimina la primera dimensión y reevalúa el árbol. Si la aptitud mejora, el árbol podado reemplaza al original y pasa a la poda de la siguiente dimensión. De lo contrario, el árbol podado se descarta y el árbol original pasa a la poda de la siguiente dimensión. El procedimiento se detiene después de podar la última dimensión.

La poda se aplica solo al mejor individuo de cada generación. Aplicarlo a todos los individuos de la población podría plantear dos problemas: (1) una demanda computacional significativamente mayor, donde se gastaría una cantidad considerable de esfuerzo en individuos que aún no serían aptos después de la poda; (2) aunque no confirmado, el peligro de provocar una convergencia prematura debido a una remoción excesiva de material genético, de la misma manera que la edición de código ha demostrado que lo causa.

Los experimentos preliminares han revelado que la poda del mejor individuo de cada generación cambia la distribución del número de dimensiones a valores más bajos (o evita que cambie a valores más altos tan fácilmente) durante la evolución, sin dañar el *ajuste*.

Elitismo

Se mencionó anteriormente que, para explorar soluciones de diferentes dimensiones, M3GP se basa en la mutación para agregar y eliminar dimensiones de los individuos, con una probabilidad bastante alta. También tiene que depender de la selección para mantener las mejores dimensiones en la población y descartar las peores. La forma de hacer esto es garantizando cierto elitismo en la supervivencia de los individuos de una generación a la siguiente. M3GP no permite que se pierda el mejor individuo de cualquier generación, y siempre se lo transfiere a la siguiente generación. Recordemos que este individuo ya está optimizado en el sentido de que pasó por la poda. Los experimentos preliminares han demostrado que el elitismo puede mejorar el *ajuste*.

2.7.5 Diagnóstico y aislamiento de fallas utilizando Programación Genética

La programación genética (GP) es una técnica de aprendizaje evolutiva que ofrece un gran potencial de clasificación. GP y otras técnicas evolutivas se han aplicado con éxito a diferentes tareas de aprendizaje supervisado como regresión [46] y tareas de aprendizaje no supervisado como agrupamiento [47]. Sin embargo, centramos nuestra revisión en la aplicación de GP a la clasificación. De hecho, GP suele realizar un proceso implícito de selección y extracción de características. La interpretabilidad se puede favorecer fácilmente mediante el uso de GP, ya que puede emplear formalismos de representación más interpretables, como reglas [48].

En [49] la densidad espectral de potencia (PSD) de las señales de vibración de la carcasa de una caja de cambios se utiliza para construir características originales. La PSD se estima mediante el periodograma (una estimación de la densidad espectral de una señal) a partir del cual se evalúan las características. Para obtener características útiles de las originales, que son de alta dimensión, se utiliza la GP para seleccionar y reducir la dimensión de las señales originales. Los resultados muestran que las características seleccionadas por la GP permitan al clasificador etiquetar correctamente los estados de operación con alta precisión.

[50] propone un método de diagnóstico de fallas para maquinaria de planta en una condición de funcionamiento inestable utilizando espectro de potencia instantánea (IPS) y programación genética (GP). IPS se utiliza para extraer las frecuencias de las características de cada estado de la máquina a partir de las señales de vibración para distinguir las fallas mediante la información de cruce relativa.

En [51], se propone un selector de características basado en GP para los datos de vibración sin procesar con el objetivo de clasificar las fallas de rodamientos. Al utilizar características

seleccionadas por GP, el clasificador ANN y SVM mejoran significativamente su precisión, en comparación con aquellos que utilizan características extraídas por métodos clásicos.

En [52] se utilizan conjuntos de características de dos máquinas diferentes para examinar el desempeño para clasificar el estado del normal del estado de falla. Los resultados se comparan con algunos otros métodos para la detección de fallas como redes neuronales artificiales (ANN) y máquinas de vectores de soporte (SVM). Para ANN y SVM, se utilizaron algoritmos genéticos para realizar la selección de características, lo cual es una función inherente de GP. En todos los casos, GP muestra un rendimiento que iguala o mejora el rendimiento de los otros métodos utilizados.

En [53] se utiliza GP para identificar tres estados de operación de válvulas de un motor diésel de seis cilindros. Los estados de operación son condición normal, holgura de válvula-taqué y fallas de fuga de gas. En el ejemplo, se han diseñado especialmente 22 funciones matemáticas y se utilizan 8 características de señal fácilmente obtenidas para construir el modelo de diagnóstico. A diferencia de los GP existentes, el árbol de diagnóstico utilizado en el algoritmo propuesto se construye aplicando un coeficiente de potencia-peso a cada característica. Los coeficientes de potencia-peso varían de forma adaptativa entre 0 y 1 durante el proceso evolutivo.

2.7.6 Software de Programación Genética

A continuación, se enlistan las principales implementaciones de GP. La **Tabla 8** presenta el tipo de licencia y el lenguaje de programación utilizado en cada implementación.

Tabla 8 Lista de implementaciones GP

Implementación	Licencia	Lenguaje
Pyvolution	Apache License 2.0	Python
TinyGP		Python
GeneXproTools	Proprietary	C++
deap	GNU Lesser GPL	Python
pySTEP	MIT License	Python
Pyevolve	Modified PSF	Python
GPLAB		MATLAB
GPTIPS	GNU GPL	MATLAB
PyRobot		Python
ICLEC	GNU GPL	Java
HeuristicLab		C#
PonyGE	GNU GPL v3	Python
inspyred	GNU GPL v3	Python

Control neuronal de voltaje de celda de combustible PEM basado en la regulación de la presión de hidrógeno

3.1 Materiales y métodos

Las etapas de desarrollo del esquema de control propuesto se describen a continuación.

1. Aplique un algoritmo de selección de características para determinar las variables necesarias para modelar y controlar el voltaje de la celda de combustible;
2. Defina las entradas del sistema del subconjunto formado por el algoritmo de selección de características y pruebe diferentes algoritmos de regresión para predecir la variable de salida;
3. Desarrolle el modelo inverso de la celda de combustible, convirtiendo las entradas del sistema en salidas. La salida del modelo de regresión se convertirá en una entrada del sistema.
4. Integre el modelo inverso con un Neurocontrol-PID para rastrear los errores y ajustar la señal de control para alcanzar el valor de referencia en la salida del sistema. La razón para integrar estos dos tipos de control es para modificar la señal de control no solo considerando el error entre la variable de salida y el valor de referencia, sino también considerando el estado de las otras variables en estado transitorio.

3.1.1 Configuración del experimento

El enfoque propuesto se aplicó a los datos de prueba de IEEE 2014 [54]. Los experimentos se llevaron a cabo en un banco de pruebas que permite trabajar las pilas de PEM en condiciones de operación constantes o variables mientras se controlan y registran datos de operación como cargas de potencia, temperaturas y tasas de estequiometría de hidrógeno y aire. Las variables monitoreadas se presentan en la **Tabla 5**.

La pila estaba formada por cinco celdas. Cada celda tenía un área activa de 100 cm². La densidad de corriente nominal de las celdas era de 0,70 A/cm², y su densidad de corriente máxima era de 1 A/cm². La prueba se realizó bajo cambios dinámicos en la corriente de carga (alrededor de 1020 h). La corriente de carga conectada fue de 70 A con oscilaciones del 10% a una frecuencia de 5 kHz. Los rangos de los parámetros operativos se muestran en la **Tabla 4**.

Tabla 4 Rango de parámetros controlados

Parameter	Range
Air flow	0 to 100 L/min
H ₂ flow	0 to 30 L/min
Gas pressure	0 to 2 bars
Temperature	20 to 80 °C
Cell current	0 to 300 A

Tabla 5 Variables monitoreadas

Variable	Description	Unit
Time	Time aging	H
Vout	Stack output voltage	V
I	Current	A
J	Current density	A/cm ²
Tin, Tout H2	Inlet and outlet H ₂ temperature	°C
Tin, Tout Air	Inlet and outlet air temperature	°C
Pin, Pout H2	Inlet and outlet H ₂ pressure	mBar
Fin, Fout H2	Inlet and outlet H ₂ flow	L/min
Fin, Fout Air	Inlet and outlet air flow	L/min
Fwat	Flow rate of cooling water	L/min
HrAIR	Inlet Hygrometry (Air)	%

3.1.2 Algoritmos de selección de características y modelos basados en datos para celdas de combustible

Las técnicas de reducción de dimensionalidad se pueden clasificar en dos grupos: selección de características y extracción de características. Cada uno tiene sus características, y su precisión depende de las características de la base de datos a analizar. Las técnicas de extracción de características logran la reducción de la dimensionalidad combinando las variables. De esta manera, pueden generar un conjunto de componentes nuevos, reduciendo la dimensionalidad de los datos y manteniendo suficiente información para describir el sistema. En algunas aplicaciones, como el análisis de imágenes, donde la precisión del modelo es más importante que la interpretabilidad, estas técnicas son muy útiles. En cambio, la selección de características reduce la dimensionalidad de los datos al eliminar variables irrelevantes y redundantes. Las técnicas de selección de características tienen como objetivo obtener un subconjunto de variables que describe con precisión las características del sistema con una degradación mínima del rendimiento. La selección de características se puede agrupar en tres categorías principales: *Filters*, *Wrappers* y *Embedded*. A continuación, se ofrece una breve descripción de sus características principales:

- Los métodos *Filter* miden la relevancia de las variables por sus correlaciones con la variable de salida;
- Los métodos *Wrappers* crean un subconjunto del conjunto de datos original usando un algoritmo de entrenamiento;
- Los métodos *Filter* son mucho más rápidos que los métodos de *Wrappers* y los *Embedded*.
- Los métodos de *Wrappers* pueden caer en el sobreajuste;
- Los métodos *Wrappers* y *Embedded* capturan las dependencias de las funciones, mientras que los métodos de filtro no.

Los principios operativos de las celdas de combustible PEM incluyen principios de electroquímica y termodinámica que con frecuencia son muy difíciles de modelar matemáticamente. Una alternativa para superar este problema es el uso de métodos de modelado basados en técnicas de inteligencia artificial. En este trabajo, se utilizaron redes neuronales para modelar y controlar las

celdas de combustible PEM porque las técnicas de aprendizaje profundo (Deep learning), en general, presentan un mejor desempeño en el modelado de sistemas altamente no lineales que los algoritmos de aprendizaje automático (Machine learning). La sección XX compara el rendimiento de diferentes algoritmos contra redes neuronales dinámicas. Estos modelos basados en datos se pueden usar como un emulador para detectar posibles fallas en los sistemas de celdas de combustible o para desarrollar un sistema de control neuronal inverso, como se muestra en [55] (Ver **Figura 24**).

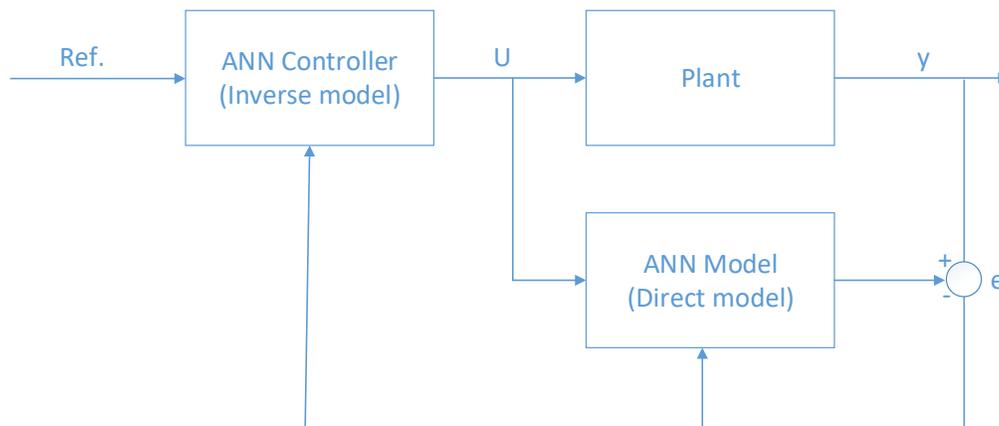


Figura 24 Control neuronal Inverso -Directo

3.2 Selección de características de celda de combustible

Para fines de investigación, toda la información recopilada durante la prueba es útil para comprender y mejorar la calidad del material y el rendimiento del diseño; estas mejoras pueden llevar a aumentar la vida útil y, por lo tanto, a reducir el costo de operación, que en este momento todavía es demasiado alto. Sin embargo, para fines de control, en aplicaciones reales, sería muy costoso instalar todos estos sensores y actuadores. El objetivo del control es identificar las variables operativas críticas y reducir el costo del sistema de control mediante la *Selección de Características*.

Se intentó entrenar un algoritmo de regresión sin aplicar la selección de características. Los malos resultados obtenidos se debieron al ruido generado por la baja correlación de algunas variables. Esta sección presenta los resultados de la aplicación de varios algoritmos de selección de características al conjunto de datos original. Los mejores resultados se obtuvieron utilizando un algoritmo de extracción de características: análisis de PCA. Por esta razón, la Sección 4.1.4 se amplió para describir cómo se seleccionaron las variables.

Métodos Filter

El método de *Correlación de Pearson* seleccionó las siguientes variables: corriente, densidad de corriente y caudales de salida de hidrógeno y aire. Estas variables se seleccionaron debido a que sus grados de correlación son superiores a 0,5. Sin embargo, aunque estas variables podrían modelar el voltaje de la celda de combustible, ninguna de ellas puede considerarse como una entrada del sistema útil para controlar la celda de combustible.

Métodos Wrapper

Se aplicaron dos métodos de envoltura para realizar la selección de funciones: *Recursive Feature Elimination* (RFE) y *Backward Elimination* (BE).

RFE seleccionó 16 variables con una precisión de modelo de 0.85. Las variables eliminadas fueron la densidad de corriente y la temperatura de salida de hidrógeno. BE seleccionó 17 variables de acuerdo con un valor p de 0.05 (estadísticamente significativo). La variable eliminada fue el flujo de entrada de hidrógeno.

La reducción de dimensionalidad lograda por ambos algoritmos, RFE y BE, no fue significativa.

Métodos Embedded

La selección se realizó utilizando *Lasso Regularization*. Si la variable es irrelevante, Lasso penaliza su coeficiente cambiándolo a cero. El mejor puntaje usando LassoCV fue 0.8617. Lasso escogió 11 variables y eliminó las otras 7 variables. La reducción lograda por este algoritmo fue altamente significativa. Sin embargo, el ajuste era apenas aceptable.

Análisis de Componentes Principales (PCA)

El análisis de PCA es un método estadístico utilizado para reducir la dimensionalidad de un conjunto de datos mientras se conserva la mayor cantidad posible de la variación presente en los datos. Para obtener más detalles sobre esta técnica y sus aplicaciones a las celdas de combustible, consulte [56].

El primer paso para realizar un análisis de PCA es hacer una estadística descriptiva que resuma la tendencia central y la dispersión de los valores; El siguiente paso es hacer una matriz de correlación, que nos permita observar qué variables tienen una relación sólida, como se muestra en la **Figura 25**.

Como puede verse, debido a la baja correlación entre las variables, es muy difícil determinar los parámetros para operar un sistema de pila de combustible [21]. Esta baja correlación se debe principalmente a que las celdas de combustible PEM tienen reacciones electroquímicas complejas con múltiples variables de entrada/salida no lineales [57]. Los sistemas con estas características son complicados de modelar con precisión y, por lo tanto, de optimizar.

