
Análisis y Predicción de Gases Contaminantes en una Turbina de Gas



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

AUTOR:

DAVID ORTIZ ARMIÑANA (dortarm@upv.edu.es)

CURSO:

**ANÁLISIS, MONITORIZACIÓN Y DIAGNÓSTICO DE PROCESOS
MULTIVARIANTES**

PROFESOR:

DR. ALBERTO JOSÉ FERRER RIQUELME

RESUMEN

En este estudio se trabajará con una turbina de gas de ciclo combinado para la generación de electricidad en una central eléctrica situada en la región noroeste de Turquía. El objetivo principal del trabajo es determinar las causas por las que se producen grandes emisiones de gases contaminantes en la turbina y proponer posibles soluciones para reducirlas.

Con la técnica PCA detectaremos un problema que ocurre con frecuencia y que obligará a dividir la base de datos en observaciones bajo control y fuera de control. Más adelante, con la técnica PLS se estudiarán las relaciones de las variables del proceso, tanto entre ellas como con las emisiones de gases, para ambas situaciones definidas.

La técnica PCA también se utilizará para llevar a cabo una monitorización del proceso en tiempo real. Con el fin de estudiar si las emisiones dependen de instantes anteriores y obtener un modelo predictivo preciso se empleará el método de PLS-Series Temporales. Por último, con la metodología de procesos por lotes se pretenderá determinar las condiciones adecuadas que debe seguir el proceso para que no se produzcan emisiones anómalamente elevadas.

INTRODUCCIÓN

Un ciclo combinado consta de dos partes, el ciclo de gas (con el que trabajaremos) y el ciclo de vapor. En el ciclo de gas se introduce aire ambiente a la turbina, pasando primero por un filtro que tiene como función limpiarlo de impurezas como polvo, arena u otros contaminantes que puedan dañar los componentes internos.

Posteriormente, el aire pasa al compresor, donde se aumenta su temperatura y su presión con el fin de que, tras mezclarlo con un combustible (gas natural), se encienda a altas temperaturas en la cámara de combustión y se generen gases calientes a altas presiones que hagan girar el eje de la turbina, conectada a un generador que transforma la energía mecánica en energía eléctrica.

Los gases calientes están compuestos por vapor de agua, dióxido de carbono, oxígeno, monóxido de carbono (CO) u óxidos de nitrógeno (NOX), entre otros. La emisión de gases nocivos en procesos de combustión, NOX y CO sobre todo, está limitada por la normativa europea, debido a que son altamente perjudiciales tanto para el medio ambiente como para la salud humana.

Los límites de emisión europeos son 30.8 mg/m^3 para el CO y 41.8 mg/m^3 para el NOx. La turbina del proceso genera picos de emisión de CO por encima del límite y valores elevados de NOX, con una media de 65.29 y un máximo de 119.91. La central quiere ajustarse lo máximo posible a los estándares europeos, por lo que es necesario determinar con exactitud en qué puntos del proceso se producen las causas que generan los fallos en las emisiones.

MATERIAL Y MÉTODOS

La base de datos consiste en el promedio horario de 11 variables cuantitativas del proceso medidas desde el 1 de enero de 2011 hasta el 31 de diciembre de 2015. No obstante, a lo largo de este tiempo se han producido parones para limpieza y reparación de la turbina de gas, por lo que en ningún año se toman mediciones en todas las horas. Durante el trabajo se utilizarán los datos de los 4 primeros años (2011-2014, 29349 observaciones) como conjunto de entrenamiento y los datos del último año como conjunto de prueba (7384 obs.).

A continuación, se enumeran y describen las variables medidas:

AT (Ambient Temperature): temperatura del aire (ambiente) que entra en el proceso (°C).

AP (Ambient Pressure): presión del aire que entra en el proceso (milibares).

AH (Ambient Humidity): humedad del aire que entra en el proceso (%).

AFDP (Air Filter Difference Pressure): diferencia de presión del aire antes y después del filtro inicial (milibares).

GTEP (Gas Turbine Exhaust Pressure): presión de los gases que salen de la turbina (milibares).

TIT (Turbine Inlet Temperature): temperatura de entrada del aire a la cámara de combustión (°C).

TAT (Turbine After Temperature): temperatura de salida de los gases de la cámara de combustión (°C).

