Asociación Argentina
de Mecánica Computacional

Mecánica Computacional Vol XLI, pp. 1065-1074
C.I. Pairetti, M.A. Pucheta, M.A. Storti, C.M. Venier (Eds.)
D. Ramajo, H. Burbridge, D. Godino (Issue eds.)
Rosario, November 5-8, 2024

SOFT SENSOR BASADO EN REDES NEURONALES RECURRENTES: APLICACIÓN AL MONITOREO DE LA PRODUCCIÓN DE CAUCHO NITRILO

SOFT SENSOR BASED ON RECURRENT NEURAL NETWORKS: APPLICATION TO MONITORING OF THE PRODUCTION OF NITRILE RUBBER

Mariano M. Perdomo[a,b], **Luis A. Clementi**[b,c] **y Jorge R. Vega**[a,b]

[a]*Instituto de Desarrollo Tecnológico para la Industria Química, Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional del Litoral, Ruta Nacional 168, Km. 0 – Paraje "El Pozo", 3000 Santa Fe, Argentina, mperdomo@intec.unl.edu.ar, https://intec.conicet.gov.ar*

[b]*Centro de Investigación y Desarrollo en Ingeniería Eléctrica y Sistemas Energéticos, Universidad Tecnológica Nacional – Facultad Regional Santa Fe, Lavaisse 610, 3000 Santa Fe, Argentina, jrvega@frsf.utn.edu.ar, https://www.frsf.utn.edu.ar/ciese*

[c]*Instituto de Investigación y Desarrollo en Bioingeniería y Bioinformática, Consejo Nacional de Investigaciones Científicas y Técnicas - Universidad Nacional de Entre Ríos, Ruta Provincial 11, Km. 10.5, 3100 Oro Verde, Argentina, laclementi@santafe-conicet.gov.ar, https://ibb.conicet.gov.ar*

Palabras clave: Sensor inferencial, proceso por lotes, caucho NBR, aprendizaje profundo.

Resumen. En Argentina, el caucho nitrilo (NBR) se produce a partir de una polimerización en emulsión en un reactor *batch*. La medición de variables de calidad del polímero (por ejemplo, con analizadores en línea o en laboratorio) no asegura un monitoreo adecuado de la reacción. En este trabajo se desarrolla un *soft-sensor* (SS) para estimar en tiempo real algunas variables de calidad del producto. La complejidad reside en las dinámicas altamente no-lineales involucradas en el proceso. Por ello, el SS propuesto utiliza redes neuronales recurrentes. La evaluación de la herramienta de estimación se realiza a través de un simulador numérico del proceso NBR ajustado a la planta industrial. Las estimaciones obtenidas en distintos escenarios de operación del reactor son promisorias. El SS podría ser implementado en la planta industrial en forma sencilla.

Keywords: Inferential sensor, batch process, NBR rubber, deep learning.

Abstract.
In Argentina, nitrile rubber (NBR) is produced in an emulsion polymerization carried out in a batch reactor. Measuring polymer quality variables (e.g., with online analyzers or in laboratory) does not ensure an adequate monitoring of the reaction. In this work, a soft-sensor (SS) is developed to online estimate some quality variables. The complexity lies in the highly non-linear dynamics involved in the process. Therefore, the proposed SS uses recurrent neural networks. The evaluation of the estimation tool is carried out through a numerical simulator of the NBR process adjusted to the industrial plant. The estimates obtained in different reactor operation scenarios are promising. The SS could be implemented in the industrial plant in a simple way.

ISSN: 2591-3522 DOI: 10.70567/mc.v41i20.104

1 INTRODUCCIÓN

El caucho nitrilo (NBR) es de gran interés para aplicaciones específicas donde se requiere de una buena resistencia a los aceites, disolventes e hidrocarburos, baja permeabilidad a gases, y aptitud para trabajar a temperaturas elevadas. El NBR tiene una importante demanda mundial en la industria automotriz y aeronáutica, proyectada en algo más de 1,5 millones de toneladas (Madhuranthakam y Penlidis, 2013; Saldivar-Guerra et al, 2020).