Las relaciones principales son las siguientes:

- La presión de entrada y salida de aire;
- La presión de entrada y salida de hidrógeno; y
- El caudal de entrada de aire y el caudal de entrada de hidrógeno.

Sin embargo, como se puede ver en la **Figura 25**, existe una correlación negativa entre el tiempo y la corriente (-0.81), así como entre el tiempo y la densidad de corriente (-0.81); Estas relaciones no son tan significativas porque solo reflejan el desgaste natural de la membrana. La variable de tiempo no se consideró durante el proceso de selección de características.

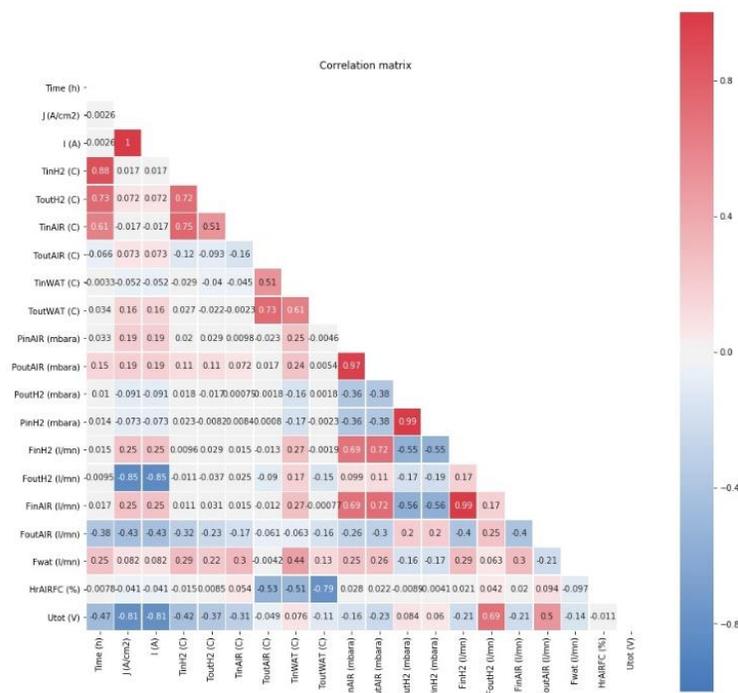


Figura 25 Matriz de correlación de pila de combustible

Después de eso, es necesario determinar el número de componentes que explican la varianza principal de los datos. Este número se obtiene por prueba y error. En este caso, cinco componentes describen la varianza de los datos correctamente (ver Figura 26).

Sin embargo, cuatro componentes explican más del 97% de la varianza. El quinto componente no es relevante, por lo que puede omitirse. En la Figura 27 se puede ver qué variables tienen un impacto importante en cada uno de los cuatro componentes (ver también la Tabla 6).

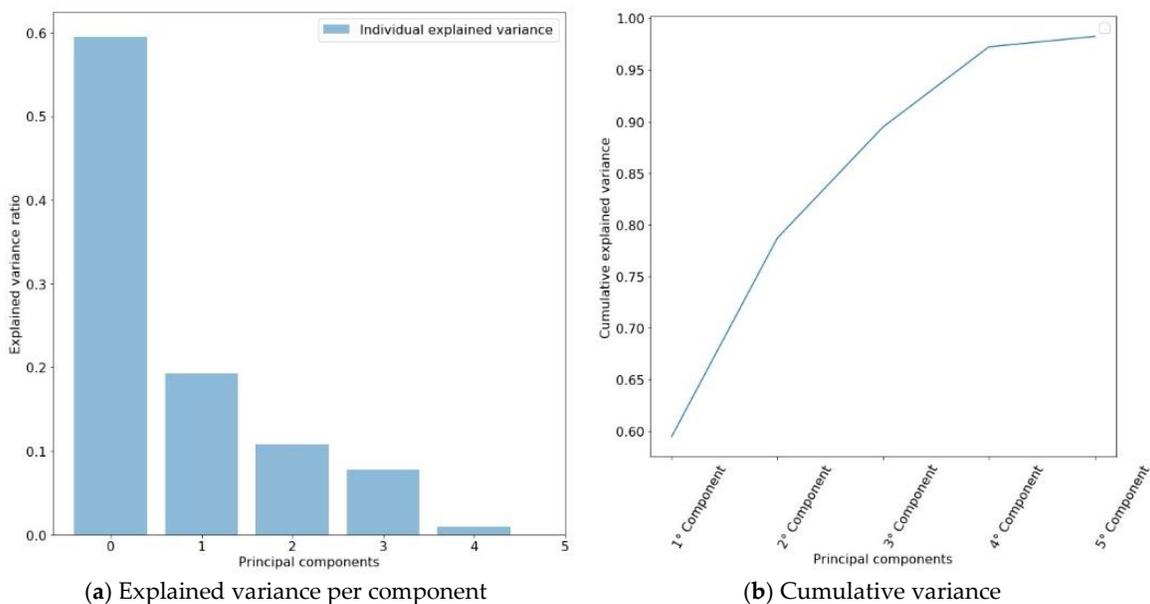


Figura 26 Varianza por componente

Tabla 6 Variables seleccionadas en el análisis PCA

Variables	Type
Current	State variable
Hydrogen inlet temp.	State variable
Air inlet temp.	State variable
Air inlet pressure	Input system
Air outlet pressure	State variable
Hydrogen inlet pressure	Input system
Hydrogen outlet pressure	State variable

Sin embargo, durante el proceso de regresión, es necesario agregar la variable de tiempo debido al desgaste natural de la membrana dependiendo de las horas de trabajo, lo que reduce de manera casi lineal el voltaje de salida. Una vez que se han identificado las principales variables de la celda de combustible, es posible crear un modelo orientado al control para rastrear el voltaje de salida.

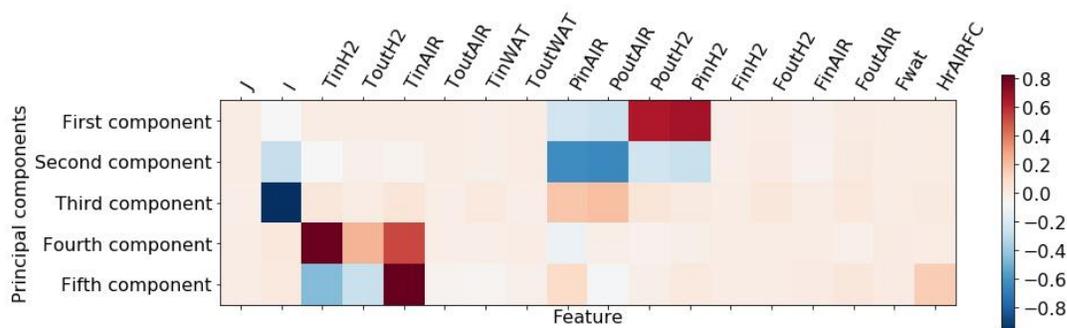


Figura 27 Impacto individual de cada variable en los componentes PCA

3.3 Modelos orientados al control basados en datos para celdas de combustible PEM

Esta sección está dividida en dos partes. La sección 3.3.1 describe los resultados obtenidos por algunos de los algoritmos de regresión más robustos utilizados en el aprendizaje automático. La Sección 3.3.2 se amplía para mostrar con más detalle el proceso de entrenamiento de la red neuronal. Las redes neuronales lograron mejores resultados que los algoritmos probados en la Sección 3.3.1, principalmente debido a su capacidad para rastrear variables no lineales y retrasos del sistema.

3.3.1 Modelado de celdas de combustible utilizando algoritmos de regresión de aprendizaje automático

Se probaron diferentes algoritmos de regresión para crear un modelo robusto orientado al control, y su rendimiento se comparó con el indicador de *Variación explicada*. Se usó el método k-fold para la validación cruzada del modelo utilizando cinco pliegues, y se estableció una semilla fija para garantizar la reproducibilidad. Los métodos comparados fueron Ridge (RID), Bayesian ridge (BYR), regresor de árbol de decisión (DTR), regresor de aumento de gradiente (GBR) y regresor forestal aleatorio (RFR). Los resultados muestran los promedios y las desviaciones estándar de la *Variación explicada*.

- RID: 0.840495 (0.075010)
- BYR: 0.840494 (0.075011)
- DTR: 0.815885 (0.131130)
- GBR: 0.860727 (0.124138)
- RFR: 0.830844 (0.120877)

La **Figura 28** compara a través de un diagrama de caja el rendimiento de los algoritmos probados. En el gráfico, se puede ver que el regresor de aumento de gradiente es el algoritmo que presenta menos variación y mejor precisión. Sin embargo, el regresor de aumento de gradiente solo alcanza una puntuación de 0,86, que es ligeramente baja para fines de control.

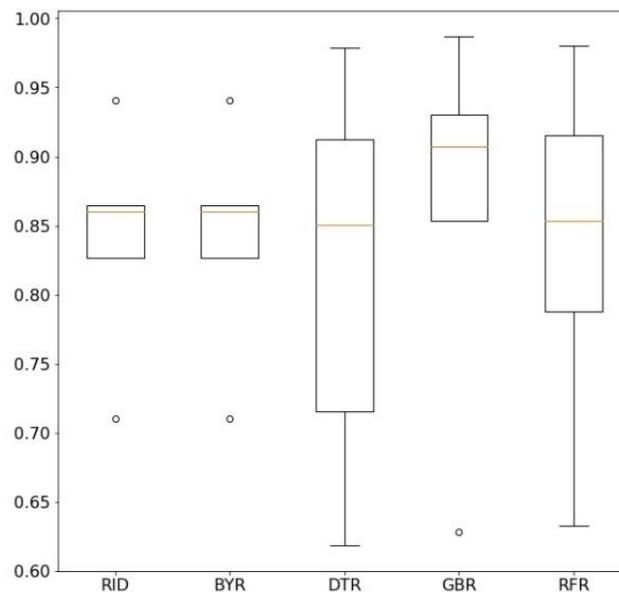


Figura 28 Comparación de algoritmos de regresión

3.3.2. Modelado de celdas de combustible basado en Redes Neuronales

Las redes neuronales se pueden clasificar según su comportamiento en el tiempo como estáticas o dinámicas. Una red neuronal estática puede modelar con alta precisión el rendimiento de una celda de combustible PEM. Sin embargo, como se puede ver en la **Figura 25**, la variable de tiempo impacta negativamente en el voltaje y la corriente de salida, incluso en condiciones de estado estable (véase también la **Figura 29**). Una red neuronal dinámica tiene en cuenta la variable de tiempo, y su estructura se puede utilizar como modelo genérico para el control del sistema [58].

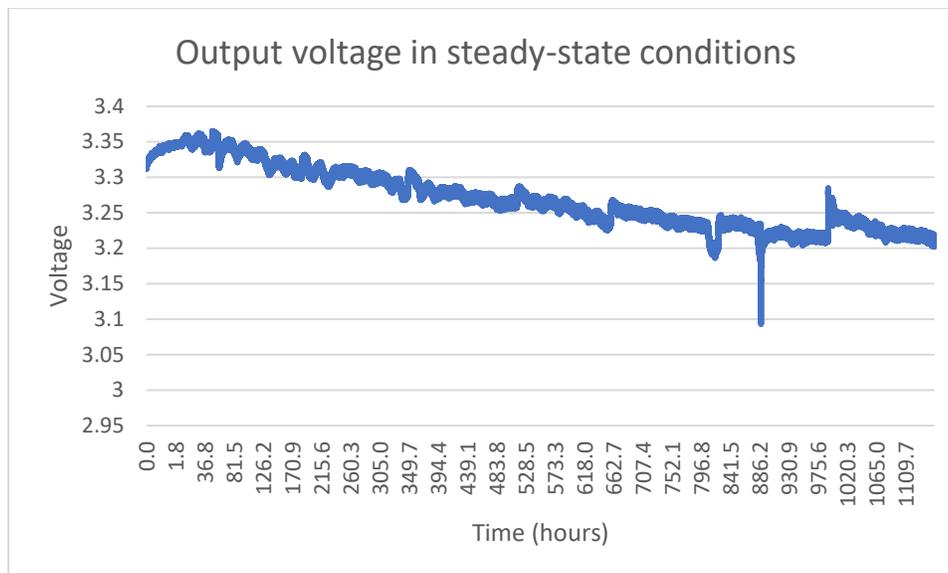


Figura 29 Voltaje de salida de la celda en función del tiempo

Se utilizó una red no lineal autorregresiva con entrada exógena externa (NARX) para modelar la pila de combustible. El proceso de validación se realizó mediante una técnica de validación cruzada (k-fold) con diez divisiones.

El conjunto de datos se divide en conjuntos de entrenamiento y validación. La capa de entrada consta de ocho entradas (las variables seleccionadas en el análisis de PCA, ver **Tabla 6**), la capa oculta tiene diez neuronas con una función de activación log-sigmoidea con dos retrasos (tiempo de muestreo de 30 seg.), y la capa de salida utiliza la función de activación de purelin para calcular el voltaje. La arquitectura de la red neuronal dinámica (DNN) se muestra en la **Figura 30**. El algoritmo de entrenamiento seleccionado fue Levenberg-Marquardt porque, en general, tiene la convergencia más rápida y reduce el costo computacional.

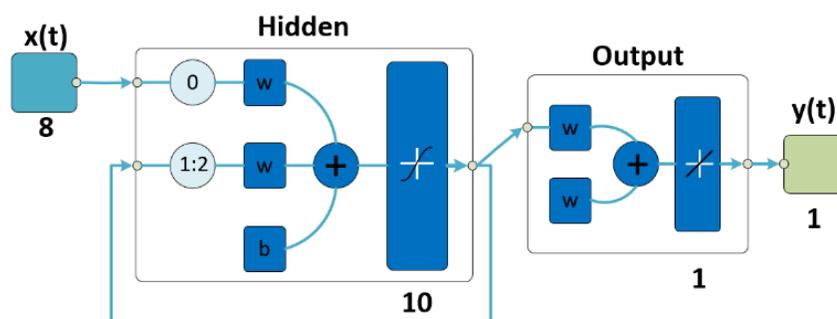


Figura 30 Estructura de la red neuronal dinámica

La alta precisión de regresión ($R^2 = 0.96$) y la rápida convergencia se deben principalmente al hecho de que en el análisis de PCA, se eliminaron las variables irrelevantes y redundantes que no tienen impacto en el voltaje de salida. Las variables eliminadas no tienen valor para fines de control. En la

Tabla 7 se presentan los puntajes de cada pliegue. En la **Figura 31**, se presenta una comparación de los valores reales con los valores predichos.