CDP (Compressor discharge pressure): presión del aire comprimido antes de mezclarse con el combustible (milibares).

TEY (Turbine energy yield): rendimiento energético de la turbina (MWH).

CO (Carbon monoxide): emisión de monóxido de carbono (mg/m^3).

NOX (Nitrogen oxides): emisión de óxidos de nitrógeno (mg/m^3).

CO y NOX serán las variables respuesta o de interés a lo largo de todo el estudio, mientras que el resto de las variables del proceso se utilizarán como explicativas. Los donantes de la base publicaron un artículo de predicción de emisiones que se puede consultar, junto con gráficas de los datos, en el Anexo. En las gráficas se puede ver que las emisiones elevadas de CO son menos frecuentes, mientras que las de NOX se producen más a menudo y siguen ciclos de emisión, con valores más altos durante los meses de invierno.

Para aplicar todas las técnicas vistas en el curso (PCA, PLS, PLS-TS y Batch-MSPC) se empleará el software Aspen ProMV, mientras que para trabajar con

la base de datos (crear conjunto de entrenamiento y de prueba, subdividir la base en datos bajo control y fuera de control, ...) se utilizará el software R.

RESULTADOS

ANÁLISIS EN COMPONENTES PRINCIPALES (PCA)

En primer lugar, se empezará con un análisis en componentes principales sobre las variables explicativas del conjunto de entrenamiento para extraer las relaciones existentes entre ellas y detectar la presencia de posibles valores anómalos. Con la función *Auto Fit* de Aspen ProMV se ajusta un modelo de 4 componentes principales por el criterio de parada de no sacar más de $J/2$ componentes, donde J es el número de variables, en nuestro caso, $J=9$.

La primera componente explica un 54.44% de la variabilidad, la segunda un 20.31%, la tercera un 10.55% y la cuarta un 6.13%. Teniendo en cuenta que tenemos 9 variables, para que una componente principal tenga un valor propio mayor que 1 debería explicar más de un $100/9=11.11\%$ ($100 \cdot \frac{\lambda}{J}$, $\lambda = 1$, $J = 9$).

La cuarta componente tiene mucho menor porcentaje, por lo que se decide eliminarla. La tercera, en cambio, tendrá un valor propio de $10.55/11.11 = 0.9495$, cercano a 1, por lo que se decide mantenerla en el modelo.

El primer paso para llevar a cabo un PCA es validar el modelo con los gráficos de la Suma de Cuadrados Residual (para detectar observaciones atípicas que rompan la estructura de correlación) y de la T^2 de Hotelling (para detectar observaciones extremas). No se observan observaciones extremas en este análisis, cuyos resultados se muestran en la Figura 1.

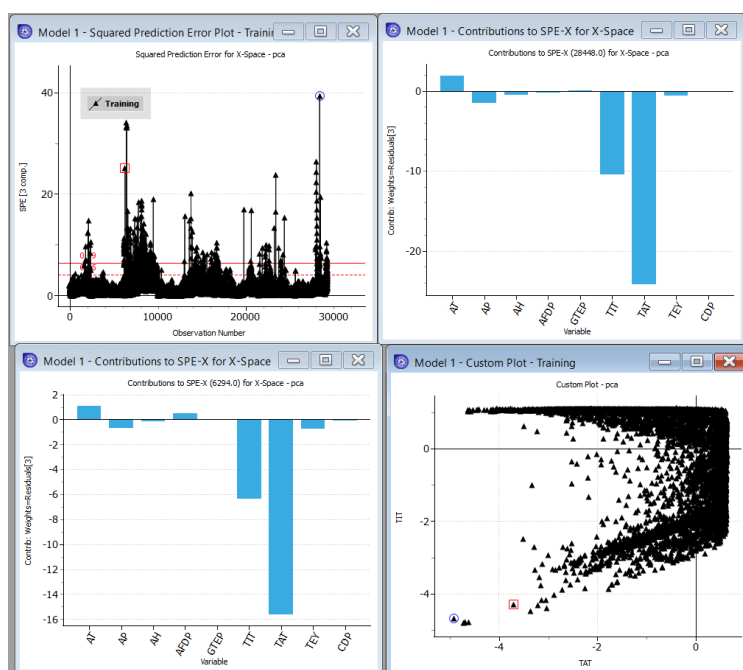


Figura 1: Gráfico de la Suma de Cuadrados Residual, contribución de datos anómalos y Custom Plot de las variables TIT y TAT.