La primera etapa para la producción de caucho NBR consiste en la obtención de un látex sintético por medio de la copolimerización en emulsión de acrilonitrilo (A) y butadieno (B), y es una etapa clave porque quedan definidas varias propiedades del producto final. En las etapas siguientes, se recuperan los monómeros residuales y el agua, y al producto seco resultante se lo enfarda. En Argentina, la planta industrial de Pampa Energía S.A. (Pto. Gral. San Martín, Santa Fe) utiliza un reactor tanque agitado operado en forma discontinua (*batch*) y en condiciones isotérmicas (a 10 °C). El NBR puede producirse en varios grados comerciales (denominados AJLT, BJLT, CJLT, DJLT), diferenciados especialmente por la fracción másica de A ligado al copolímero (p_A). Otra variable importante es la conversión másica de monómeros (x). Ambas son variables de calidad del producto final, y deben ubicarse dentro de rangos preestablecidos. En el reactor industrial se miden, en tiempo real, únicamente la temperatura de la emulsión (T) y el caudal del flujo refrigerante de salida del reactor (F_R). Las variables de calidad x y p_A se miden en laboratorio, con la desventaja de los considerables retardos de estimación y elevados tiempos de muestreo. Disponer de estimaciones en tiempo real de x y p_A permite mejorar la supervisión del avance de la reacción, y de implementar estrategias de lazo cerrado con adiciones continuas de reactivos (operación semicontinua del reactor), con el objetivo de regular las variables de calidad. Un ejemplo es la regulación de p_A por adiciones de A (Clementi et al., 2018). Por estas razones, es de interés disponer de un estimador en línea (*soft sensor*: SS) para este proceso industrial.

Un SS es una herramienta computacional diseñada para estimar variables de proceso de difícil medición en tiempo real (Perera et al., 2023). Son de particular relevancia para la estimación de variables de calidad en procesos de la industria química, ya sean continuos, semicontinuos o discontinuos, permitiendo una mejor supervisión y control del proceso, la optimización de la producción y el aseguramiento de la calidad del producto final. Los SS emplean modelos matemáticos y algoritmos basados en datos obtenidos de sensores, datos históricos de producción y/o de otras fuentes de información del proceso. En la industria, las mediciones son aportadas por los medidores más usuales: de presión, de temperaturas y de caudales. En el caso de procesos de polimerización se han reportado resultados recientemente (Perdomo et al., 2024, Sanseverinatti et al., 2023). La necesidad de capturar relaciones no lineales en los conjuntos de datos torna atractivo el uso de las redes neuronales artificiales (ANN) para diseñar un SS. Además, las ANN poseen flexibilidad para incorporar técnicas que mejoren su capacidad de generalización, robustez ante ruidos y adaptabilidad. En aplicaciones donde se desea modelar procesos dinámicos complejos, las redes neuronales recurrentes (*recurrent neural networks*: RNN) son de interés por su principio de operación (Perera et al., 2023).

En este contexto, en el presente trabajo se implementan técnicas de estimación en tiempo real de x y p_A a partir de otras mediciones típicas del proceso NBR. Las estrategias de estimación utilizadas se basan en RNNs. Se estudia el impacto de la consideración de metodologías de regularización para efectuar estimaciones de las variables de calidad, incluso ante la presencia de distintos tipos posibles de perturbaciones en la planta. La base de datos utilizada para su entrenamiento es generada a partir de un modelo matemático detallado disponible (Vega et al., 1997). El estudio se extiende a dos grados comerciales del caucho

NBR con diferentes valores de p_A al final de la reacción: i) AJLT ($p_A \cong 20\%$); y ii) BJLT ($p_A \cong 32{,}5\%$). En ambos casos, p_A decrece a lo largo de la reacción, y entonces los algoritmos de estimación son útiles para su monitoreo y posterior control.

2 CONCEPTOS PRELIMINARES

2.1 Redes neuronales recurrentes

Las RNN están diseñadas específicamente para procesar datos secuenciales. Estas redes poseen conexiones recurrentes para permitir la persistencia de la información, conformando un tipo de memoria. Esta propiedad las torna adecuadas para tareas donde el contexto temporal y el orden de los datos son cruciales, como en el procesamiento de series temporales. En una RNN, las salidas de una capa se retroalimentan como entradas a la misma capa en el siguiente paso temporal, permitiendo que la red tenga en cuenta tanto la entrada actual como la información procesada de entradas anteriores. Sin embargo, las RNN estándar pueden enfrentar problemas de aprendizaje a largo plazo, por lo cual se han desarrollado variantes como las redes LSTM (*Long Short-Term Memory*) y GRU (*Gated Recurrent Unit*) para mitigar estos problemas y mejorar su desempeño (Goodfellow et al., 2016). Este tipo de redes neuronales, al ser empleadas con una estructura *many-to-one* permiten aprovechar grandes cantidades de datos no etiquetados, algo que usualmente ocurre en el ámbito industrial, donde las mediciones de sensores industriales tienen una frecuencia de muestreo mucho mayor que la correspondiente a las mediciones de variables de calidad en laboratorios.

2.2 Métodos de regularización

Al entrenar una ANN, se busca un modelo que sea capaz de generalizar sus predicciones ante nuevos datos no presentados durante el entrenamiento. En este contexto surgen los métodos de regularización, los cuales son utilizados para aumentar la capacidad de generalización de los modelos (Bishop and Bishop, 2023). Los métodos de regularización son estrategias diseñadas para reducir el error de testeo, posiblemente a expensas de un incremento del error de entrenamiento. De esta manera, se evita el sobreajuste y se mejora la capacidad del modelo para realizar predicciones precisas en datos desconocidos. En el contexto del aprendizaje profundo existen varias técnicas de regularización. A continuación, se mencionan y definen brevemente aquellas utilizadas en el presente trabajo.