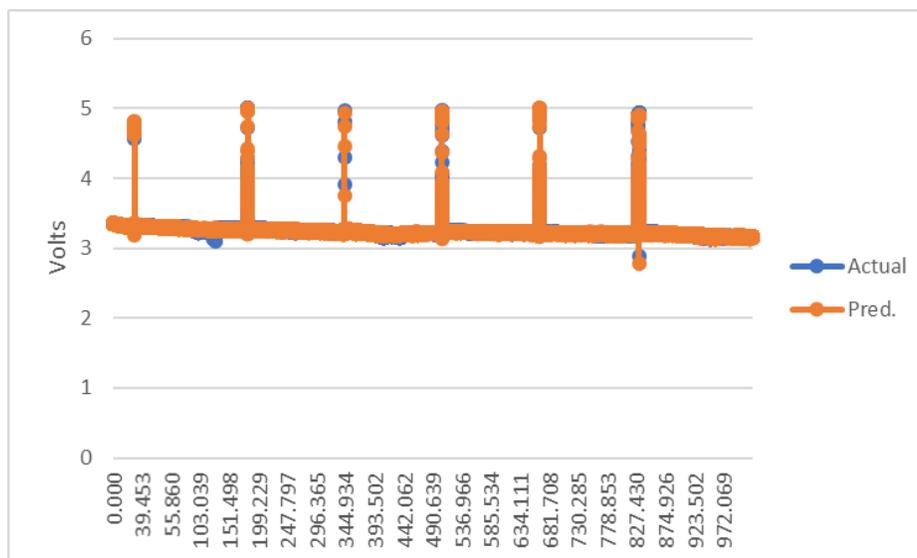


Figura 31 Valores reales contra valores predichos

Tabla 7 Puntuación de regresión de cada pliegue

Fold	Score
1	0.955929840882997
2	0.953074662444409
3	0.959505398134269
4	0.957813889951952
5	0.958048357252375
6	0.958116362163286
7	0.959355574634461
8	0.960026345242201
9	0.966767102061810
10	0.971974506261939
Ave.	0.960061203902970

3.4 Esquema de control híbrido

Según las ocho variables principales identificadas en el análisis de PCA, solo dos pueden considerarse como entradas del sistema: las presiones de entrada de hidrógeno y oxígeno. Sin embargo, al analizar la **Figura 27**, se puede ver que la presión de hidrógeno es la variable que más afecta el rendimiento de la celda de combustible. Las variaciones en la presión de entrada de aire pueden considerarse no representativas para fines de control si se mantienen dentro de un rango específico de operación.

Mantener constante o seguir una referencia no es el objetivo de este enfoque de control para el voltaje de salida de la celda de combustible; esto se debe a que el voltaje de salida no depende solo del suministro de los reactivos. La carga tiene una correlación negativa retrasada en el nivel de voltaje

y en condiciones de estado transitorio es la variable que más lo impacta. La carga (corriente) puede considerarse como una perturbación externa. Para lo anterior, se necesita un control MISO (entradas múltiples, salida única) para suministrar la presión de hidrógeno óptima a la celda de acuerdo con las condiciones de operación, como la temperatura, la corriente o la presión del aire.

El controlador neuro-PID es un enfoque de control ya probado en casos de recuperación de fallas del sistema, como inundaciones, secado y fallas auxiliares, como las de un compresor[24]. Se propuso un esquema de neurocontrol en Serie-PID (con un modelo inverso de la celda de combustible) para suministrar la presión de hidrógeno óptima teniendo en cuenta los valores de las principales variables en condiciones transitorias (ver **Figura 32**). El autoajuste del control PID se realizó de acuerdo con el método propuesto por Omatu et al. [59].

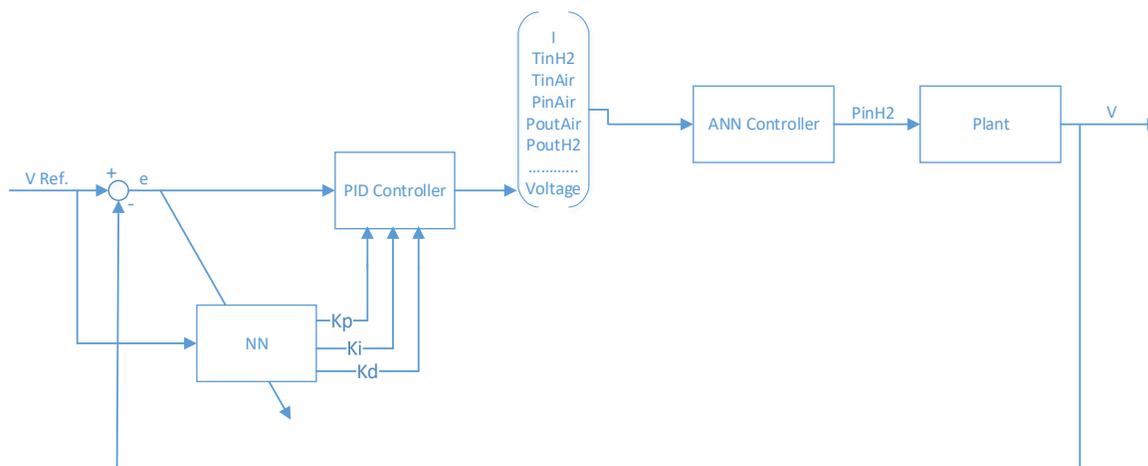


Figura 32 Neuro-controlador en serie PID

El controlador ANN es el modelo inverso de la planta; Esto significa que el voltaje de salida de la planta se convirtió en entrada, y la presión de hidrógeno se convirtió en la salida del sistema. El voltaje nominal es de 3.3 voltios en condiciones de estado estable; sin embargo, este valor nominal depende de los cambios en la carga y su efecto sobre la temperatura. El controlador ANN no solo considera el error entre el valor de voltaje nominal y el valor real, sino que también considera los valores de las variables seleccionadas en el análisis de PCA para estimar la señal de control, en este caso, la presión de hidrógeno. El entrenamiento se llevó a cabo siguiendo el enfoque descrito en la sección 3.3.2.

En la **Figura 33** se comparan el voltaje, la corriente y la presión de hidrógeno para ambos controladores, el control convencional y el neuro-control Serie-PID. Ambos controladores lograron un rendimiento similar en condiciones estables y transitorias. La principal diferencia es la reducción de la presión de hidrógeno en el estado estacionario. Esta reducción en la presión provoca una disminución en el flujo de hidrógeno, que a su vez disminuye el consumo de hidrógeno. Es necesario recordar que la diferencia es solo de unos 45 mbar. En aplicaciones prácticas, se necesitaría un actuador de alta precisión (costoso) para controlar estas pequeñas diferencias.

El algoritmo de entrenamiento deriva el error parcialmente, por lo que cada neurona actualiza su peso de acuerdo con su proporción en ese error. Si el controlador neuro-PID solo considera la brecha entre el voltaje deseado y el valor real sin tener en cuenta los cambios en las variables seleccionadas en el análisis de PCA, la señal de control, la presión de entrada de hidrógeno, no

estabilizará el rendimiento de la celda de combustible. Las celdas de combustible PEM deben funcionar en condiciones de estado estacionario para evitar fallas prematuras, como el *starvation* debido a un suministro de gas inadecuado o una demanda de carga transitoria excesiva [60].

Se requiere un sistema de gestión de energía para entregar un voltaje fijo al equipo para que pueda funcionar correctamente. Este documento propone un enfoque práctico para estabilizar el rendimiento de la celda de combustible en condiciones transitorias a un costo de control mínimo, enfocando la atención en las variables que más impactan en el desempeño de la celda y eliminando mediciones innecesarias. Sin embargo, este control puede mejorarse si la presión de entrada de aire también está regulada. Un proceso de inicio/apagado incorrecto puede causar daños acelerados o permanentes en la capa de catalizador. Estas consideraciones deben incluirse en el proceso de control para mejorar el control propuesto.

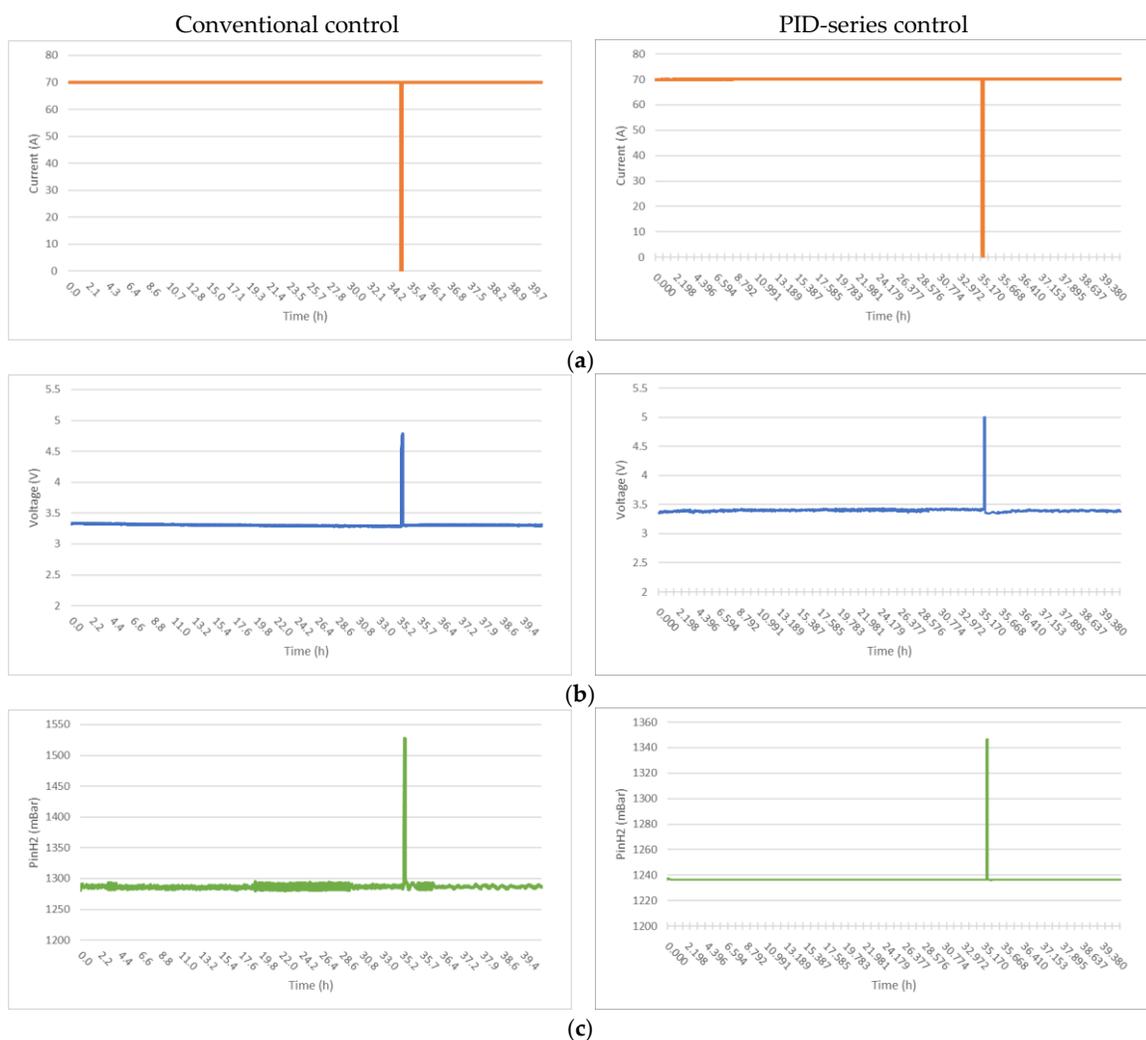


Figura 33 Análisis de rendimiento del Neurocontrol Serie-PID y control convencional. (a) Se conectó la misma carga a ambos controladores. A las 35.42 h se cambió la carga para evaluar el efecto sobre el

voltaje de salida y la presión de suministro de hidrógeno; **(b)** La tensión de salida alcanzó el mismo nivel en estado estacionario en ambos casos prácticamente. Cuando la carga conectada se redujo (hasta circuito abierto) el voltaje de salida aumentó; **(c)** La presión de hidrógeno en el control convencional (izquierda) oscilaba (en estado estacionario) principalmente entre 1280 y 1295 mBar, mientras que en el neurocontrol (derecha), la presión de hidrógeno permaneció prácticamente constante a 1238 mBar en el estado estacionario

3.5 Conclusiones parciales

En este documento, desarrollamos un enfoque de control basado en datos para celdas de combustible PEM para minimizar el costo de control. Se usaron varios algoritmos de selección de características para la reducción de dimensionalidad. El análisis de componentes principales (PCA) obtuvo los mejores resultados al eliminar las variables irrelevantes y redundantes. Las variables seleccionadas (**Tabla 6**) pueden describir con alta precisión el rendimiento de la celda de combustible PEM. Se compararon algunos de los algoritmos de regresión más potentes para predecir el voltaje de salida de la celda. Sin embargo, las redes neuronales obtuvieron la mayor precisión ($R^2 = 0.96$) debido a su capacidad para mapear relaciones complejas no lineales. Con las variables seleccionadas, se desarrolló un modelo inverso de la celda de combustible utilizando redes neuronales para desarrollar un controlador neuro-PID. Se integró un control de la serie PID con el modelo inverso para regular la entrada del sistema (presión de entrada de hidrógeno) al considerar los valores de las otras variables. El nivel de voltaje de la celda de combustible no depende solo del suministro de los reactivos, y en condiciones transitorias, la carga es la variable que más impacta en el rendimiento de la celda de combustible. Este método es una forma práctica de ahorrar tiempo de modelado matemático y reducir la cantidad de sensores en el sistema de control.

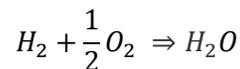
Diagnóstico de fallas para sistemas de celdas de combustible PEM mediante programación genética

El diagnóstico de fallas se realizó reproduciendo la simulación de modos fallas presentada en [61]. El aporte de este trabajo respecto al diagnóstico es la clasificación de los estados de operación utilizando arboles de falla. Las ramas del árbol son las mediciones de las variables de estado y los nodos son funciones lógicas interpretables que sirven para entender mejor el comportamiento de la falla.

En la sección 4.1 se describe el modelo matemático utilizado para la simulación, en 4.2 se describe la estructura del sistema a analizar, en 4.3 se describen las fallas a diagnosticar, en 4.4 se realiza un análisis intuitivo de cada modo de falla utilizando la gráfica de violín, en 4.5 se discuten los resultados del método M3GP y en 4.6 se compara el desempeño de este algoritmo con otros métodos publicados en la literatura.

4.1 Modelo Dinámico PEMFC

Para analizar las fallas en los sistemas PEMFC, primero se debe estudiar su modelo dinámico. En general, se sabe que una PEMFC es una celda de combustible de baja temperatura, alrededor de 55 ° C, por lo que el producto de la reacción es agua líquida. La reacción general en la PEMFC se representa de la siguiente manera:



4.1.1 Voltaje de salida

El voltaje de salida de una pila PEMFC puede expresarse mediante siguiente ecuación:

$$V_{out} = N_{cell} * (E_{cell} - V_{act,cell} - V_{ohm,cell} - V_{conc,cell})$$

Donde V_{out} es el voltaje de salida de una pila PEMFC, N_{cell} es el número de las celdas en la pila, V_{cell} es el voltaje de salida de una sola celda, E_{cell} es el potencial reversible de una sola celda.

Según la ecuación de Nernst, el potencial reversible se modifica de la siguiente manera [12]:

$$E_{cell} = E_{0,eq} - k_E(T - 297.15) + \frac{RT}{2F} \ln[\dot{P}_{H_2} * \dot{P}_{O_2}^{0.5}]$$

Donde $E_{0,eq}$ es el potencial de referencia estándar en estado estándar (297.15 °K y 1 atmosfera de presión), k_E es la constante empírica, T es la temperatura de operación, \dot{P}_{H_2} es la presión de hidrógeno y $\dot{P}_{O_2}^{0.5}$ es la presión parcial de oxígeno.