Observando primero el gráfico de la Suma de Cuadrados Residual se pueden ver una gran cantidad de observaciones muy por encima del límite de control al 99%, por lo que claramente se trata de observaciones atípicas que rompen la estructura de correlación de los datos. Para encontrar las causas por las que se producen estas anomalías se grafican las contribuciones de las observaciones atípicas, en las que se detecta valores demasiado bajos de las variables TIT y TAT (temperatura de entrada y salida a la cámara de combustión). En el Custom Plot se observa que estas observaciones rompen la estructura de correlación entre ambas variables.

Además, como este hecho no se produce de forma puntual, sino que se repite en instantes de tiempo ampliamente distanciados, se puede considerar que esta turbina tiene un problema estructural que se menciona de forma puntual en el artículo: cuando la temperatura de entrada a la turbina (TIT) es demasiado baja, la combustión se vuelve incompleta, lo que deriva en mayor formación de monóxido de carbono (CO).

No obstante, lo que no se menciona en este artículo y que se puede deducir de este primer análisis es que cuando la temperatura de entrada a la turbina es demasiado baja también se puede obtener una temperatura de salida de los gases muy baja (producida por la combustión incompleta), con las consecuencias que esto puede tener en el resto del proceso.

En este punto del estudio y sabiendo que existe un problema en la turbina que afecta a la emisión de los gases contaminantes, se decide dividir la base de datos en dos partes, una parte donde se tengan los datos correspondientes a emisiones que se consideran bajo control para esta turbina y otra parte donde se tengan emisiones fuera de control.

Para llevar a cabo esta subdivisión se podrían utilizar los límites impuestos por la Unión Europea, que recordamos eran 30.8 mg/m^3 para el CO y 41.8 para el NOX. Sin embargo, la media de emisiones de CO para esta turbina es muy inferior (2.37), por lo que se produciría una sobreestimación del límite, derivando en poca capacidad para detectar salidas de control. Por otro lado, la media de emisiones de NOX es muy superior (65.29), lo que llevaría a una subestimación del límite de control, derivando en un exceso de falsas alarmas.

Para subdividir la base de datos se consideran como incorrectas aquellas observaciones con niveles de emisiones anómalos a nivel univariante en cualquiera de las dos variables de interés, es decir, aquellas que superen el valor determinado por la suma del tercer cuartil y 1.5 veces el rango intercuartílico de cualquiera de las dos variables de interés. En el caso del CO el valor es 4.35095, mientras que para el NOX es 92.725. En total hay 3445 observaciones fuera de control, lo que supone un 11.74% del conjunto total de entrenamiento. Con un PCA sobre las variables respuesta en el conjunto bajo

control se ha comprobado que no existen observaciones anómalas a nivel multivariante.

PARTIAL LEAST SQUARE REGRESSION (PLS)

Con el fin de encontrar las relaciones entre las variables explicativas y con las variables respuesta en ambos escenarios, se llevarán a cabo dos análisis utilizando el método PLS. Comenzaremos con la situación correspondiente a cuando las emisiones de gases contaminantes se encuentran fuera de control. En primera instancia se obtienen cuatro componentes principales para la matriz de las variables explicativas, pero se reduce el número a tres porque la cuarta no explica demasiada variabilidad. La primera componente explica un 39.567% de la variabilidad total, la segunda un 4.971% y la tercera un 9.414%, es decir, la tercera componente explica un porcentaje de la variabilidad total mayor que la segunda. Visualizamos primero los gráficos de validación del modelo en la Figura 2.