- *Weight Decay*. Incorpora un término de penalización en la función de error para reducir los valores de los pesos resultantes del entrenamiento. Así, se sesga al modelo para reproducir respuestas más suaves. El término regularizador se compone por la suma del cuadrado de los parámetros del modelo.
- *Early Stopping*. Analiza la curva de aprendizaje y selecciona el conjunto de parámetros del modelo que asegura el menor error de validación. Para ello, se fija el hiperparámetro "paciencia"; es decir, el número de épocas sin disminución del error de validación que esperará el algoritmo de *early stopping* antes de detener el proceso de optimización.
- *Dataset Augmentation* (DA). Considera que un modelo de aprendizaje automático tendrá mejor capacidad de generalización al ser entrenado con una mayor cantidad de datos. Además, como la cantidad de datos es limitada y en muchos casos escasa, ese déficit se compensa con la creación de datos falsos a incorporar en el conjunto de entrenamiento. La dificultad para crear instancias de entrenamiento dependerá del dominio de aplicación. Una de las técnicas para hacer DA es la inyección de ruido en los datos de entrada, teniendo en cuenta que muchas tareas pueden ser resueltas incluso con un ruido en las

entradas, pudiendo mejorar así la robustez de la red ante ruidos. Por esas razones, esa técnica suele denominarse *noise robustness*.

- *Model Averaging*. Propone promediar las estimaciones de diferentes modelos entrenados para resolver la misma tarea. Se espera que las contribuciones de los términos de variancia de todos los modelos tiendan a cancelarse, conduciendo así a mejores predicciones. Para esto, los modelos son entrenado con alguna técnica que asegure la existencia de una variabilidad entre sus predicciones (por ejemplo, *bagging* o *boosting*).
- *Dropout*. Se considera una forma implícita de aproximar un *model averaging* de una gran cantidad de modelos. Presenta la ventaja de no tener que entrenar individualmente múltiples modelos y tampoco evaluar las predicciones de todos los modelos considerados. Consiste en eliminar nodos de la red de manera aleatoria durante el entrenamiento. Así, para cada nueva instancia presentada a la red, se efectúa una nueva eliminación aleatoria de los nodos de la red. Esta modificación durante el entrenamiento permite afrontar problemas de nodos sobre-especializados y de co-adaptación.

3 METODOLOGÍA

3.1 Proceso NBR

Se toma como referencia el proceso industrial NBR desarrollado en Argentina por la empresa Pampa Energía, situada en Puerto General San Martin (provincia de Santa Fe). La copolimerización en emulsión de los comonómeros A y B se lleva a cabo en un reactor tanque agitado de unos 18.000 litros, operado en forma discontinua (o *batch*) e isotérmica. Inicialmente, el reactor se carga con una receta específica, conformada por los ingredientes: A, B, iniciador (I), emulsificante (E), modificador (X) y agua (W) (Vega et al., 1997). Las recetas dependen del grado de caucho que se pretende producir. La producción de un lote usualmente demora entre 8 y 10 horas.

Para lograr un caucho con una calidad adecuada, se debe asegurar que determinadas variables de calidad se ubiquen dentro de un rango preestablecido. A los efectos de limitar el número de ramificaciones en las cadenas poliméricas, se procede a: *i*) regular la temperatura de la emulsión en ∼10 °C utilizando un sistema de refrigeración que manipula el caudal de flujo refrigerante en el reactor, y *ii*) detener la reacción cuando x alcanza valores del 75~80%. El grado comercial de NBR queda prácticamente determinado por el valor final de p_A, requiriéndose idealmente que dicha variable se mantenga uniforme a lo largo de la reacción, para lograr un producto final homogéneo. Dicha uniformidad puede lograrse mediante la operación semi-continua del reactor por adición de caudales de alimentación de A o B, según el grado comercial de caucho NBR que se pretenda producir.

