

APRENDIZAJE AUTOMÁTICO PARA PROCESAMIENTO DE GRANDES VOLÚMENES DE DATOS EN MEDICIONES DE MEZCLAS DINÁMICAS DE GASES

Vorobioff J.^{ab}, Noé M.C.^a, Vainer M.N.^a, Orlando N.^a, Checozzi F.^{ab}, Garron Pinto T.^a, Antaurco G.^a, Nieva F.G.^a, Morgada M.^{a,c}, Boggio N.^{abcd}, Filipussi D.^{ab}, Nonino D.^a, Rinaldi C.^{abcd}

^aFacultad Regional Buenos Aires, Universidad Tecnológica Nacional, Argentina, ^bInstituto de Nanociencia y Nanotecnología (CNEA – CONICET), Argentina

^cCONICET, Argentina, ^dUNSAM, Argentina, e-mail: jvorobioff@frba.utn.edu.ar

Introducción

La adquisición masiva de datos de mediciones de gases enfrenta desafíos debido a las respuestas no repetitivas de los sensores, afectadas por el uso, la temperatura, la humedad y el estado previo del sensor. En este trabajo se proponen estrategias para mejorar y optimizar los sistemas sensoriales mediante el análisis de datos de 16 sensores expuestos a mezclas de etileno y CO. Se emplean técnicas de reducción de datos, aprendizaje automático y regresión para optimizar el procesamiento y mejorar la fiabilidad de los sensores en la monitorización ambiental.

Metodología

Se recopilan datos de 16 sensores de gas Figaro que miden mezclas de etileno y monóxido de carbono (CO) en el aire, con mediciones continuas durante un periodo de 12 horas. Las mediciones se realizan en un banco de gases, donde las concentraciones de las mezclas se modifican aleatoriamente cada 80 a 120 segundos. El conjunto de datos incluye todas las posibles transiciones: incremento, decremento o reducción a cero de la concentración de un gas mientras la del otro se mantiene constante. Para las mediciones, se emplea una cámara de 60 ml con un flujo constante de 300 ml/min. Los datos provienen del conjunto de datos del Laboratorio ChemoSignals (Fonollosa J., 2015). Se lleva a cabo una regresión utilizando el método de mínimos cuadrados parciales (PLS), mostrando la varianza acumulada en función del número de componentes PLS. Se observa una varianza acumulada cercana a 1 con 8 componentes. Se calculan las regresiones en Python mediante redes neuronales, OLS, PCR, Ridge, Lasso, Adaboost, árboles de decisión, Redes convolucionales, KNN, entre otros.

Resultados y Conclusiones

- El análisis de regresión mediante **PLS** revela una alta varianza explicada con un número óptimo de componentes, asegurando una precisión correcta en las predicciones de las concentraciones de gases, **reduciendo la carga computacional**.
- En la **matriz de correlación** se observa alta correlación entre muchos sensores, indicando redundancia y la posibilidad de **reducir físicamente el número de sensores utilizados**. La matriz de correlación muestra patrones significativos entre los sensores y las concentraciones de CO y etileno, destacando sensores con baja correlación que podrían indicar fallos o mediciones de parámetros distintos. Esto sirve para **detectar sensores con fallas**. Como trabajos futuros se pueden utilizar otros métodos de selección de variables.
- Se obtiene una buena regresión, con **coeficientes de determinación: $R^2 = 0,941$** .