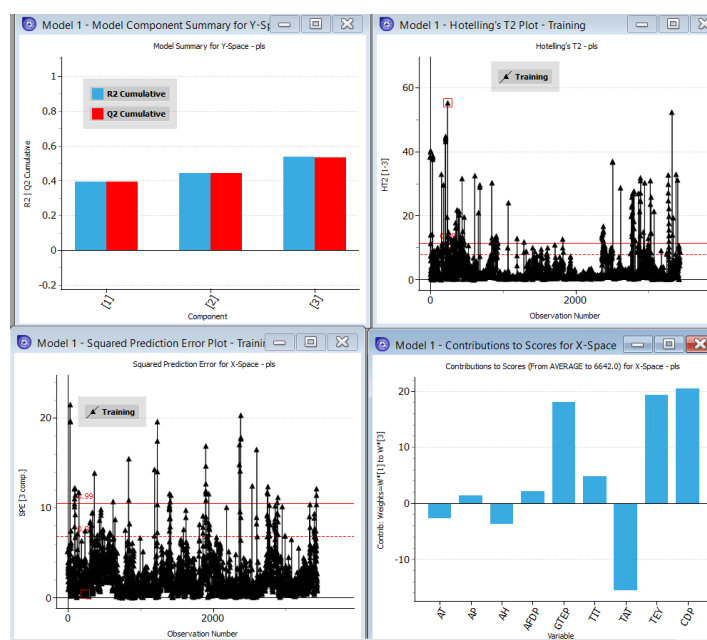


Figura 2: Gráficos de validación del modelo PLS.

En el gráfico de Suma de Cuadrados Residual las observaciones no se alejan más de dos veces el límite de control al 99%, pero en el gráfico de la T^2 de Hotelling sí, por lo que tenemos presencia de observaciones extremas. Como se puede ver en el gráfico de contribución, estas anomalías se producen a nivel multivariante con más de dos variables implicadas. A continuación, se procede a depurar el modelo hasta que no se visualicen observaciones anómalas. Con este modelo se obtienen gráficos de pesos w^*c/w^*c (Figura 3) idénticos a los del primer modelo, el no depurado. Los gráficos que aparecen a continuación son de la primera componente frente a la segunda y frente a la tercera.

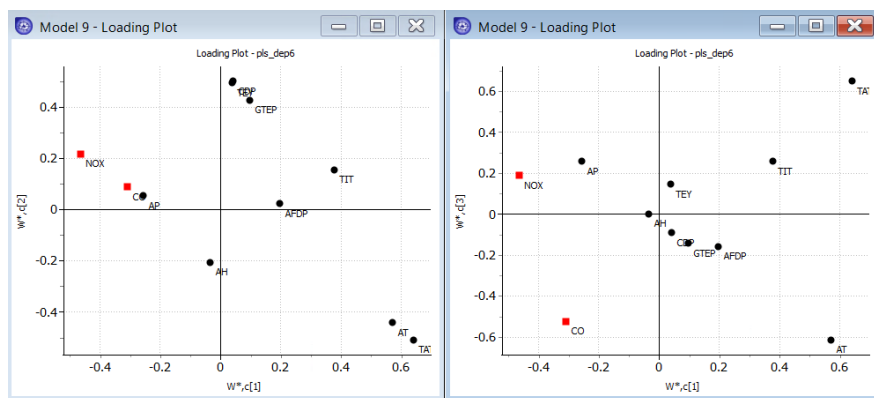


Figura 3: Gráficos de pesos w^*c/w^*c para el conjunto fuera de control.

Como se puede observar en los gráficos de la Figura 3, en la primera componente las variables AT, TIT y TAT están correlacionadas positivamente, y a su vez negativamente con las variables de interés, CO y NOX. De estos pesos en la primera componente podemos deducir que las temperaturas de entrada bajas a la turbina, que provocan combustión incompleta, en la primera componente están relacionadas con temperaturas ambientales del aire bajas. Además, se corrobora que valores demasiado reducidos de la variable TIT generan valores reducidos de TAT.

Recordamos que la primera componente explica un 40% de variabilidad, luego existe un gran número de observaciones en las que la combustión incompleta puede estar relacionada con temperaturas ambientales demasiado bajas, y en estos casos se producen emisiones elevadas tanto de CO como de NOX, las emisiones están muy correlacionadas. Sin embargo, es conocido el hecho de que la formación de NOX no puede deberse a la combustión incompleta. En las otras componentes no se ve que las variables AT y TIT estén correlacionadas positivamente. Junto con la Figura 1 podemos deducir que hay veces que la turbina no consigue calentar suficientemente el aire, aunque no sea frío. Además, con gráficos de loadings no se puede intuir ninguna relación más de la variable TIT con otras variables explicativas.