3.2 Base de datos

Actualmente, no se dispone de un conjunto de datos experimentales directos del proceso de síntesis del látex; pero, se cuenta con un modelo matemático de la reacción, basado en primeros principios, el cual se ha ajustado a valores de la planta (Vega et al., 1997), permitiendo simular su comportamiento de manera realista. Se considera, además, el sistema de refrigeración del reactor encargado de controlar la temperatura de la emulsión (Sangoi, 2021). A partir de dicho modelo, se simuló el proceso de copolimerización en diferentes condiciones, y se registraron las variables de interés, emulando así el proceso de muestreo y obtención de datos experimentales. Esta aproximación permitió obtener un conjunto de datos representativo para entrenar y desarrollar modelos predictivos, a pesar de la ausencia de datos de proceso reales. En particular, se estudian los grados industriales de mayor interés comercial

del caucho NBR (AJLT y BJLT), y se realizan simulaciones utilizando diferentes recetas de carga del reactor y registrando las variables de interés. Es importante mencionar que la utilización del modelo del proceso es independiente de la metodología propuesta para el desarrollo del SS. En caso de poseer una base de datos obtenida directamente de los sensores del proceso, se puede prescindir completamente del modelo.

La base de datos para entrenar, validar y testear las RNNs a implementar, emula diferentes estados de operación del proceso. Cada estado de operación se caracteriza principalmente por el grado comercial de NBR a sintetizar. Las recetas base utilizadas para producir un lote de producto AJLT y uno de BJLT en el reactor de la planta se muestran en la Tabla 1. A partir de dichas recetas, se simularon variaciones en las cantidades de los ingredientes de carga, con el objetivo de representar variaciones que ocurren típicamente en el proceso, las cuales son de difícil detección (se producen, por ejemplo, por errores en el pesaje de los ingredientes y/o por variaciones en las purezas de los reactivos). Se consideraron 16 variaciones equiespaciadas entre +2% y -2% para los pesos de A, B y X, y entre +10% y -10% para los pesos de I y E. Estas variaciones se simulan secuencialmente para cada ingrediente por separado, conformando 162 simulaciones en total. Cada lote tiene una duración de 600 minutos.

Grado	A	B	E	I	X	W
AJLT	937,3	4999	248,4	0,273	22,60	10104
BJLT	1783	4284	235,7	0,250	22,18	11064

Tabla 1: Recetas en kg para producción de un lote de grados AJLT y BJLT.

Con respecto a las variables de proceso registradas, se asume la disponibilidad de sensores industriales de T y F_R. Para estos sensores, se considera un período de muestreo de un minuto. Además, las variables registradas se contaminaron con ruido aleatorio Gaussiano de media nula y desvío estándar del 1% con relación a la media de cada variable, permitiendo conformar una base de datos representativa de un caso de estudio cercano a la realidad, donde los valores reales de las variables exhiben generalmente incertidumbres de medición. Por otra parte, con respecto a las variables de calidad del látex, x y p_A, se considera que se toman muestras de laboratorios cada 30 minutos, disponiéndose así de 20 muestras por cada lote. Los errores de estimación en laboratorio se introdujeron considerado para p_A la adición de ruido Gaussiano de media nula y desviación estándar tal que las estimaciones pueden tener un error de $\pm$2% (en términos absolutos) con respecto al valor real. Por otra parte, a x se le adicionó un ruido Gaussiano de media nula y desvío estándar dependiente del valor real obtenido en la simulación. La desviación estándar se asume de variación lineal con x, considerando: i) para x = 30%, las estimaciones tienen errores de $\pm$2% en términos absolutos, y ii) para x = 70%, las estimaciones tienen errores de $\pm$0,5% en términos absolutos. Además, se fijan los límites de ambas variables en base a sus definiciones de manera que siempre estén comprendidas con valores entre 0 y 100%. Luego, en el caso de los ingredientes, se consideran los valores presentados en la tabla 1 como entradas para los modelos de estimación. Finalmente, del total de la base de datos generada, se seleccionó aleatoriamente el 50%, 25% y 25% de los registros para constituir los conjuntos de datos de entrenamiento, validación y testeo, respectivamente. Es decir que se cuenta con 82, 40 y 40 registros de lotes en estado normal para los conjuntos de entrenamiento, validación y testeo, respectivamente.

3.3 El SS propuesto

Se propone implementar el SS mediante una RNN, dado el potencial de este tipo de redes para capturar información del pasado del proceso. Además, se entrenan las RNNs con técnicas

orientadas a incrementar la capacidad de generalización.

Se selecciona la arquitectura GRU (Cho et al., 2014) por sus ventajas con respecto a otras arquitecturas de RNN, como las redes de Elman y LSTM. Se evalúa el desempeño de las RNNs considerando: el número de capas ocultas comprendidas en {2, 4, 6}; la dimensionalidad de sus capas ocultas en {50, 100, 1000}; y magnitudes de *learning rate* inicial de {0.0001, 0.00001, 0.000001}. Además, se selecciona una ventana temporal de la secuencia de entrada a la red GRU de 30 minutos, un tamaño *batch* de 32, un *Early Stopping* con una paciencia de 100 *epochs,* se utiliza el método Adam para actualizar los parámetros durante el proceso de optimización, y se define el error medio cuadrático (MSE) como función de pérdida. Los datos son estandarizados de manera que conjunto total de datos de entrenamiento tenga media nula y desviación estándar de 1.