V_{act} es la caída de voltaje debido a la pérdida de activación de la celda de combustible, es definido por la ecuación de Tafel de la siguiente manera:

$$V_{act} = \xi_0 + a(T - 297.15) + T * b \ln(I)$$

Donde ξ_0 , a , b son constantes empíricas, I es la densidad de corriente de la PEMFC. V_{ohm} la caída de voltaje debido a la resistencia óhmica del PEMFC, se expresa de la siguiente manera:

$$V_{ohm} = I * (R_{ohm} + k_{RI}I - k_{RT}T)$$

Donde R_{ohm} es la parte constante de la resistencia óhmica del PEMFC, k_{RI} y k_{RT} son las constantes empíricas.

En el proceso de reacción en PEMFC, difusión en masa desde ellos canales de flujo de gas a las superficies del catalizador para la reacción influyen en concentración del gas.

V_{conc} es la caída de voltaje debido a lo anterior descripción:

$$V_{conc} = \frac{RT}{2F} \ln \left(1 - \frac{I}{I_{limit}} \right)$$

Donde I_{limit} es el límite de densidad de corriente. I es la corriente general que consiste en I_{out} (la densidad de corriente de salida) e I_{in} (la pérdida de corriente interna asociada con el cruce de combustible).

4.1.2 Transferencia de calor

La razón principal de la variación de temperatura es la transferencia de calor en el PEMFC, que se define de la siguiente manera [13]:

$$\dot{q}_{net} = \dot{q}_{chem} - \dot{q}_{elec} - \dot{q}_{sens} - \dot{q}_{loss} - \dot{q}_{cool}$$

Donde \dot{q}_{net} es la energía térmica neta, \dot{q}_{chem} es la energía química de la reacción del gas, \dot{q}_{elec} es la energía que consume la carga, \dot{q}_{sens} es el calor sensible, \dot{q}_{loss} es la pérdida de calor debido al intercambio de calor entre la pila y el aire circundante, y \dot{q}_{cool} es la eliminación de calor debido al sistema de enfriamiento artificial.

El cambio de temperatura de la pila depende de la energía térmica neta. La siguiente ecuación puede calcular la temperatura:

$$M * C \left(\frac{dT}{dt} \right) = \dot{q}_{net}$$

Donde M es la masa de la pila PEMFC y C es la capacidad de calor específica de la pila. El valor de C puede ser determinado por los componentes de la pila, como el grafito y el acero inoxidable. La capacidad calorífica específica de la pila se estima a partir de los datos de las referencias [9,13].

Cuando ocurren algunas fallas, el estado estable de PEMFC desaparecerá. El modelo transitorio es significativo para esta investigación. El modelo transitorio obtenido mediante el acoplamiento del modelo electroquímico en estado estacionario y el modelo térmico en función del tiempo con el nuevo término de acumulación.

En los experimentos, se simula una pila PEMFC de 500 W para el diagnóstico de fallas. En la **Tabla 8** se presentan los parámetros de esta pila. Basado en el análisis previo del modelo PEMFC y estos parámetros de experiencia, un modelo transitorio de la pila PEMFC de 500 W está construido con el software Python.

Tabla 8 Parámetros de la pila PEMFC

Parametro	Unidad	Valor
Potencia	W	500
No. De celdas		36
Temp ambiental	°C	25.00
Dimensiones	cm	22.5x12.5x17.5
Peso	kg	2.5
E_0	V	44.15
k_E	V/K	0.00085
λ_e	Ω	0.00333
t_e	s	80
ξ_0	V	20.145
a	V/K	-0.1373
b	V/K	-0.000193
R_{ohm}	Ω	0.2793
I_{limit}	A/cm ²	0.468
k_{RI}		0.001872
k_{RT}		-0.0023712
R	J/mol.K	8.3143
F	C/mol	96,487
C	J/kg.K	35

4.2 Estructura del sistema PEMFC a diagnosticar

La **Figura 34** ilustra un sistema PEMFC completo que consta de una pila PEMFC, una carga, un subsistema de suministro de aire, un subsistema de suministro de hidrógeno y un subsistema de enfriamiento de la pila [20-22]. En este sistema, el controlador mantiene la P_{H_2} y la P_{O_2} alrededor de 1.5 atm cambiando la válvula de contrapresión correspondiente. El sistema de enfriamiento de la pila mantiene la temperatura de la pila por debajo de 55 °C. Se adopta una carga resistiva electrónica para controlar la densidad de corriente de la pila PEMFC a un valor especificado [12,14].

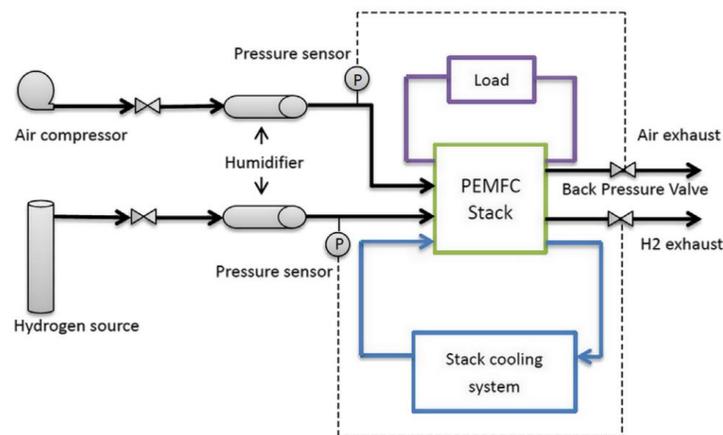


Figura 34 Sistema típico de una celda de combustible

4.3 Descripción de fallas

En los sistemas PEMFC, hay dos tipos de fallas: fallas que pueden detectarse mediante mediciones de sensores y fallas que no pueden detectarse directamente. Ambos tipos de fallas influyen en el funcionamiento normal de los sistemas PEMFC. Sin embargo, en la mayoría de los casos, la detección de fallas en sistemas comerciales PEMFC se limita a la detección de fallas del primer tipo. Este documento se centra en la detección de fallas del segundo tipo.

De acuerdo con la discusión sobre el modo dinámico PEMFC, cuatro tipos de fallas en el sistema se analizan de la siguiente manera.

Falla en el sistema de enfriamiento de la pila F1

El sistema de enfriamiento de la pila se utiliza para mantener la temperatura de la pila dentro del rango aceptado. En experimentos, la temperatura de funcionamiento normal de la pila es de aproximadamente 55 °C. El aumento de temperatura mejora el rendimiento de la pila de PEMFC, sin embargo, este aumento de temperatura seca el MEA rápidamente, lo que provocará su degradación [6,12].

La evolución de los parámetros bajo la falla en el sistema de enfriamiento de la pila se muestra en la **Figura 35**. Durante los primeros 15 minutos, la pila PEMFC funciona bien. La falla en el sistema de enfriamiento de la pila ocurre en el minuto 15, por lo que el q_{cool} disminuye rápidamente y la temperatura de la pila aumenta obviamente (Ec. (10)).

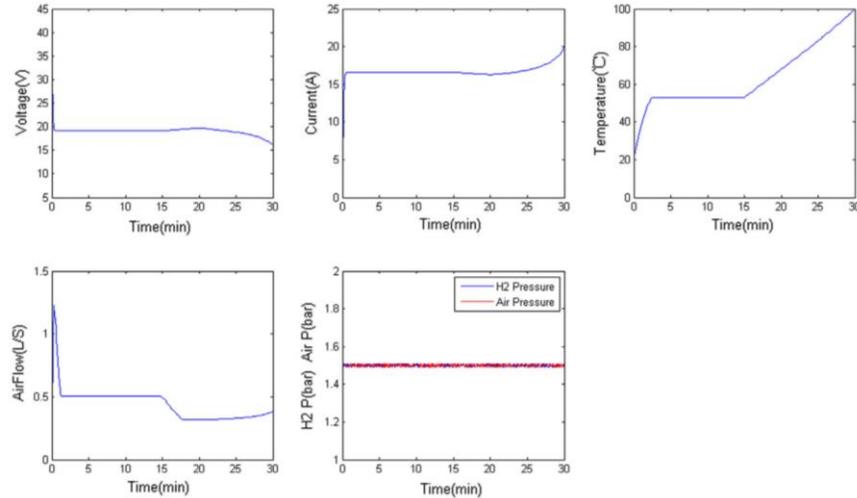


Figura 35 Evolución de los parámetros bajo falla en el sistema de enfriamiento de la pila

Falla en el sistema de suministro de aire F2

Hay dos fallas principales en el sistema de suministro de aire: fuga de aire y la falla en el compresor de aire. Ambas fallas conducen a las siguientes consecuencias: primero, la presión del aire cae obviamente, lo que disminuye el efecto de la reacción en el PEMFC. En segundo lugar, las fallas conducen a la disminución del oxígeno que debe reaccionar con el hidrógeno, y luego la disminución del oxígeno conduce a la disminución de la corriente directamente. La última consecuencia es la acumulación de agua líquida que debe ser arrastrada por el flujo de aire. Una gran acumulación de agua inunda los electrodos, lo que dificulta la difusión del gas y afecta el rendimiento de la pila [4,6].

La **Figura 36** ilustra la evolución de los parámetros cuando la falla en el sistema de suministro de aire se produce. La presión del aire se reduce de 1.5 atm a 1.2 atm en el minuto 15.

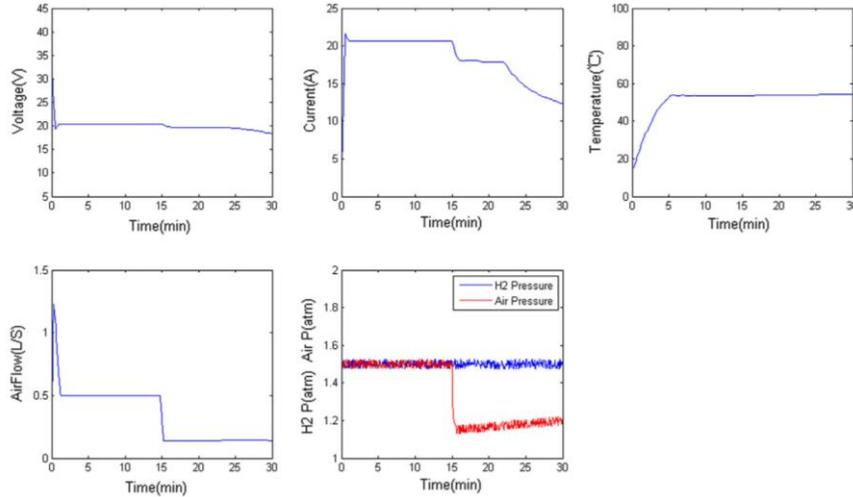


Figura 36 Evolución de los parámetros bajo falla en el sistema de suministro de aire

Incremento del cruce de combustible F3

En el caso del fenómeno de cruce de combustible, una pequeña cantidad de hidrógeno puede transferirse a través de la membrana y reaccionar directamente con el oxígeno para producir solo calor. En condiciones normales, solo unos pocos iones de hidrógeno se transfieren a través de la membrana, por lo que I_{in} es muy pequeño. El aumento de la temperatura local aumenta la degradación de MEA [6].

La **Figura 37** ilustra la evolución de los parámetros bajo el aumento del cruce de combustible. Durante el aumento del combustible cruzado dentro del PEMFC, hay un aumento en la temperatura de la pila.

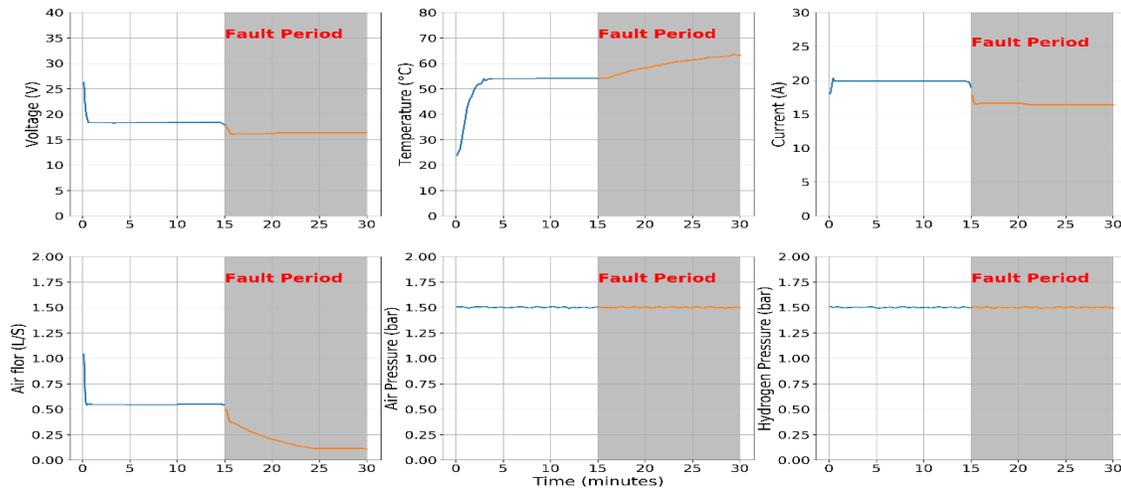


Figura 37 Evolución de los parámetros bajo aumento de combustible cruzado

Falla en el sistema de suministro de hidrógeno F4

En teoría, cuanto mayor es la presión de hidrógeno, mejor es el rendimiento de la pila. Cuando ocurre la falla en el sistema de suministro de hidrógeno, la reducción de la presión de hidrógeno reduce la densidad de corriente directamente y degrada el rendimiento de la pila. La presión de hidrógeno debe mantenerse a 1,5 atm. La **Figura 38** ilustra que la presión de hidrógeno se reduce de 1.5 atm a 1.2 atm en el minuto 15, y luego otros parámetros cambian obviamente

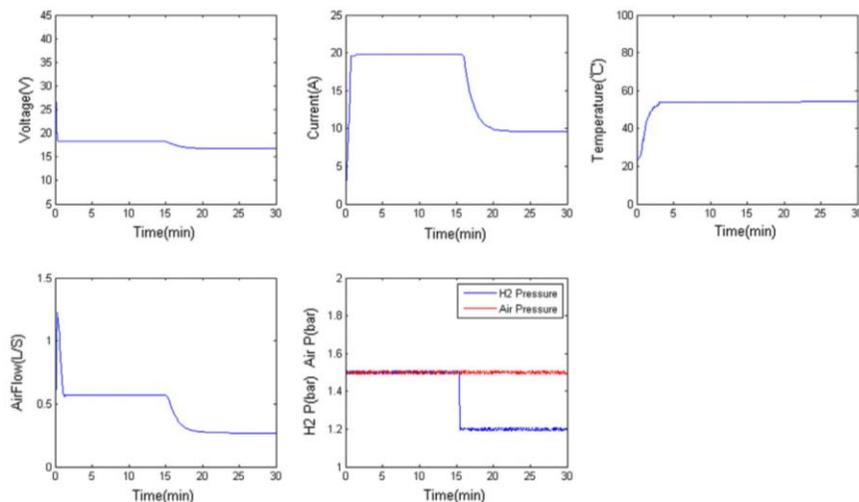


Figura 38 Evolución de los parámetros bajo falla en el sistema de suministro de hidrógeno

4.4 Etiquetado de datos y configuración de parámetros M3GP

Este trabajo se centra en la detección de fallas que no se pueden medir directamente pero que se pueden inferir en función de la distribución de valores de las variables del sistema. La simulación de cada falla se ejecutó durante 30 minutos. Desde el minuto 15 al 30 se provocó la avería. Se tomaron 200 muestras de cada experimento. Las primeras 100 muestras de la simulación se etiquetaron como estado saludable y las últimas 100 se etiquetaron como estado de falla. Los estados bajo análisis y su ubicación en el sistema PEMFC se enumeran en la **Tabla 9**.