En cuanto a las variables respuesta, por la Figura 3 el CO parece estar correlacionado negativamente con las variables TIT y TAT (combustión incompleta), aunque más con TAT. Visualizando los datos podemos encontrar observaciones con altas emisiones de CO con valores de TIT altos y de TAT bajos, por lo que se puede deducir que la combustión incompleta también se produce en ocasiones aún con temperaturas de entrada altas. Para el NOX, la variable más relacionada es la temperatura ambiental de forma negativa, pero esta relación no parece ser causa-efecto, no es una relación razonable.

Pasamos ahora a realizar un modelo PLS para el conjunto de datos bajo control. Las componentes principales que se obtienen con Aspen ProMV son cuatro. En el gráfico de la T^2 de Hotelling destaca una observación, la número 22270, por tener valores bajos en las variables TIT y TAT, luego se trata de una

observación en la que se ha producido una combustión completa y se elimina. El nuevo modelo tiene un $R^2 = 52.983\%$ y un $Q^2 = 52.970$.

Como se puede ver en el gráfico de la Suma de Cuadrados Residual de la Figura 4, destaca una observación, la 13802, que tiene valores elevados en la variable AFDP, la diferencia de presión del aire antes y después de salir del filtro, algo que puede deberse a obstrucciones y necesidad de limpieza. No obstante, con un gráfico de Raw Data vemos que este hecho se repite varias veces durante las mediciones totales, por lo que no se elimina este dato.

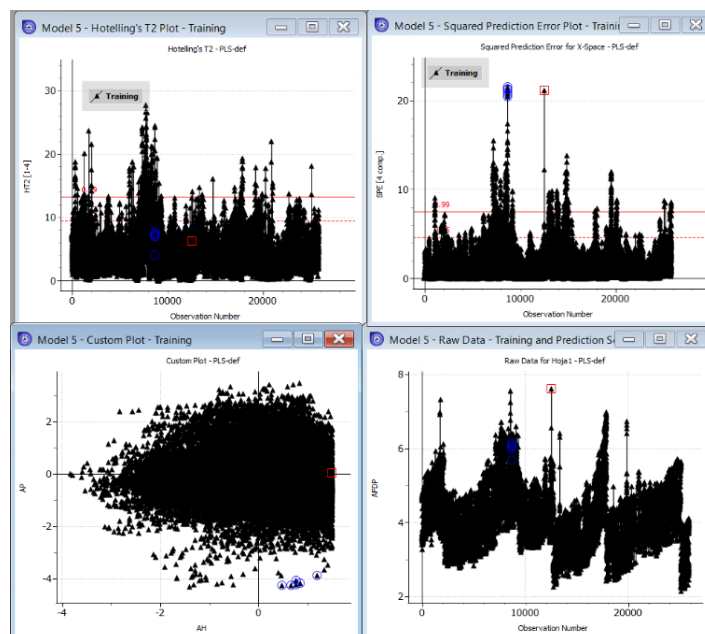
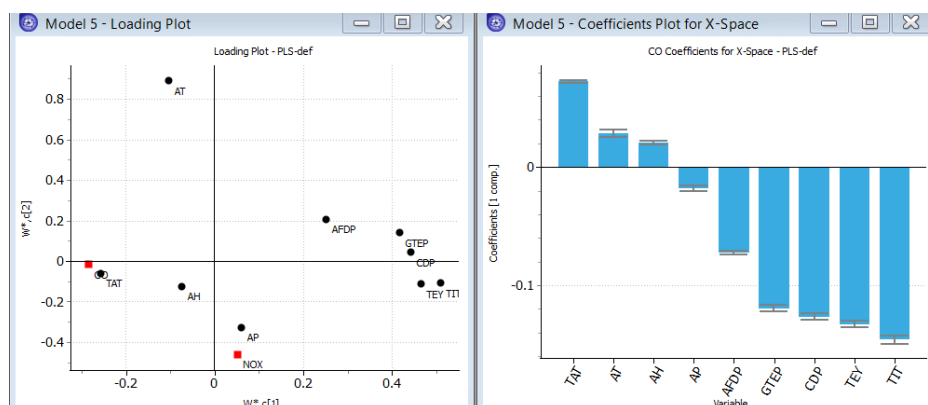


Figura 4: Validación del modelo PLS.