Por otro lado, se pretende que el SS tenga capacidad de generalización, de forma que permita estimar adecuadamente las variables de calidad para estados de operación no presentados en los datos de entrenamiento. Las situaciones más complejas se presentan en general durante la operación del proceso en presencia de anomalías o errores que producen dinámicas que difieren sustancialmente de las operaciones en estado normal. En consecuencia, se analizan distintos métodos de regularización, implementados con el objetivo de otorgar robustez a las estimaciones obtenidas en condiciones de operación atípicas del proceso. En particular, se estudian las siguientes 10 alternativas de regularización:

- Caso 1. Se utiliza para establecer una línea base de comparación. Consiste en una red GRU cuyo método de regularización empleado es el *early stopping*, con una paciencia de 100 *epochs*. El resto de los casos analizados contienen a este caso, y contemplan técnicas adicionales de regularización.
- Caso 2. Se usa *weight decay* con un valor de *lambda* de 0.000001.
- Caso 3. Se usa *dropout* con p=0.2 para la capa de entrada y p= 0.5 para las capas ocultas.
- Caso 4. Se usa DA- Se crean nuevos datos a partir de los disponibles en la base de datos. Las modificaciones se efectúan en una de las variables de entrada y comienza a aplicarse desde un tiempo comprendido entre los 0 y 500 minutos del lote, seleccionado aleatoriamente. Los datos creados son:
  - *Linear drifts*. Para su creación, se selecciona una pendiente de la deriva como un porcentaje del valor de la variable en el momento que inicia la modificación. Se utilizan valores porcentuales de 0.05%, 0.10% y 0.20% para entrenamiento, validación y testeo, respectivamente. Entonces, desde ese instante en adelante, se les suma a los datos iniciales una recta con una pendiente relativa al valor desde el cual inicia la modificación y al valor porcentual que se seleccionó. A partir de los datos de cada lote de producción se crean dos nuevos datos de lotes con *linear drifts* (1 para cada sensor de entrada). Entonces se tienen 164, 80 y 80 datos de lotes con *linear drift* para los conjuntos de entrenamiento, validación y testeo, respectivamente.
  - *Off-sets* constantes. Para su creación, se selecciona un *off-set* a aplicar como un porcentaje del valor de la variable en el momento que inicia la modificación. Se utilizan valores porcentuales de 10%, 15% y 20% para entrenamiento, validación y testeo, respectivamente. Entonces, desde ese instante en adelante, se les suma a los datos iniciales una recta horizontal con una magnitud igual al off-set relativo al valor desde el cual inicia la modificación y al valor porcentual que se seleccionó. A partir de los datos de cada lote de producción se crean cuatro nuevos datos de lotes con *off-sets* (2 para cada sensor de entrada). Entonces, se tienen 328, 160 y 160 datos de lotes con *off-sets* para los conjuntos de entrenamiento, validación y testeo, respectivamente.
  - Datos ruidosos. Consiste en sumar ruido Gaussiano de media nula y desviación

estándar igual a un porcentaje de la desviación estándar de los datos iniciales del lote de producción. Para esto se debe definir previamente el porcentaje de la desviación estándar a utilizar. Para F_R, se utilizan valores porcentuales de 20%, 40% y 80% para entrenamiento, validación y testeo, respectivamente. Para T, se utilizan valores porcentuales de 100%, 200% y 400% para entrenamiento, validación y testeo, respectivamente. Entonces, desde ese instante en adelante, se les suma a los datos iniciales el ruido Gaussiano de media nula y desviación estándar relativa a la desviación original de los datos del lote. A partir de los datos de cada lote de producción se crean dos nuevos datos de lotes con ruido (1 para cada sensor de entrada). Entonces se tienen 164, 80 y 80 datos de lotes con *off-sets* para los conjuntos de entrenamiento, validación y testeo, respectivamente.

Los datos obtenidos con DA a partir de los conjuntos iniciales de entrenamiento y validación se agregan a los efectos de poder considerar estos nuevos datos creados en el proceso de entrenamiento y validación. Luego, los datos obtenidos con DA a partir del conjunto de datos iniciales de testeo se incorporan en este último para comparar a todos los modelos (entrenados con o sin DA).