Tabla 9 Estados de operación bajo análisis

Descripción estado de operación	Ubicación	Notación
Saludable	Sistema entero	NO
Incremento de temp	Subsistema temp	F1
Decremento en la presión de aire	Subsistema de aire	F2
Incremento paso de combustible	MEA	F3
Decremento presión de hidrogeno	Subsistema de hidrogeno	F4

La **Tabla 10** resume los parámetros establecidos para ejecutar el algoritmo M3GP. Estos parámetros se utilizaron para generar arboles de falla tanto individuales como globales. Se seleccionaron funciones lógicas como operadores para mejorar la interpretabilidad de los árboles.

Tabla 10 Parámetros de operación MEGP

Variable	Descripción
Función de ajuste	M3GP precisión
Operadores	Crossover, mutación
Tamaño población	200
Funciones	plus, minus, myif, and, min, and max.
Terminales	voltage, current (load), temperature, airflow, air pressure, hydrogen pressure
Conjunto terminales	Rand
Tipo de prob. operadores	Fixe
Muestreo	lexictour

4.4 Análisis de fallas

Una manera practica de visualizar el impacto que tiene un modo de falla en las variables de estado del sistema, en este caso la celda de combustible, es la utilización de la gráfica de violín, ya que nos permite entender de manera intuitiva como se distribuyen los valores de las variables de estado durante la simulación de la falla, así como interpretar la correlación entre ellas. En la **Figura 39** se puede observar el comportamiento de las variables de estado que se utilizaran para el diagnóstico y aislamiento de fallas, tanto en estado saludable como en los cuatro modos de fallas a estudiar. Las 6 variables de estado son:

- Voltaje
- Corriente (carga)
- Temperatura
- Flujo de aire
- Presión de aire
- Presión de hidrógeno

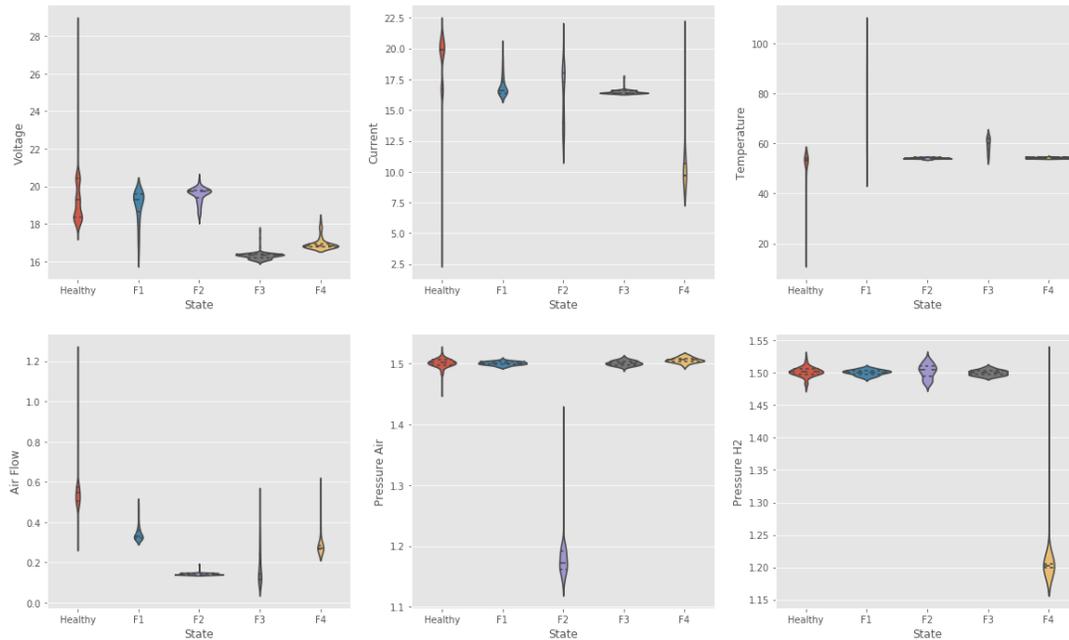


Figura 39 Distribución de las variables de estado en estado saludable y en modo de falla

Falla en el sistema de enfriamiento de la pila F1

En la **Figura 40**, que corresponde a la falla 1, se puede observar cómo esta falla se puede diagnosticar por el efecto negativo que tiene el incremento de temperatura en el voltaje, la cual es la variable de salida o variable de control. También se puede observar que el incremento de temperatura, disminuye el flujo de aire y aumenta la corriente de la celda. La presión de aire e hidrógeno prácticamente no se ven alteradas en este modo de falla. Este comportamiento característico de cada modo de falla se conoce como firma de falla y es la base para el desarrollo del árbol de falla que se utilizara de manera online para el diagnóstico.

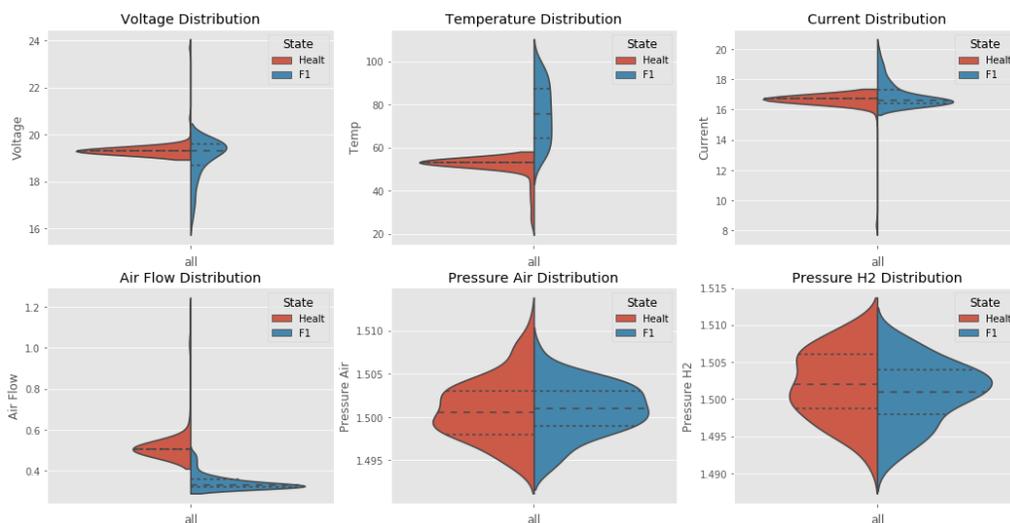


Figura 40 Distribución de valores en el modo de falla F1

Falla en el sistema de suministro de aire F2

Obviamente esta falla reduce el flujo de aire debido a que la presión de aire paso de 1.5 a 1.2 atm, como puede verse en la **Figura 41**. Esta falla no puede ser diagnosticada en base al cambio en el voltaje o la temperatura de celda, debido a que la distribución de los valores de estas variables se concentra en rangos similares. Sin embargo, en este modo de falla, la reducción los valores de la presión y flujo de aire se concentran en distintos rangos que en modo saludable. Este modo de falla también afecta negativamente la densidad de corriente de la celda.

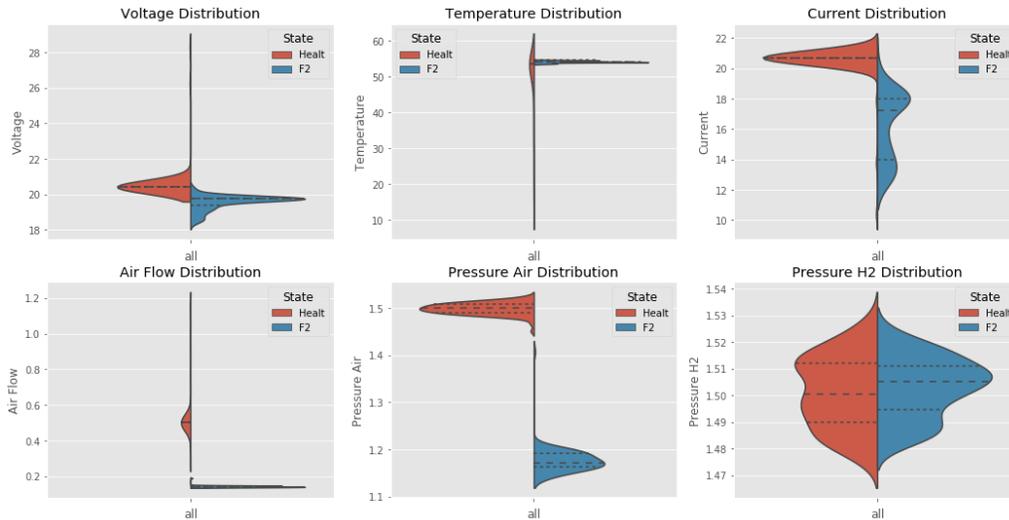


Figura 41 Distribución de valores en el modo de falla F2

Incremento del cruce de combustible F3

Como se comentó en la sección anterior este modo de falla provoca que el nivel de voltaje de salida de la celda disminuya debido a la falta de moléculas de oxígeno en el cátodo. Este modo de falla también reduce la densidad de corriente y genera aumento de temperatura, ver **Figura 42**. Lo característico de esta falla es la reducción en el flujo de aire sin la reducción de la P_{O_2} . El comportamiento de estas variables permite el diagnóstico (clasificación) correcto de este modo.

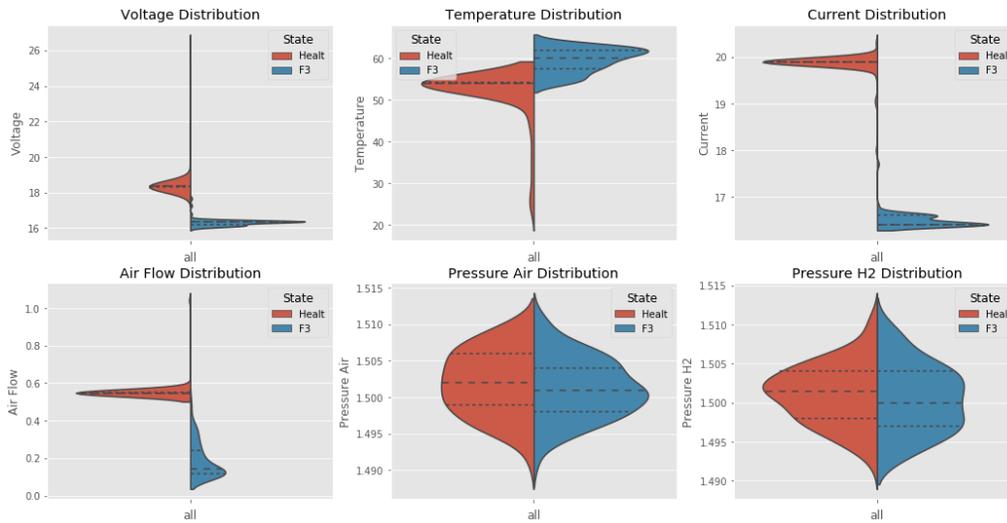


Figura 42 Distribución de valores en el modo de falla F3

Falla en el sistema de suministro de hidrógeno F4

La presión de hidrógeno pasó de 1.5 a 1.2 atm, ver **Figura 43**. El déficit de hidrógeno en el ánodo provoca disminución la reacción de oxidación en el cátodo, lo cual reduce la generación de corriente y disminuye el nivel de voltaje. En este modo de falla la temperatura y la presión de aire no presentan cambios significativos.

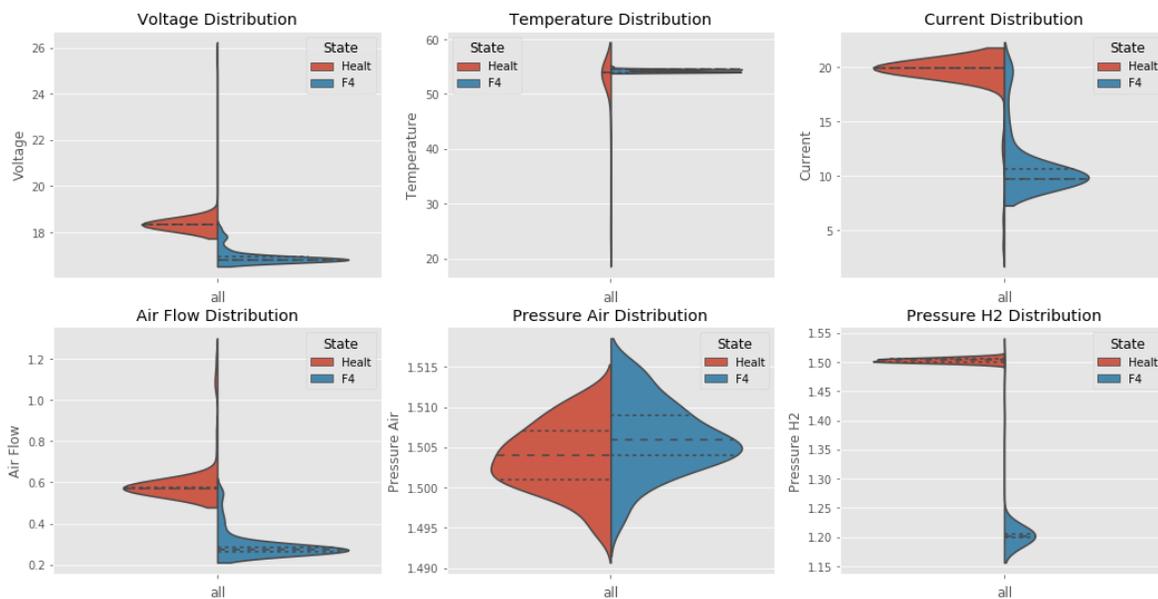


Figura 43 Distribución de valores en el modo de falla F4

4.5 Diagnostico mediante arboles de falla utilizando el algoritmo M3GP

Los árboles de decisión son un método de aprendizaje automático popular para problemas de clasificación y regresión. Dado un vector de características de valor real (por ejemplo, mediciones de sensores), un árbol de decisión utiliza un conjunto de reglas basadas en umbrales para determinar si se ha producido una falla. En este trabajo, las únicas características que se pasan al árbol de decisiones son los valores de sensor sin procesar. Sin embargo, es una práctica común diseñar otras características adaptadas al dominio del problema específico. Las reglas se aprenden durante una fase de entrenamiento en la que los ejemplos etiquetados (las etiquetas le dicen al árbol qué ejemplos corresponden a fallas de sensores) se muestran en el árbol. Una vez entrenadas, las reglas son, en esencia, una serie de declaraciones *if-else* con expresiones condicionales que verifican si las características están por encima o por debajo de diferentes valores de umbral.