En la Figura 4 se observa también que hay algunas observaciones atípicas del gráfico de la Suma de Cuadrados Residual. Con el gráfico de contribuciones se puede ver que estos datos toman valores bajos para la presión ambiental, pero no es un hecho puntual, sino que se repite en otros instantes de tiempo distanciados, luego se deduce que se trata de condiciones ambientales específicas que se dan en un determinado momento de cada año, por lo que se decide no eliminarlas del modelo. En la Figura 5 se hace un gráfico de cargas del modelo PLS ajustado para las componentes 1 y 2.



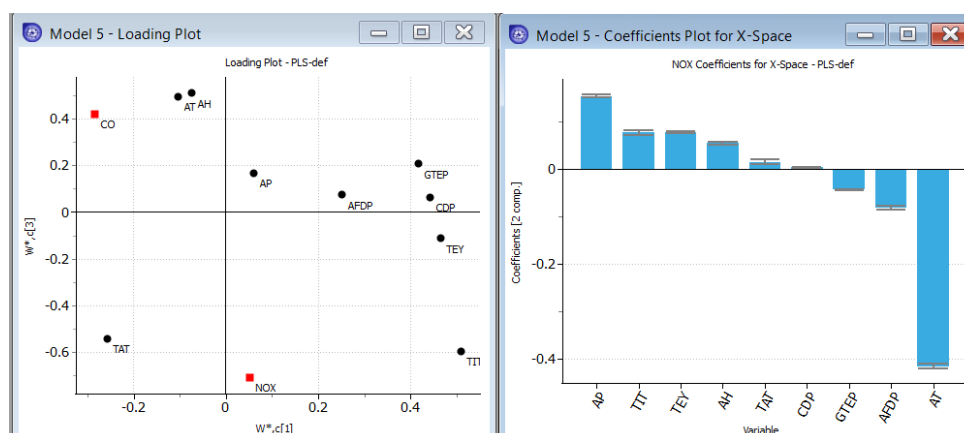


Figura 5: Resultados del modelo PLS ajustado al conjunto bajo control.

En el gráfico de pesos w^*c/w^*c de arriba se observan las siguientes relaciones:

- La primera componente principal explica la relación de la variable de interés CO con las variables TAT, TIT, TEY, CDP, GTEP y AFDP. Las 5 últimas están correlacionadas positivamente entre ellas, mientras que CO y TAT tienen alta correlación positiva pero negativa con el resto de las variables.
- La segunda componente principal explica la relación de la variable de interés NOX con las variables AT y AP. La emisión de NOX está correlacionada de forma positiva con la presión y de forma negativa con la temperatura ambiental.

Estas relaciones de las variables respuesta con las explicativas se pueden corroborar con un análisis de los coeficientes en las componentes respectivas. En los coeficientes para la variable CO en la primera componente, las variables que más peso tienen son TIT, TEY, CDP, GTEP, AFDP, con coeficientes negativos, y TAT, con coeficiente positivo. Para el NOX, las variables que tienen coeficientes más altos en la segunda componente son AT (negativo) y AP (positivo). Las variables ambientales están correlacionadas con el NOX, pero de nuevo estas relaciones carecen de causalidad. Todas las variables tienen algún coeficiente significativo en alguna componente, por lo que no se plantea eliminar ninguna. Además, se puede deducir que las dos variables respuesta están en direcciones ortogonales de variabilidad, luego no están relacionadas.

PLS-SERIES TEMPORALES

Como se ha visto en la introducción del trabajo, los datos del proceso son promedios de 1 hora y están ordenados por antigüedad, por lo que se tratan de una serie temporal. En consecuencia, parece razonable pensar que los valores que tome una variable en un instante van a depender de los valores que se tomen en instantes anteriores, con especial interés en las variables respuesta, luego tiene sentido aplicar la técnica de PLS-Series Temporales.

No obstante, estamos trabajando con el conjunto bajo control, al que se le han eliminado observaciones. Por esto, en el PLS-TS tan solo consideramos dos

decalajes para todas las variables, ya que si introducimos más decalajes pueden haber muchas observaciones que no tengan valores para el decalaje 3 o más. Además, por los parones que se han producido para limpiezas y reparaciones, no podemos asegurar que dos observaciones consecutivas sean de dos horas seguidas. Este primer modelo tiene un $R^2 = 81.742\%$ y un $Q^2 = 81.709\%$, es decir se explica mucha más variabilidad y la capacidad predictiva es mucho mayor que para modelos anteriores. Para determinar las variables más relevantes en un modelo PLS se utiliza el VIP, como en la Figura 6.