- Caso 5. Se usa *noise robustness*. Esta técnica se implementa de manera que, a cada dato presentado a la red durante el entrenamiento, antes de iniciar la pasada *forward*, se le suma un ruido aleatorio con una distribución Gaussiana de media nula y desviación estándar de entre 10% y 30% para F_R y de entre 50% y 150% para T, con respecto a la desviación estándar del *batch* de series temporales pasado a la red. Como se observa, durante las distintas épocas, para los datos de un mismo lote de producción, la red verá distintos valores de las entradas afectadas por la suma del ruido. Se pretende de esta manera generar robustez con respecto a la adición de ruido en los sensores de temperatura y fluido refrigerante. A cada dato presentado a la red durante la validación, antes de iniciar la pasada *forward*, se le suma un ruido aleatorio con una distribución Gaussiana de media nula y desviación estándar de entre 30% y 50% para F_R y de entre 150% y 250% para T con respecto a la desviación estándar del *batch* de series temporales pasado a la red.
- Caso 6. Se considera la combinación de los métodos de regularización utilizados para los casos previos. Se usa *weight decay*, *dropout*, DA y *noise robustness* con las mismas configuraciones descriptas en los casos previos.
- Caso 7. Se usa *dropout*, DA y *noise robustness* con las mismas configuraciones descriptas en los casos previos.
- Caso 8. Se usa *dropout* y *noise robustness* con las mismas configuraciones descriptas en los casos previos.
- Caso 9. Se usa *dropout* y DA con las mismas configuraciones descriptas en los casos previos.
- Caso 10. Se usa DA y *noise robustness* con las mismas configuraciones descriptas en los casos previos.

Luego de obtener el conjunto de modelos en base a las configuraciones descriptas previamente, se selecciona el mejor modelo de cada caso en base a su error de estimación para el conjunto de datos de validación. Dicho conjunto se conforma, en esta etapa, con los datos de validación de la base de datos disponible junto con los datos de validación creados con la metodología descripta de DA. Con esa estrategia, se pretende seleccionar aquellos modelos que se comporten aceptablemente en casos de operación normal y además donde se presenten fenómenos atípicos. En esta etapa se selecciona como modelo final a aquel caso que represente los menores errores de estimación.

Finalmente se evalúan los mejores modelos de cada configuración en el conjunto de datos de testeo. En esta etapa, dichos datos se conforman con los datos iniciales de testeo de la base de datos disponible, por aplicar las técnicas de DA en el conjunto previamente mencionado y un nuevo subconjunto de datos "de fallas". El nuevo subconjunto pretende representar casos de alta exigencia para el SS donde las entradas son significativamente distintas a las utilizadas durante las etapas de entrenamiento y validación. Así, este subconjunto representa algunos casos poco probables y de alta exigencia en el proceso. Estos datos de fallas normalmente no se disponen durante la etapa de desarrollo de un modelo por los altos esfuerzos orientados a que esos escenarios no se produzcan. Por estas razones, para emular un escenario de estas características, se crean datos falsos de fallas que representen una operación de naturaleza distinta a la utilizada en las etapas de entrenamiento y validación. A continuación, se detallan las fallas creadas. Las modificaciones se efectúan a partir de un tiempo comprendido entre los 0 y 500 minutos del lote, seleccionado aleatoriamente.

- Caída de sensores.: Se considera que las mediciones de T o F_R pasan a valer 0 mientras el lote se sigue desarrollando normalmente.
- Ruido excesivo en ambos sensores. Se considera la introducción de un ruido de medición en ambos sensores al mismo tiempo. Se usa la metodología descripta para crear datos ruidosos, pero esta vez se aplica en ambas entradas. Se usan valores porcentuales de entre 20% y 80% para el flujo refrigerante y de entre 100% y 400% para la temperatura.

4 RESULTADOS Y DISCUSIONES

En la Tabla 2 se presenta el RMSE total del conjunto de datos de validación para el mejor conjunto de hiperparámetros en cada caso. Se observa que los casos 10 y 9 presentan los menores valores de RMSE, seguidos por los casos 7 y 6. En base a esos resultados, se selecciona el modelo del caso 9 como la mejor opción. Esta decisión contempla la mayor diversidad de representación que puede lograr el caso 9 al hacer uso de *dropout* y DA en comparación con el caso 10, el cual hace uso de DA *y noise robustness*. Esto es así, puesto que *noise robustness* tiene similitudes con los datos ruidosos del DA. Los mejores hiperparámetros del modelo obtenido para el caso 9 son: un número de capas ocultas de 4, una dimensionalidad de sus capas ocultas de 50, y una taza de aprendizaje inicial de 0.0001.

Caso	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
RMSE	0.135	0.132	0.133	0.039	0.137	0.037	0.037	0.137	0.036	0.035

Tabla 2: RMSE en datos de validación.

Para comparar los resultados del modelo seleccionado en base al análisis previo con otras alternativas, en la Tabla 3 se presentan los valores de RMSE para los 10 casos en los datos de testeo, diferenciando entre el total de los datos, los datos normales, los datos creados con técnicas de DA y los datos de fallas. En la Tabla 3, se observa que el caso 9 tiene el segundo mejor valor de RMSE en el total de los datos de testeo (luego del caso 4). Con respecto al análisis de las clasificaciones de los diferentes tipos de lotes de producción, con distintas condiciones de operación, el caso 9 seleccionado se sitúa entre las mejores alternativas presentadas. En general se observa que el uso de métodos de regularización mejora los resultados del caso base, verificándose entonces su utilidad para aplicaciones de este tipo.