El FDI se puede realizar de dos maneras. Generando FTs individuales (por tipo de falla) o generando un FT global, el cual sirve para clasificar cualquier estado de operación. El desempeño global de este algoritmo se puede ver en la matriz de confusión mostrada en la **Figura 44**. La precisión fue de 0.97 con solo 0.0275 de clases mal clasificadas. La falla más difícil de clasificar es F1, ya que durante su transición el efecto en el desempeño de la celda es casi nulo. El árbol de falla generado para clasificar todos los modos de operación de la celda, se puede ver en la **Figura 45**. Este algoritmo genera una población inicial de manera aleatoria, por lo cual se pueden obtener distintos arboles con puntajes de clasificación similares. El proceso se puede repetir hasta encontrar un árbol congruente que ayude a interpretar la interacción de las variables de estado en los distintos modos de falla. El árbol generado sirve para clasificar todas las fallas, sin embargo, se pueden generar arboles individuales por tipo de falla para poder analizar la causa raíz de cada falla.

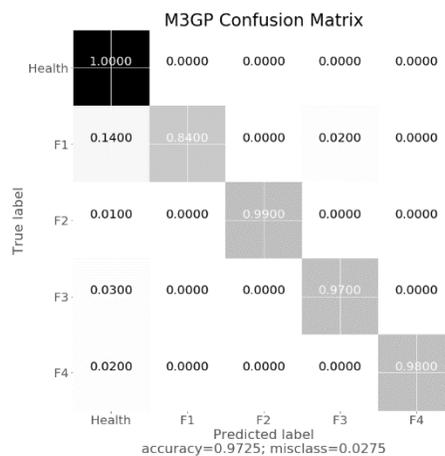


Figura 44 Matriz de confusión M3GP

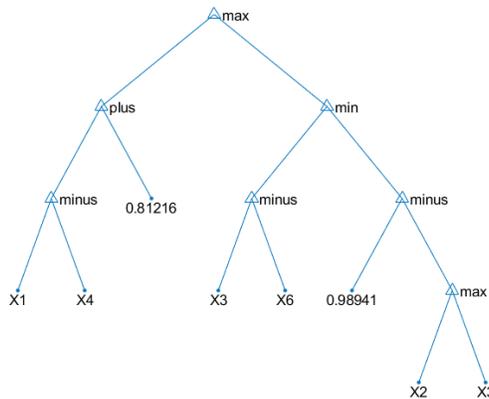


Figura 45 Árbol de falla M3GP

El proceso de entrenamiento solo tomó 3 generaciones, la evolución del proceso se puede observar en la **Figura 46**. El bajo costo computacional permite la implementación de este algoritmo en un sistema embebido para el control y diagnóstico de fallas de manera on-line.

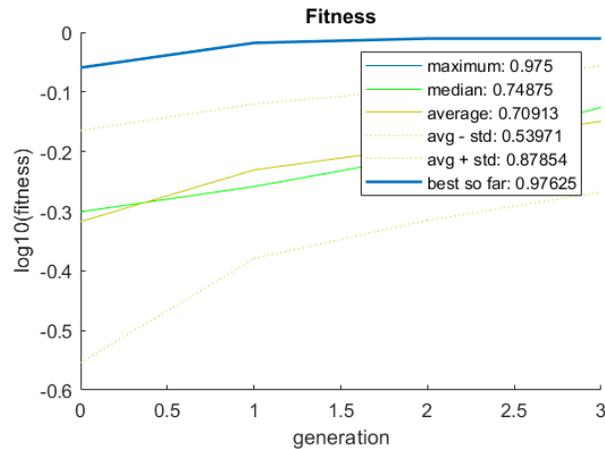


Figura 46 Evolución del algoritmo M3GP

En la **Figura 47** se observa el árbol generado para diagnosticar la falla F1. En el árbol se relaciona mediante la función “my-if” el voltaje con la temperatura de celda, después, el resultado de esa función se compara con el flujo de aire para determinar si el sistema PEMFC tiene problemas en el sistema de enfriamiento, esta relación es congruente con la **Figura 40** en la cual se observa que la distribución de los valores de estas tres variables se concentra en rangos diferentes cuando la celda opera sin problemas de enfriamiento.

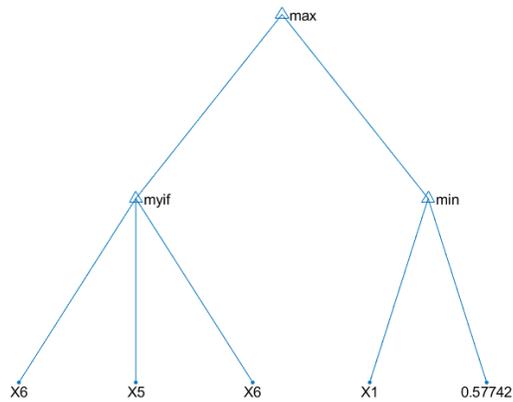


Figura 47 Árbol de falla F1

La **Figura 48** muestra el árbol generado para F2 (falla en el suministro de aire). Este árbol utiliza las variables flujo de aire y presión de aire para el diagnóstico, lo cual es congruente con la distribución de valores mostrada en la **Figura 41**. Estos árboles individuales son muy útiles cuando se presenta un modo de falla desconocido, ya que nos ayudan a entender el comportamiento de la falla y a distinguir su efecto en la operación de la celda. El proceso de entrenamiento solo tomó 2 generaciones y se detuvo debido a que se alcanzó una clasificación con un 100% de precisión.

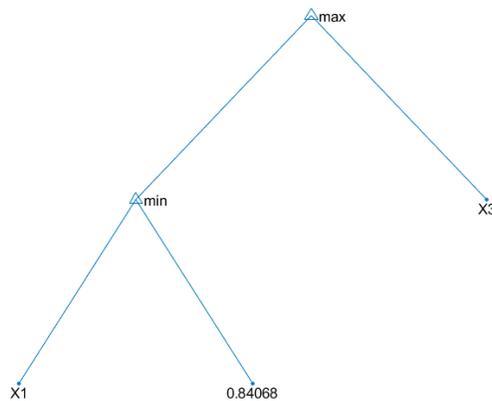


Figura 48 Árbol de falla F2

En la **Figura 49** se muestra el árbol de falla de F3. Este árbol relaciona las variables flujo de aire, corriente y temperatura mediante la función "If-then-else", las cuales son las variables más afectadas en este modo de falla, ver **Figura 42**. El proceso de entrenamiento solo tomó una generación y se detuvo debido a que se alcanzó una clasificación con un 100% de precisión.

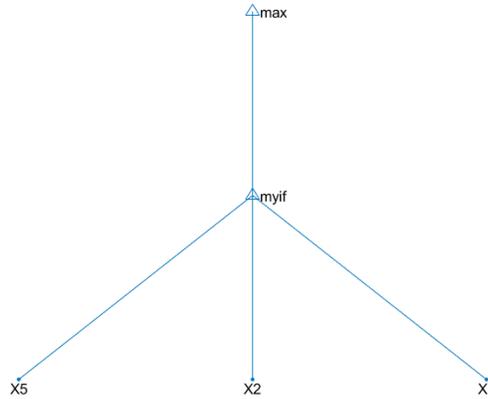


Figura 49 Árbol de falla F3

En la **Figura 50** se muestra el árbol de falla de F4. Este árbol relaciona las variables flujo de aire y presión de hidrogeno con la función “max”, las cuales son las variables que más varían su distribución de valores en este modo de falla, ver **Figura 43**. El proceso de entrenamiento tomo 3 generaciones y alcanzó una clasificación con un 99% de precisión.

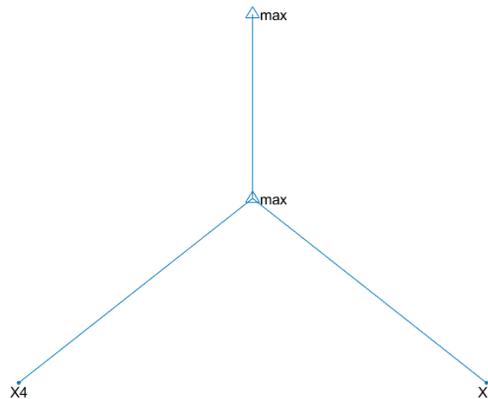


Figura 50 Árbol de falla F4

4.6 Comparación de métodos de diagnostico

En esta sección se van a comparar el desempeño de una red neuronal complementada con *linear discriminant analysis* (LDA) y *principal component análisis* (PCA), con el desempeño del algoritmo M3GP para el diagnóstico de fallas en el sistema PEMFC. Los métodos de extracción de características, LDA y PCA, se han utilizado para redimensionar el espacio original de los datos con el objetivo de mejorar la precisión de la clasificación. Los primeros 3 componentes de cada método de extracción explican más del

90% de correlación de los datos. Estos 3 componentes se utilizaron como vector de entrada para la red neuronal.

4.6.1 Reporte de clasificación

En el caso de diagnóstico de fallas tanto el índice de medición *precision* p como *recall* r son importantes. Intuitivamente, la *precision* es la capacidad del clasificador de no etiquetar como positiva una muestra que es negativa, y *recall* es la capacidad del clasificador de encontrar todas las muestras positivas. La medida F-1 puede interpretarse como una media armónica ponderada de la *precision* y la *recall*. Estas medidas alcanzan su mejor valor en 1 y su peor puntaje en 0. Supongamos que el algoritmo detecta un falso positivo, es decir, diagnostica una falla cuando en realidad el sistema no ha fallado. Si este falso positivo fuera una falla en el sistema de enfriamiento, la acción correctiva solo sería incrementar el flujo de agua, lo cual no impacta en el desempeño de la celda. Sin embargo, si el falso positivo fuera la falla F3 (Incremento del cruce de combustible), la acción correctiva sería parar el sistema para inspeccionar la membrana de la celda y los sellos de pila. Esta acción correctiva detendría la operación del sistema y generaría costos por el reemplazo de piezas en buen estado. La misma analogía aplica en el caso de los falsos negativos, es decir, el algoritmo no diagnostica falla, pero en realidad la celda está fallando, ver **Tabla 11**. Por lo anterior descrito se selecciona como medida de desempeño el índice F-1.

Tabla 11 Sensibilidad y especificidad

<i>Diagnostico</i>	<i>Estado de la celda</i>	
	<i>Falla</i>	<i>Saludable</i>
Positivo	Verdadero positivo	Falso positivo
Negativo	Falso negativo	Verdadero negativo

A continuación, se muestran los reportes de clasificación de cada método evaluado.

4.6.2. Método LDA-NNA para el diagnóstico de fallas

La **Figura 51** muestra el espacio redimensionado de la base de datos original después de realizar la técnica LDA. La dimensión original de los datos era 6-D. La dimensión se redujo a 3-D, debido a que los primeros 3 componentes explicaban más del 90% de la correlación de los datos. Como puede verse en la **Figura 51** la reducción de dimensiones permite aislar fácilmente las fallas F2 y F4, sin embargo, la falla F1 y el estado saludable están traslapados.

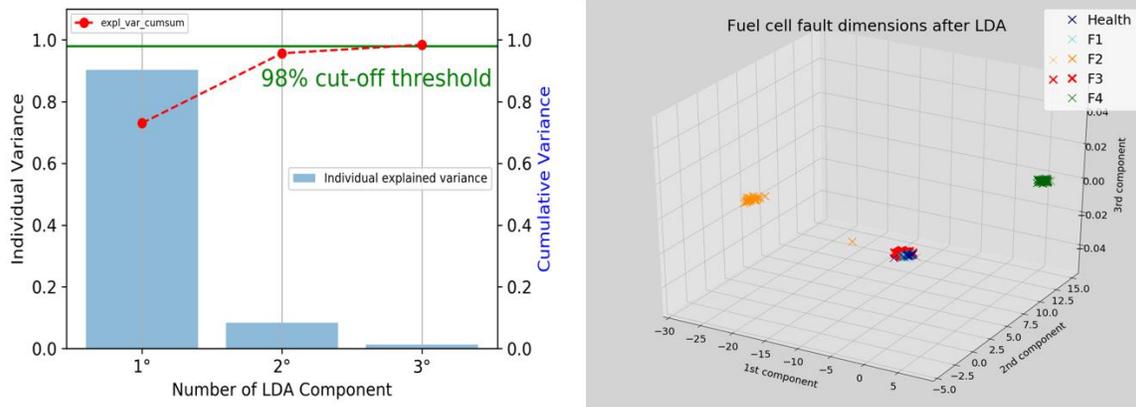


Figura 51 Dimensión del espacio de falla de la celda de combustible después de LDA

La **Figura 52** muestra la matriz de confusión donde se puede observar la dificultad del algoritmo para diferenciar la falla F1 del estado saludable de la celda. Sin embargo, a pesar de las dificultades para identificar la falla F1, el algoritmo presenta un excelente desempeño para diagnosticar las fallas F2, F3 y F4.

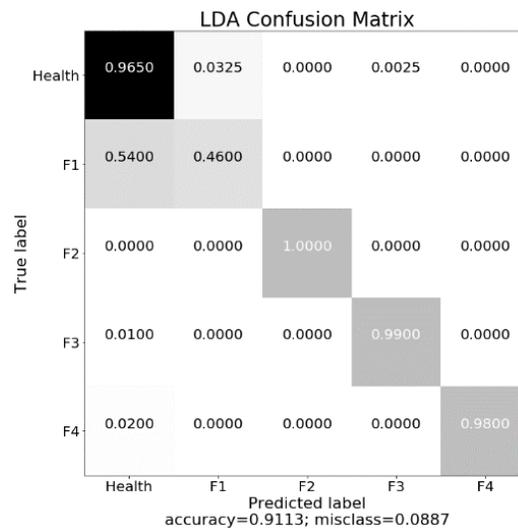


Figura 52 Matriz de confusión LDA-ANN

La **Tabla 12** muestra el reporte de clasificación. La presión de este método es de 0.91 con un *puntaje promedio ponderado* de 0.90 para F-1. Este método no tiene problema para diagnosticar las fallas F2, F3 y F4, sin embargo, tiene muchos problemas para diferenciar el estado saludable de la falla F1.

Tabla 12 Reporte de clasificación LCD-ANN

	Precision	Recall	F1-score	Support
Health	0.87	0.96	0.92	400
F1	0.78	0.46	0.58	100
F2	1	1	1	100
F3	0.99	0.99	0.99	100
F4	1	0.98	0.99	100

Accuracy			0.91	800
Macro avg	0.93	0.88	0.89	800
Weighted avg	0.91	0.91	0.90	800

4.6.3. Método PCA-NNA para el diagnóstico de fallas

La **Figura 53** muestra el espacio redimensionado de la base de datos original después de realizar la técnica PCA. La dimensión original de los datos era 6-D. La dimensión se redujo a 3-D, debido a que los primeros 3 componentes explicaban más del 90% de la correlación de los datos. Como puede verse en la **Figura 53** la reducción de dimensiones permite aislar fácilmente las cuatro fallas. Este método solo tiene problemas menores para diagnosticar las fallas durante la etapa de transición del estado saludable a las fallas F1, F2 y F4.