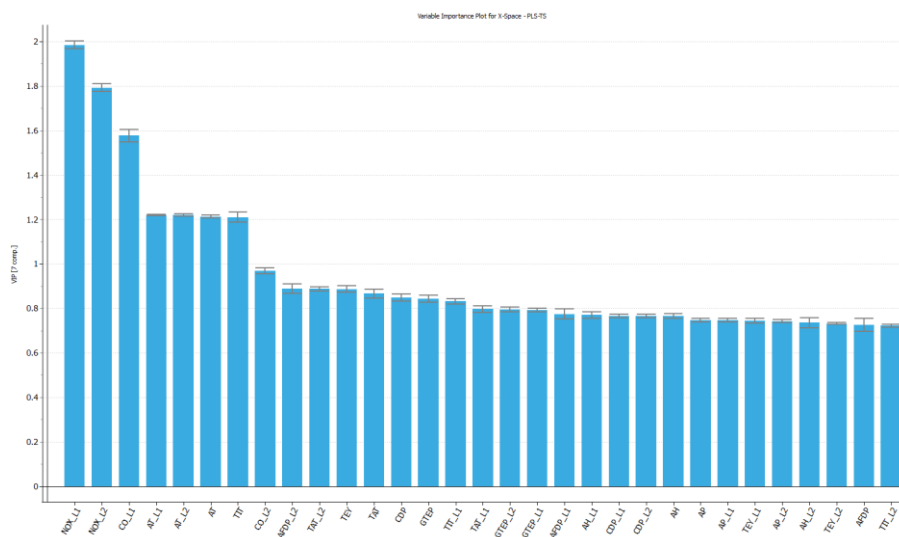
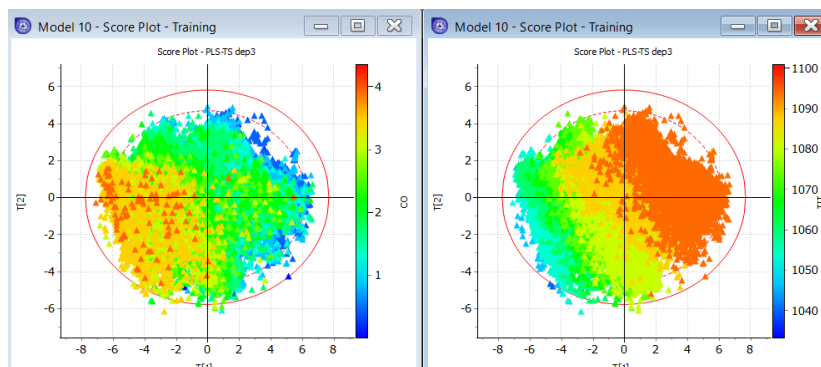


Figura 6: VIP de las variables en el modelo PLS-Series Temporales.

Aquellas variables con un valor del VIP mayor que 1 son muy relevantes en el modelo, mientras que aquellas con VIP menor que 0.8 son muy poco relevantes. Siguiendo esta norma se van eliminando paulatinamente retardos y variables con valores del VIP significativamente menores que 0.8, es decir, tales que el límite superior del intervalo Jackknife no sobrepase dicho valor.

Con el modelo depurado final se tiene un $R^2 = 81.524\%$ y un $Q^2 = 81.504\%$, ligeramente menor que para el modelo original, pero con menos variables y retardos, en concreto las variables AP, AH y AFDP ya no aparecen en el análisis. Del *Variable Importance Plot* de la Figura 7 se puede deducir que los dos retardos de la variable NOX y el primero del CO son las variables que mayor importancia tienen en el modelo, seguidas de la variable TIT.