La Figura 1 corresponde a 4 casos de ejemplo para distintos tipos de lotes de testeo. Las gráficas muestran las estimaciones del modelo seleccionado (caso 9). A efectos de comparación, se muestran además las estimaciones del caso base (caso 1), y de los tres

mejores casos obtenidos restantes (Tabla 2): casos 10, 7 y 6. Los casos de ejemplo a, b, c y d corresponden a: un *linear drift* en el sensor de T, un *off-set* en el sensor de F_R, una caída del sensor de F_R, y presencia de ruido en ambos sensores en simultáneo, respectivamente. Por razones de espacio, no se muestra una mayor diversidad de casos de ejemplo. Como se observa, el caso base (caso 1) tiene, en general, un menor desempeño, siendo incapaz de afrontar eventos exigentes durante la producción del lote. Por otra parte, se observa que, ante los distintos tipos de fallas presentadas, los casos que utilizan métodos de regularización superan al caso base. Si bien se observa un comportamiento similar en algunos casos de ejemplo, el caso 9 seleccionado se ubica dentro de las mejores opciones de SS, específicamente ante caídas de los sensores. Este es un tipo de fenómeno no observado durante el entrenamiento, demostrando las bondades del uso de *dropout* para abordar situaciones que pueden haber pasado desapercibidas durante el diseño de los SS. Aun así, en el ejemplo c) ningún modelo puede estimar correctamente a p_A. Solo la primera estimación una vez ocurrida la falla de 3 modelos es aceptable. Esto puede deberse a que la ventana temporal seleccionada no permite analizar la relación dinámica entre las variables de entrada en pasos pasados de importancia para identificar que se trata de una falla del sensor y no de un cambio real de la variable.

Caso	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Total	0.154	0.143	0.156	0.073	0.159	0.097	0.093	0.166	0.082	0.090
Normal	0.074	0.076	0.069	0.019	0.057	0.025	0.019	0.064	0.017	0.021
DA	0.148	0.143	0.149	0.062	0.155	0.053	0.054	0.158	0.063	0.053
Fallas	0.188	0.161	0.193	0.109	0.191	0.181	0.169	0.206	0.120	0.166

Tabla 3: RMSE en datos de testeo.