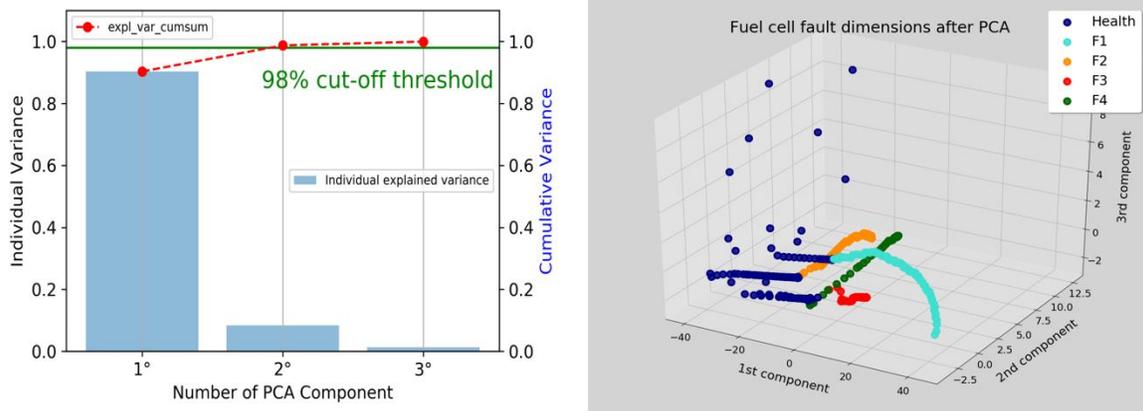


Figura 53 Dimensión del espacio de falla de la celda de combustible después de PCA

La **Figura 54** muestra la matriz de confusión donde se puede observar los problemas menores para diferenciar las fallas F1, F2 y F4 en su etapa inicial. El desempeño del algoritmo identificando los cuatro modos de falla es excelente, sin embargo, una precisión del 0.99 puede sugerir sobre-entrenamiento.

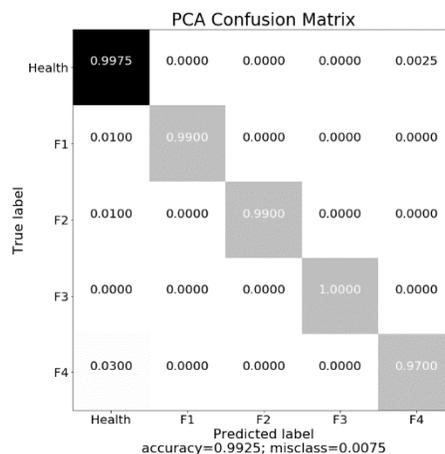


Figura 54 Matriz de confusión PCA-ANN

La **Tabla 13** muestra el reporte de clasificación. La presión de este método es de 0.99 con un *puntaje promedio ponderado* de 0.99 para F-1. Este método presenta de manera global un excelente desempeño para identificar cualquier estado de la celda.

Tabla 13 Reporte de clasificación PCA-ANN

	Precision	Recall	F1-score	Support
Health	0.99	1	0.99	400
F1	1	0.99	0.99	100
F2	1	0.99	0.99	100
F3	1	1	1	100
F4	0.99	0.97	0.98	100

Accuracy			0.99	800
Macro avg	1	0.99	0.99	800
Weighted avg	0.99	0.99	0.99	800

4.6.3. Comparación de desempeño

La principal diferencia del método M3GP respecto a los dos anteriores, es que este método no hace una *extracción de características* para reducir la dimensión de la base de datos sino una *selección de características*, descartando aquellas variables que no tienen valor al momento del diagnóstico. Este algoritmo al igual que los dos anteriores tiene problemas menores al diagnosticar fallas en la etapa de transición, como puede observarse en la matriz de confusión (ver **Figura 44**). La falla más difícil de diagnosticar para este algoritmo es F1. Sin embargo, el promedio ponderado de F-1-score es de 0.97, el cual es un excelente resultado, ver **Tabla 14**.

Tabla 14 Reporte de clasificación M3GP

	Precision	Recall	F1-score	Support
Health	0.95	1	0.98	400
F1	1	0.84	0.91	100
F2	1	0.99	0.99	100
F3	0.98	0.97	0.97	100
F4	1	0.98	0.99	100

Accuracy			0.97	800
Macro avg	0.99	0.96	0.97	800
Weighted avg	0.97	0.97	0.97	800

La **Tabla 15** muestra el resumen de desempeño de los métodos de clasificación evaluados con el índice F1-score. El promedio ponderado de los 3 métodos supera el 90%, sin embargo, el método LDA-ANN tiene muchos problemas para diferenciar F1 del estado saludable. El índice F1-score del algoritmo M3GP para la falla F1 (falla en el sistema de enfriamiento) es de 0.91. Sin embargo, este índice no es malo, ya que los estados mal clasificados son estados de transición. Cuando el incremento en la temperatura es más evidente y el nivel de voltaje empieza a disminuir el algoritmo no tiene problemas en detectar el modo de falla. F1 es la falla más difícil de diagnosticar debido a que el incremento en la temperatura tiene un efecto retardado en el desempeño de la celda, lo cual es congruente con los resultados presentados en [62]. Los tres métodos no presentan dificultad para diagnosticar F2, F3 y F4.

El algoritmo con el desempeño más bajo fue LDA-ANN con un *puntaje promedio ponderado* F1-score de 0.90; en segundo lugar, el algoritmo M3GP con 0.97 y con una diferencia mínima, en primer lugar, el algoritmo PCA-ANN con 0.99.

La principal ventaja de M3GP sobre los dos métodos anteriores, LDA-ANN y PCA-ANN, es que este método no es una caja negra, este método entrega como resultado un gráfico que ayuda a los investigadores a entender el comportamiento de cada falla. Ayuda a aislar el efecto de cada modo de falla en las variables del sistema. Este árbol de falla permite, en algunos casos, corregir la señal de control para reestablecer el estado de operación saludable de la celda. En las fallas transitorias (ver **Tabla 2**), por ejemplo, en el secado de membrana, se puede incrementar el nivel de humidificación de los reactantes, una vez que se diagnostica este modo de falla.

Tabla 15 Comparación de los métodos de clasificación

F1-score	LDA-ANN	PCA-ANN	M3GP
Health	0.92	0.99	0.98
F1	0.58	0.99	0.91
F2	1	0.99	0.99
F3	0.99	1	0.97
F4	0.99	0.98	0.99
Accuracy	0.91	0.99	0.97
Macro avg	0.89	0.99	0.97
Weighted avg	0.90	0.99	0.97

4.6.4. Conclusiones parciales

En esta sección se realizó un diagnóstico de fallas en un sistema PEMFC utilizando el algoritmo de cómputo evolutivo *Programación Genética M3GP*. Esto con el objetivo de generar árboles de falla que ayuden a los investigadores a entender el comportamiento de las fallas, esto es particularmente útil cuando se presenta una falla nueva en el sistema. Estos árboles permiten aislar a las variables afectadas por los distintos tipos de fallas. Para validar el desempeño del diagnóstico y aislamiento de fallas se simularon 4 fallas, incluyendo diferentes componentes de un sistema PEMFC. El desempeño de M3GP fue medido utilizando el índice F1. Además, el bajo costo de computación (memoria y tiempo de computación necesarios) hace que el enfoque sea adecuado como herramienta de diagnóstico en línea. Es importante señalar que el algoritmo M3GP genera una población inicial aleatoria de árboles, esto puede producir varios árboles con un desempeño similar. El proceso debe de repetirse hasta que el árbol generado presente una relación lógica entre las variables. Se pueden generar arboles por cada tipo de falla para analizar su comportamiento individual o se puede generar un árbol general para clasificar todos los modos de falla. En la actualidad, el trabajo de codificar este enfoque en una tarjeta embebida y las pruebas en línea están en proceso. Además, se está generando un algoritmo para la detección de modos de operación de manera no supervisada utilizando el mismo enfoque.

Conclusiones y perspectivas

Para promover la comercialización de celdas de combustible, se debe reducir el costo de operación, así como mejorar la durabilidad y confiabilidad. Por alcanzar esta meta, el objetivo principal de esta tesis es diseñar estrategias de control y de diagnóstico basadas en datos para sistemas PEMFC.

Los principales aportes de este proyecto se pueden resumir de la siguiente manera.

1. Determinar las variables de estado necesarias para modelar y controlar un sistema PEMFC. Esto con el objetivo de reducir el número de sensores instalados y por ende reducir el costo de operación. Esto se realizó comparando distintos métodos de *selección de características* como *métodos de filtrado*, *métodos de envoltura* y *métodos integrados*. Sin embargo, la precisión obtenida con estos métodos no fue adecuada para un modelo orientado al control de la celda de combustible. Por lo anterior se utilizó el análisis PCA para seleccionar las variables que más impactan la operación de la celda.

2. Desarrollar un modelo que permita simular el desempeño de la celda. Una vez determinadas las variables de estado críticas que describen la dinámica de las celdas de combustible, estas se utilizaron como vector de entrada en algunos de los más robustos algoritmos de regresión usados en *machine learning*. El algoritmo de regresión con el mejor desempeño fue ANN.

3. Diseñar un sistema de control que mejore el consumo de combustible. De las ocho variables principales identificadas en el análisis de PCA, solo dos pueden considerarse como entradas del sistema: las presiones de entrada de hidrógeno y oxígeno. Sin embargo, la presión de hidrógeno es la variable que más afecta el rendimiento de la pila de combustible. Las variaciones en la presión de entrada de aire pueden considerarse no representativas para fines de control si se mantienen dentro de un rango específico de operación. Se propuso un esquema de neuro-control en serie PID (con un modelo inverso de la pila de combustible) para suministrar la presión de hidrógeno óptima teniendo en cuenta los valores de las principales variables en condiciones transitorias.

4. Diseñar un árbol de falla que sirva para diagnosticar y aislar fallas. Utilizando principios de selección natural se identificaron las variables afectas según el tipo de falla. Se diseñaron arboles individuales para determinar cada falla, así como un árbol de falla global, capaz de determinar cualquier estado de operación en la celda de combustible. Se utilizaron reglas lógicas para mejorar la interpretabilidad del diagnóstico. Una vez entrenadas, las reglas son, en esencia, una serie de declaraciones *if-else* con expresiones condicionales que verifican si las características están por encima o por debajo de diferentes valores de umbral.

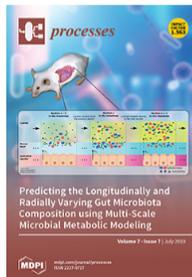
A continuación, se describen las direcciones de investigación relacionadas, que merecen una investigación futura.

1. Desarrollar un proceso de diagnóstico no supervisado capaz de diagnosticar fallas no conocidas online.
2. Obtener un modelo orientado al control basado en datos utilizando regresión simbólica. Esto con el objetivo de utilizar las técnicas de tolerancia a fallos establecidas en la teoría de control de sistemas no lineales.
3. El diagnóstico de fallas se puede combinar con la estrategia de control. Hasta donde sabemos, la investigación de este aspecto acaba de comenzar. Deben realizarse más esfuerzos para el control de tolerancia a fallos para los sistemas PEMFC.

Productos Académicos

Artículo JCR

- Andrés Morán-Durán, Albino Martínez-Sibaja, José Pastor Rodríguez-Jarquín, Rubén Posada-Gómez and Oscar Sandoval González, "PEM Fuel Cell Voltage Neural Control Based on Hydrogen Pressure Regulation", *Processes* 2019, 7(7), 434 <https://doi.org/10.3390/pr7070434>



Processes, Volume 7, Issue 7 (July 2019) – 83 articles

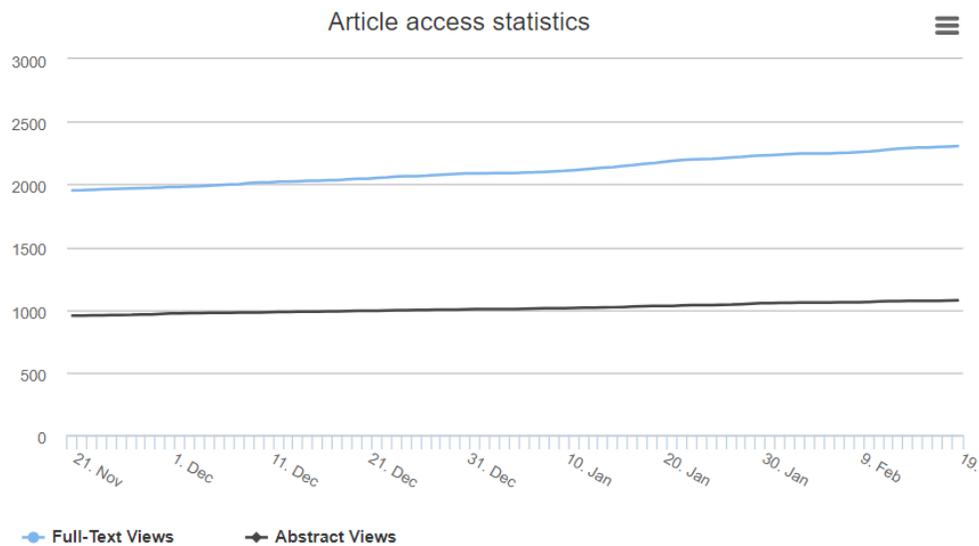
- Issues are regarded as officially published after their release is announced to the **table of contents alert mailing list**.
- You may **sign up for e-mail alerts** to receive table of contents of newly released issues.
- PDF is the official format for papers published in both, html and pdf forms. To view the papers in pdf format, click on the "PDF Full-text" link, and use the free **Adobe Reader** to open them.

Article Metrics

Citations



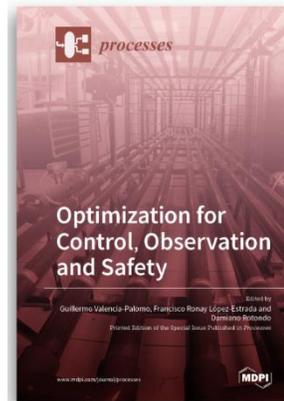
Article Access Statistics



- Se planea enviar el artículo “Diagnóstico de fallas para sistemas de celdas de combustible PEM mediante programación genética” a la “International Journal of Hydrogen Energy” durante este mes de junio.
 - **Estatus: en proceso de revisión de idioma**

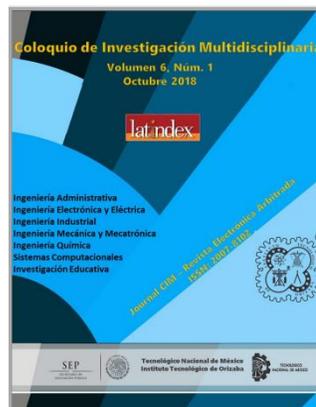
Capítulo de libro (Edición especial)

- **Optimization for Control, Observation and Safety**
 - ISBN 978-3-03928-440-5 (Pbk); ISBN 978-3-03928-441-2 (PDF)
<https://doi.org/10.3390/books978-3-03928-441-2> (registering DOI)



Congresos

- A. Morán Durán, A. Martínez Sibaja, J. Pastor Rodríguez Jarquin, R. Posada Gómez, “Control de Reactantes para celdas de combustible PEM alimentadas con Bio-Hidrógeno”, COLOQUIO DE INVESTIGACIÓN MULTIDISCIPLINARIA JOURNAL CIM, Volumen 6, Numero 1, octubre 2018 ISSN: 2007-8102



Certificaciones

- Se obtuvo la certificación por parte de **Publon's Academy** como revisor por pares para investigadores de carrera temprana, desarrollado junto con académicos y editores expertos.



Colaboraciones

- Se recibió la invitación de la revista *Mathematical and Computer Modelling of Dynamical Systems* para participar como revisor por pares

On the combination of kernel principal component analysis

Mathematical and Computer Modelling of Dynamical Systems <onbehalf@manuscriptcentral.com> para mí

26 dic. 2019 12:32

inglés > español Traducir mensaje Desactivar para: inglés x

26-Dec-2019

Dear Professor Andrés Morán Durán:

Thank you for reviewing the above manuscript, entitled "On the combination of kernel principal component analysis and neural networks for process indirect control" for Mathematical and Computer Modelling of Dynamical Systems.