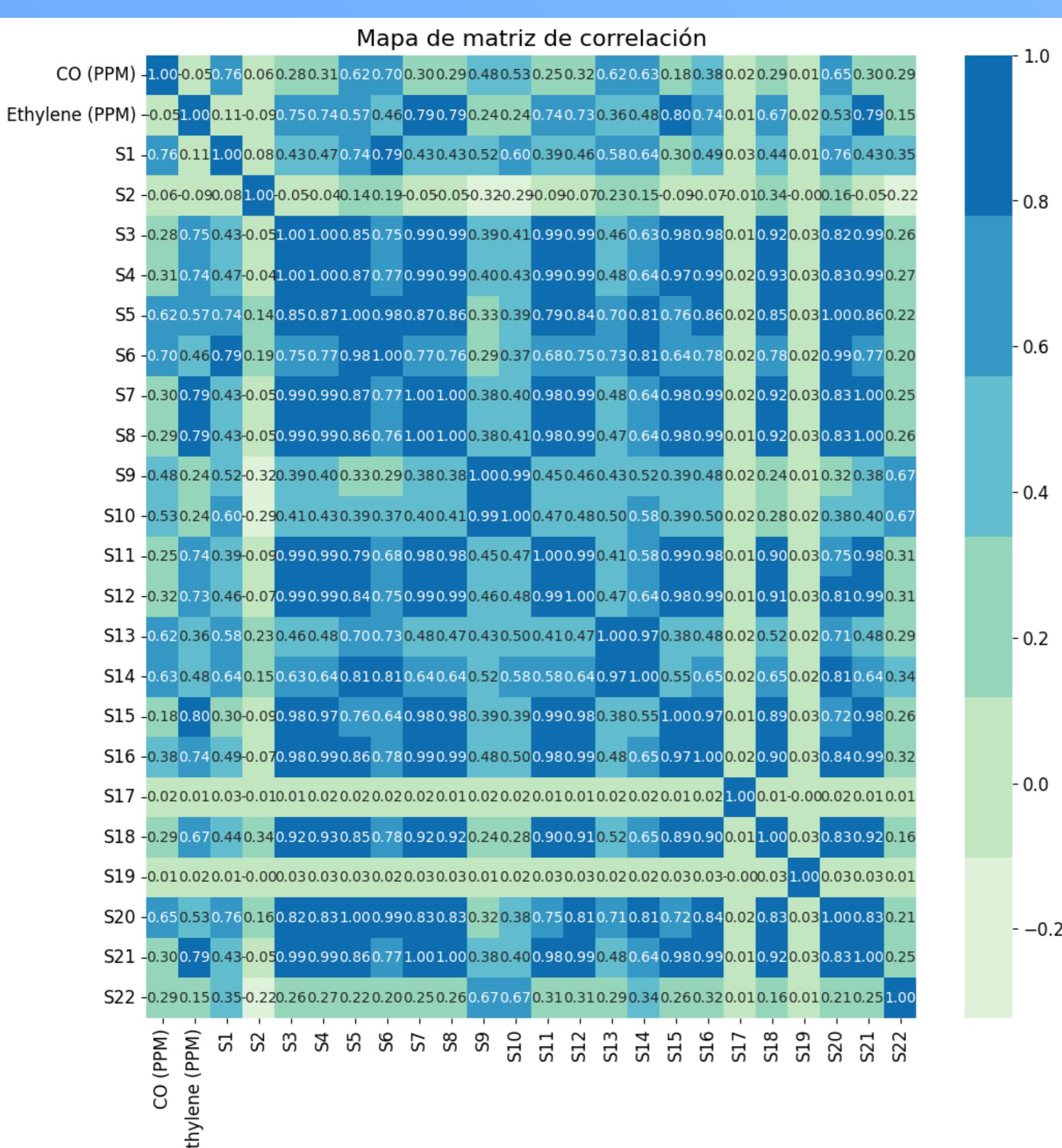


Figura. Matriz de correlación de los sensores y las concentraciones de gases

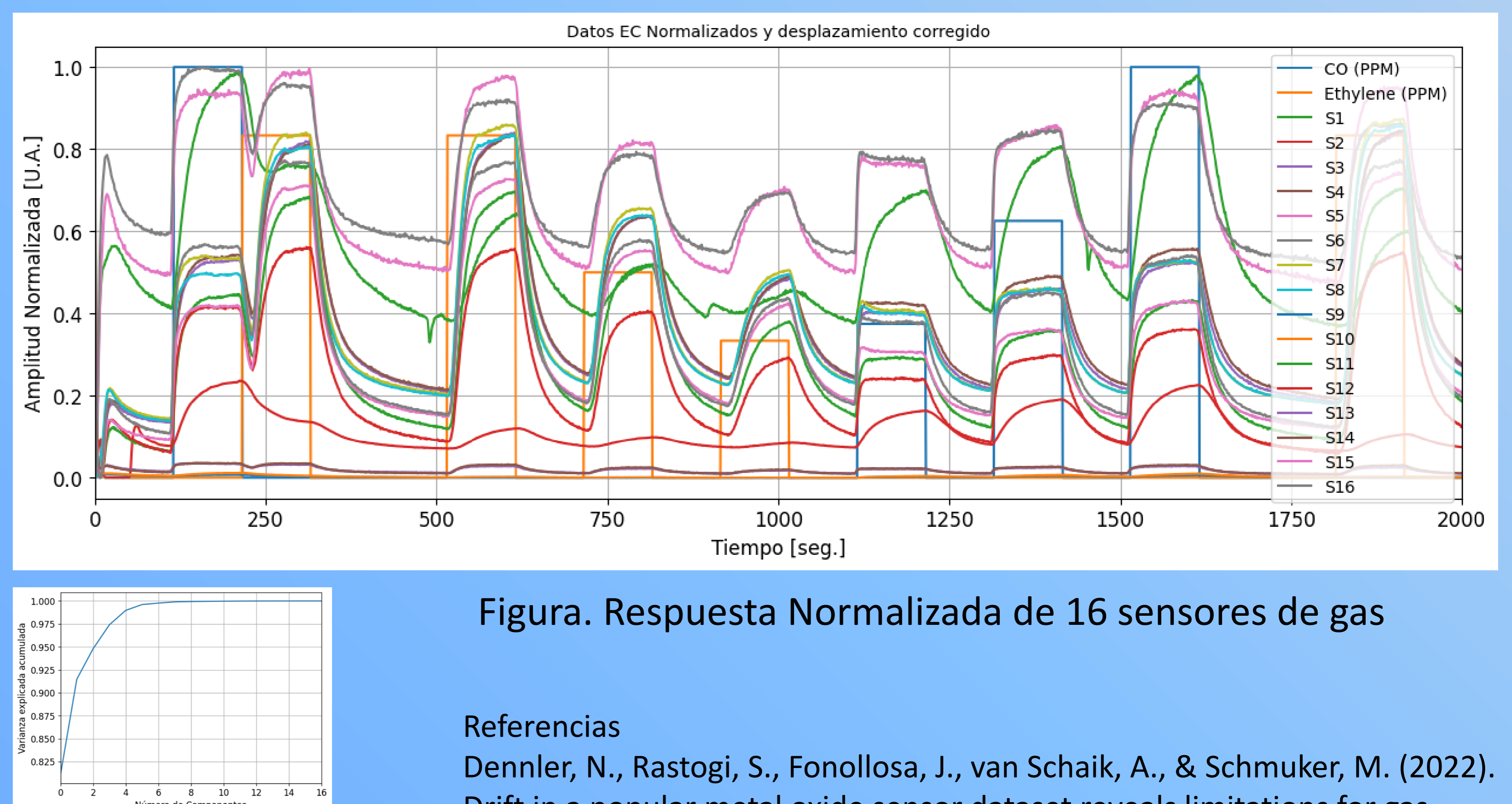


Figura. Respuesta Normalizada de 16 sensores de gas

Figura. Varianza acumulada respecto al número de componentes PLS

Referencias

Dennler, N., Rastogi, S., Fonollosa, J., van Schaik, A., & Schmuken, M. (2022). Drift in a popular metal oxide sensor dataset reveals limitations for gas classification benchmarks. *Sensors and Actuators B: Chemical*, 361, 131668. <https://doi.org/10.1016/j.snb.2022.131668>