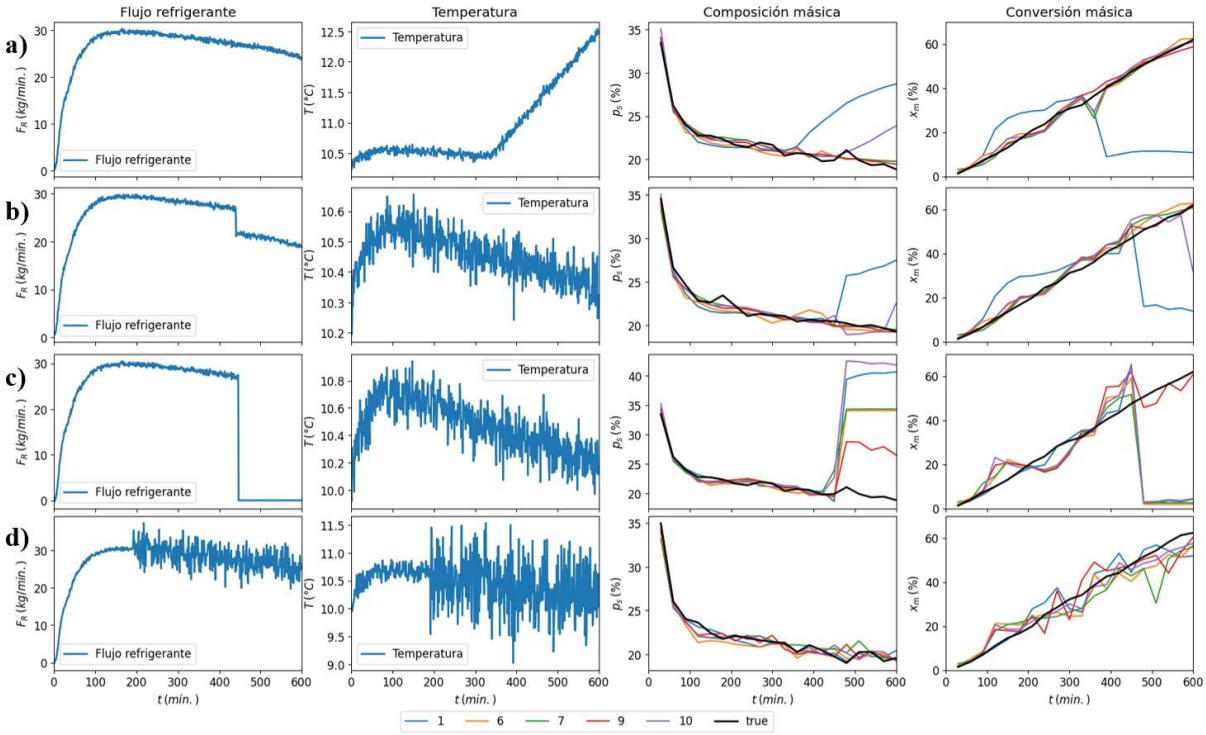

Figura 1: Casos de ejemplo de lotes de testeo. a) linear drift en el sensor de T, b) off-set en el sensor de F_R, c) caída del sensor de F_R, d) ruido simultáneo en ambos sensores.

5 CONCLUSIONES

En el presente trabajo se comprobó la factibilidad de desarrollar un SS para estimar variables de calidad de importancia en el proceso de producción de látex para el caucho NBR, tipo AJLT o BJLT. El SS desarrollado hace uso de RNN, las cuales son aptas para procesar la naturaleza dinámica y no lineal del proceso abordado. Además, con este tipo de arquitectura se aprovechan en mayor medida aquellos datos recolectados del proceso que no se encuentran etiquetados, es decir, que no cuentan con valores numéricos de variables de calidad asociados. Además, en el desarrollo del SS se evaluó el uso de los siguientes métodos de regularización: *weight decay*, *data augmentation*, *dropout* y *noise robustness*, los cuales permiten efectuar estimaciones robustas aun ante la presencia de determinadas fallas exigentes del proceso. La principal ventaja del enfoque utilizado es que no se requiere tener datos del proceso en estados de falla para poder efectuar estimaciones robustas ante la presencia de éstas. Lo cual, sería inviable, dado que por razones técnicas y económicas los procesos no podrían operarse en esas condiciones. Se destacan principalmente las ventajas aportadas por los métodos de regularización *data augmentation* (que puede incluir a *noise robustness*) y *dropout*. Esto muestra la importancia de recrear situaciones exigentes de estimación de las variables de calidad durante el proceso de aprendizaje para obtener SS con mayor robustez.

REFERENCIAS

Bishop, C. M., and Bishop, H., Deep learning: Foundations and concepts. *Springer Nature*, 2023.

Cho, K., Van Merriënboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., and Bengio Y., Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078[cs.CL]*, 2014.

Clementi, L. A., Suvire, R. B., Rossomando, F. G., and Vega, J. R., A Closed-Loop Control Strategy for Producing Nitrile Rubber of Uniform Chemical Composition in a Semibatch Reactor: A Simulation Study. *Macromolecular Reaction Engineering*, 12:1700054, 2018.

Goodfellow I., Bengio Y., and Courville A., Deep Learning. *MIT press*, 2016.

Perera, Y. S., Ratnaweera, D. A. A. C., Dasanayaka, C. H., and Abeykoon, C., The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: A critical review. *Engineering Applications of Artificial Intelligence*, 121:105988, 2023

Sangoi, E., SBR and NBR industrial processes energetically coupled: dynamic simulation, estimation of variables and new power contributions to the energy system of the industrial complex" *Ph.D. dissertation, National Technological Univ., Santa Fe, Argentina*, 2021.

Sanseverinatti, C. I., Perdomo, M. M., Clementi, L. A., and Vega, J. R., An Adaptive Soft Sensor for On-Line Monitoring the Mass Conversion in the Emulsion Copolymerization of the Continuous SBR Process. *Macromolecular Reaction Engineering*, 17:2300025, 2023.

Madhuranthakam, C. M. and Penlidis, A. Improved Operating Scenarios for the Production of Acrylonitrile-Butadiene Emulsions. *Polym. Eng. & Sci.*, 53:9-20, 2013.

Perdomo, M. M, Clementi, L. A and Vega, J. R., Estimation of quality variables in a continuous train of reactors using recurrent neural networks-based soft sensors. *Chemometrics and Intelligent Laboratory Systems*, 253:105204, 2024.

Saldívar-Guerra, E., Infante-Martínez, R. and Islas-Manzur, J. M, Mathematical Modeling of the Production of Elastomers by Emulsion Polymerization in Trains of Continuous *Reactors. Processes*, 8:1508, 2020.

Vega, J.R., Gugliotta, L.M., Bielsa, R.O., Brandolini, M.C. and Meira, G.R., Emulsión Copolymerization of Acrylonitrile and Butadiene. Mathematical Model of an Industrial Reactor. *Ind. Eng. Chem. Res*. 36:1238-1246, 1997.