We greatly appreciate the voluntary contribution that each reviewer gives to the Journal. We hope that we may continue to seek your assistance with the refereeing process for Mathematical and Computer Modelling of Dynamical Systems, and hope also to receive your own research papers that are appropriate to our aims and scope.

In appreciation, you are eligible for a 30% discount on Routledge, CRC Press, and Taylor & Francis books. For details on how to claim please go to www.tandf.co.uk/journals/pdf/books_discount_postcard.pdf

Sincerely Yours,
Dr Troch
Associate Editor, Mathematical and Computer Modelling of Dynamical Systems
inge.troch@tuwien.ac.at

*** Your review has been registered. If you receive any reminders for this paper after this email, please disregard them. ***

Referencias

- [1] U.S. Energy Information Administration, "International Energy Outlook 2016," 2016.
- [2] S. N. A. Rahman, M. S. Masdar, M. I. Rosli, E. H. Majlan, and T. Husaini, "Overview of Biohydrogen Production Technologies and Application in Fuel Cell," *Am. J. Chem.*, vol. 5, pp. 13–23, 2015.
- [3] S. M. Kotay and D. Das, "Biohydrogen as a renewable energy resource – Prospects and potentials," *Int. J. Hydrogen Energy*, vol. 33, pp. 258–263, 2008.
- [4] U. Lucia, "Overview on fuel cells," *Renew. Sustain. Energy Rev.*, vol. 30, no. November, pp. 164–169, 2014.
- [5] "Fuel Cell Technologies Office Multi-Year Research, Development, and Demonstration Plan | Department of Energy." [Online]. Available: <https://www.energy.gov/eere/fuelcells/downloads/fuel-cell-technologies-office-multi-year-research-development-and-22>. [Accessed: 05-Feb-2020].
- [6] S. V. Puranik, A. Keyhani, and F. Khorrami, "State-Space Modeling of proton exchange membrane fuel cell," *IEEE Trans. Energy Convers.*, vol. 25, no. 3, 2010.
- [7] F. Gao, B. Blunier, and A. Miraoui, *Proton Exchange Membrane Fuel Cells Modeling*. Great Britain: Wiley, 2012.
- [8] J. Larminie and A. L. Dicks, *Fuel Cell Systems Explained, 2^o.*, vol. 93, no. 1–2. Chichester,: Wiley, 2000.
- [9] N. Yousfi-Steiner, P. Moçotéguy, D. Candusso, D. Hissel, A. Hernandez, and A. Aslanides, "A review on PEM voltage degradation associated with water management: Impacts, influent factors and characterization," *J. Power Sources*, vol. 183, no. 1, pp. 260–274, 2008.
- [10] J. T. Pukrushpan, A. G. Stefanopoulou, and H. Peng, *Control of fuel cell power systems*, 2nd ed. London: Springer-Verlag London, 2005.
- [11] W. R. W. Daud, R. E. Rosli, E. H. Majlan, S. A. A. Hamid, R. Mohamed, and T. Husaini, "PEM fuel cell system control: A review," *Renew. Energy*, vol. 113, pp. 620–638, 2017.
- [12] C. Ainscough, D. Peterson, and E. Miller, "Hydrogen Production Cost From PEM Electrolysis," 2014.
- [13] R. I. Salim, H. Noura, and A. Fardoun, "A review on fault diagnosis tools of the proton exchange Membrane Fuel Cell," *2013 Conf. Control Fault-Tolerant Syst.*, pp. 686–693, 2013.

- [14] W. Schmittinger and A. Vahidi, "A review of the main parameters influencing long-term performance and durability of PEM fuel cells," *J. Power Sources*, vol. 180, no. 1, pp. 1–14, 2008.
- [15] W. M. Yan, C. Y. Chen, S. C. Mei, C. Y. Soong, and F. Chen, "Effects of operating conditions on cell performance of PEM fuel cells with conventional or interdigitated flow field," *J. Power Sources*, vol. 162, no. 2 SPEC. ISS., pp. 1157–1164, 2006.
- [16] F. Da Costa Lopes, E. H. Watanabe, and L. G. B. Rolim, "A Control-Oriented Model of a PEM Fuel Cell Stack Based on NARX and NOE Neural Networks," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5155–5163, 2015.
- [17] G. Napoli, M. Ferraro, F. Sergi, G. Brunaccini, and V. Antonucci, "Data driven models for a PEM fuel cell stack performance prediction," *Int. J. Hydrogen Energy*, vol. 38, no. 26, pp. 11628–11638, 2013.
- [18] N. S. Sisworahardjo, T. Yalcinoz, M. Y. El-Sharkh, and M. S. Alam, "Neural network model of 100 W portable PEM fuel cell and experimental verification," *Int. J. Hydrogen Energy*, vol. 35, no. 17, pp. 9104–9109, 2010.
- [19] I. S. Han and C. B. Chung, "Performance prediction and analysis of a PEM fuel cell operating on pure oxygen using data-driven models: A comparison of artificial neural network and support vector machine," *Int. J. Hydrogen Energy*, vol. 41, no. 24, pp. 10202–10211, 2016.
- [20] A. Kheirandish, N. Shafiabady, M. Dahari, M. S. Kazemi, and D. Isa, "Modeling of commercial proton exchange membrane fuel cell using support vector machine," *Int. J. Hydrogen Energy*, vol. 41, no. 26, pp. 11351–11358, 2016.
- [21] S. M. Pourkiaei, M. H. Ahmadi, and S. M. Hasheminejad, "Modeling and experimental verification of a 25W fabricated PEM fuel cell by parametric and GMDH-type neural network," *Mech. Ind.*, vol. 17, no. 1, p. 105, 2015.
- [22] A. U. Chávez-Ramírez *et al.*, "High power fuel cell simulator based on artificial neural network," *Int. J. Hydrogen Energy*, vol. 35, no. 21, pp. 12125–12133, 2010.
- [23] R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois, and F. Gao, "Data-driven proton exchange membrane fuel cell degradation predication through deep learning method," *Appl. Energy*, vol. 231, no. September, pp. 102–115, 2018.
- [24] C. Lin-kwong-chon, B. Grondin-pérez, J. A. Kadjo, C. Damour, and M. Benne, "A review of adaptive neural control applied to proton exchange membrane fuel cell systems," *Annu. Rev. Control*, vol. 47, pp. 133–154, 2019.
- [25] S. M. Rakhtala, R. Ghaderi, and A. Ranjbar Noei, "Proton exchange membrane fuel cell voltage-tracking using artificial neural networks," *J. Zhejiang Univ. Sci. C*, vol. 12, no. 4, pp. 338–344, 2011.
- [26] A. Abbaspour, A. Khalilnejad, and Z. Chen, "Robust adaptive neural network

- control for PEM fuel cell," *Int. J. Hydrogen Energy*, vol. 41, no. 44, pp. 20385–20395, 2016.
- [27] M. Aliasghary, "Control of PEM Fuel Cell Systems Using Interval Type-2 Fuzzy PID Approach," *Fuel Cells*, vol. 18, no. 4, pp. 449–456, 2018.
- [28] A. Kheirandish, F. Motlagh, N. Shafiabady, M. Dahari, and A. Khairi Abdul Wahab, "Dynamic fuzzy cognitive network approach for modelling and control of PEM fuel cell for power electric bicycle system," *Appl. Energy*, vol. 202, pp. 20–31, 2017.
- [29] A. Benmouna, M. Becherif, D. Depernet, F. Gustin, H. S. Ramadan, and S. Fukuhara, "Fault diagnosis methods for Proton Exchange Membrane Fuel Cell system," *Int. J. Hydrogen Energy*, pp. 1–10, 2016.
- [30] R. A. de Ingeniería, "conocimiento heurístico | Real Academia de Ingeniería." [Online]. Available: <http://diccionario.raing.es/es/lema/conocimiento-heuristico>. [Accessed: 10-Feb-2020].
- [31] R. J. Patton, "Encyclopedia of Systems and Control," *Encycl. Syst. Control*, no. Isermann 2006, pp. 1–8, 2019.
- [32] A. Scheer, H. Kruppke, and R. Heib, *Springer-Verlag Berlin Heidelberg GmbH*, 1st editio. Berling: Springer-Verlag Berlin Heidelberg GmbH, 2001.
- [33] R. Petrone *et al.*, "A review on model-based diagnosis methodologies for PEMFCs," *Int. J. Hydrogen Energy*, vol. 38, no. 17, pp. 7077–7091, 2013.
- [34] A. Zeller, O. Rallières, J. Régnier, and C. Turpin, "Diagnosis of a hydrogen/air fuel cell by a statistical model-based method," *2010 IEEE Veh. Power Propuls. Conf. VPPC 2010*, 2010.
- [35] A. Hernandez, D. Hissel, and R. Outbib, "Modeling and fault diagnosis of a polymer electrolyte fuel cell using electrical equivalent analysis," *IEEE Trans. Energy Convers.*, vol. 25, no. 1, pp. 148–160, 2010.
- [36] S. De Lira, V. Puig, and J. Quevedo, "LPV model-based fault diagnosis using relative fault sensitivity signature approach in a PEM fuel cell," *IFAC Proc. Vol.*, no. 1, pp. 528–533, 2009.
- [37] S. De Lira, V. Puig, and J. Quevedo, "Robust LPV model-based sensor fault diagnosis and estimation for a PEM fuel cell system," *Conf. Control Fault-Tolerant Syst. SysTol'10 - Final Progr. B. Abstr.*, pp. 819–824, 2010.
- [38] M. Buchholz, M. Eswein, and V. Krebs, "Modelling PEM fuel cell stacks for FDI using linear subspace identification," *Proc. IEEE Int. Conf. Control Appl.*, pp. 341–346, 2008.
- [39] S. S. Araya, F. Zhou, S. L. Sahlin, S. Thomas, C. Jeppesen, and S. K. Kær, "Fault characterization of a proton exchange membrane fuel cell stack," *Energies*, vol. 12,

- no. 1, 2019.
- [40] S. Wasterlain, D. Candusso, F. Harel, D. Hissel, and X. Franois, "Development of new test instruments and protocols for the diagnostic of fuel cell stacks," *J. Power Sources*, vol. 196, no. 12, pp. 5325–5333, 2011.
- [41] H. Wang, X. Z. Yuan, and H. Li, *PEM Fuel Cell Diagnostic Tools*. Boca Raton: CRC Press, 2011.
- [42] J. Hua, J. Li, M. Ouyang, L. Lu, and L. Xu, "Proton exchange membrane fuel cell system diagnosis based on the multivariate statistical method," *Int. J. Hydrogen Energy*, vol. 36, no. 16, pp. 9896–9905, 2011.
- [43] J. R. Koza, *Genetic Programming, On the Programming of Computers by means of Natural Selection*. Massachusetts Institute of Technology, 1992.
- [44] V. Ingalalli, S. Sara, M. Castelli, and L. Vanneschi, "A Multi-dimensional Genetic Programming Approach for Multi-class Classification Problems," in *Genetic Programming. EuroGP 2014. Lecture Notes in Computer Science, vol 8599*, Springer, Berlin, Heidelberg, 2014, pp. 48–60.
- [45] L. Muñoz, S. Silva, and L. Trujillo, "M3GP – Multiclass Classification with GP," in *Genetic Programming*, Machado P. et al., Ed. Springer, Cham, 2015, pp. 78–91.
- [46] M. Oltean and L. Dioşan, "An autonomous GP-based system for regression and classification problems," *Appl. Soft Comput. J.*, vol. 9, no. 1, pp. 49–60, 2009.
- [47] I. De Falco, E. Tarantino, A. Della Cioppa, and F. Fontanella, "An innovative approach to genetic programming-based clustering," *Adv. Soft Comput.*, vol. 34, pp. 55–64, 2006.
- [48] P. G. Espejo, S. Ventura, and F. Herrera, "A Survey on the Application of Genetic Programming to Classification," *IEEE Trans. Syst. Man. Cybern.*, vol. 40, no. 2, pp. 121–144, 2010.
- [49] J. Xuan, H. Jiang, T. Shi, and G. Liao, "Gear Fault Classification Using Genetic Programming and Support Vector Machines," *Int. J. Inf. Technol.*, vol. 11, no. 9, pp. 19–27, 2005.
- [50] P. Chen, M. Taniguchi, T. Toyota, and Z. He, "Fault diagnosis method for machinery in unsteady operating condition by instantaneous power spectrum and genetic programming," *Mech. Syst. Signal Process.*, vol. 19, no. 1, pp. 175–194, 2005.
- [51] H. Guo, L. B. Jack, A. K. Nandi, and S. Member, "Feature Generation Using Genetic Programming With Application to Fault Classification," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 35, no. 1, pp. 89–99, 2005.
- [52] L. Zhang, L. B. Jack, and A. K. Nandi, "Fault detection using genetic programming," *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 271–289, 2005.

- [53] W. X. Yang, "Establishment of the mathematical model for diagnosing the engine valve faults by genetic programming," *J. Sound Vib.*, vol. 293, no. 1–2, pp. 213–226, 2006.
- [54] "IEEE PHM 2014 Data Challenge," 2014. [Online]. Available: <http://eng.fclab.fr/ieee-phm-2014-data-challenge/>. [Accessed: 22-May-2019].
- [55] P. E. M. Almeida and M. G. Simões, "Neural optimal control of PEM fuel cells with parametric CMAC networks," *IEEE Trans. Ind. Appl.*, vol. 41, no. 1, pp. 237–245, 2005.
- [56] L. Placca, R. Kouta, D. Candusso, J. F. Blachot, and W. Charon, "Analysis of PEM fuel cell experimental data using principal component analysis and multi linear regression," *Int. J. Hydrogen Energy*, vol. 35, no. 10, pp. 4582–4591, 2010.
- [57] A. K. Al-Othman, N. A. Ahmed, F. S. Al-Fares, and M. E. AlSharidah, "Parameter Identification of PEM Fuel Cell Using Quantum-Based Optimization Method," *Arab. J. Sci. Eng.*, vol. 40, no. 9, pp. 2619–2628, 2015.
- [58] M. Hatti and M. Tioursi, "Dynamic neural network controller model of PEM fuel cell system," *Int. J. Hydrogen Energy*, vol. 34, no. 11, pp. 5015–5021, 2009.
- [59] S. Omatu, M. Yoshioka, T. Kosaka, and H. Yanagimoto, "Neuro-PID Control of Speed and Torque of Electric Vehicle," *Int. J. Adv. Syst. Meas.*, vol. 3, no. 1, pp. 82–91, 2010.
- [60] R. E. Rosli *et al.*, "Reactant Control System for Proton Exchange Membrane Fuel Cell," *Procedia Eng.*, vol. 148, pp. 615–620, 2016.
- [61] M. Shao, X. J. Zhu, H. F. Cao, and H. F. Shen, "An artificial neural network ensemble method for fault diagnosis of proton exchange membrane fuel cell system," *Energy*, vol. 67, pp. 268–275, 2014.
- [62] Z. Li, R. Outbib, S. Giurgea, D. Hissel, and Y. Li, "Fault detection and isolation for Polymer Electrolyte Membrane Fuel Cell systems by analyzing cell voltage generated space," *Appl. Energy*, vol. 148, pp. 260–272, 2015.