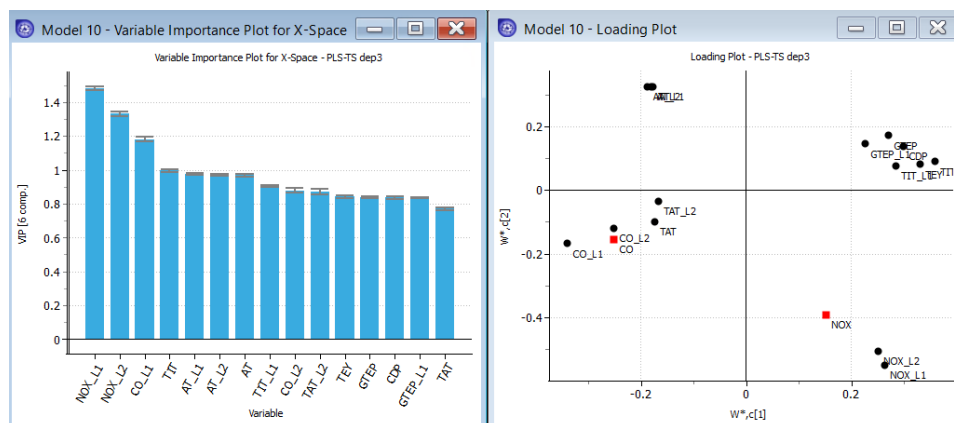


Figura 7: Modelo de PLS-Series Temporales depurado para el conjunto bajo control.

En el gráfico de pesos w^*/w^* de la Figura 7 se vuelven a ver dos claras direcciones de variabilidad: la variable CO está correlacionada positivamente con sus retardos y con la variable TAT, y negativamente con las variables TIT, TEY, GTEP, CDP y sus retardos. La relación entre TIT y CO se puede comprobar gráficamente coloreando el gráfico de scores según ambas variables, donde se observa cómo, a medida que disminuye TIT, las emisiones de CO tienden a aumentar y viceversa. Por otro lado, la variable NOX está correlacionada positivamente con sus retardos y negativamente con la temperatura ambiental.

Con el modelo de PLS-Series Temporales entrenado se realiza una predicción de las emisiones del año 2015 que se encuentran bajo control, es decir, que no tienen emisiones anómalas de CO ni de NOX. En la Figura 8, con dos gráficos de valores predichos frente a reales, se pueden consultar los ajustes de la predicción de ambas variables respuesta. Podemos observar que el modelo predice mejor las emisiones de NOX, pues la nube de puntos de ajusta mejor a la recta ideal. Para esta variable se tiene un $R^2 = 85.832\%$ y un RMSE = 2.91709, valores que indican que el modelo predice realmente bien las emisiones de este gas. Por la nube de puntos más dispersa y el menor $R^2 = 77.2158\%$ se deduce que las emisiones de CO se predicen peor, aunque el RMSE es más bajo porque los valores de las emisiones también lo son.

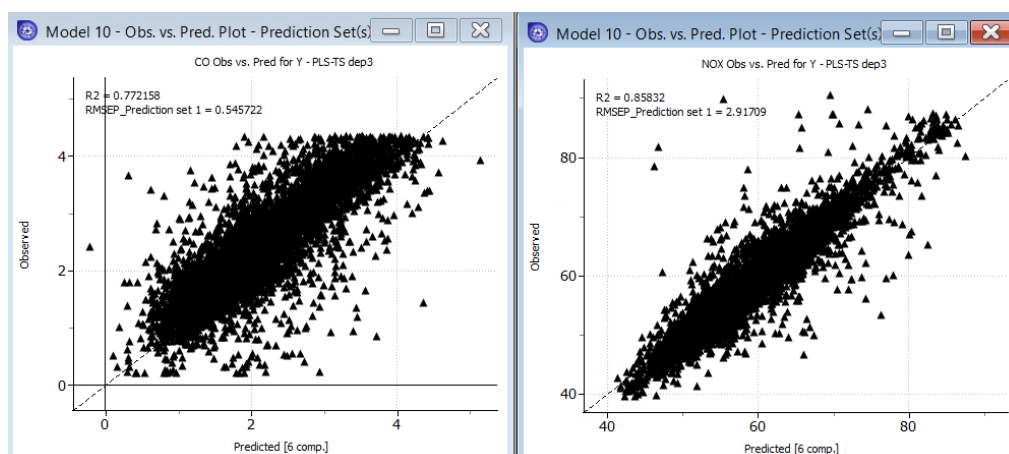


Figura 8: Gráficos de predichos vs reales en las emisiones de gases contaminantes.

No obstante, si se quiere conseguir un modelo que sea capaz de predecir observaciones bajo control y fuera de control habría que entrenarlo con todos los datos. Para ello entrenamos un modelo PLS-TS con 5 decalajes por estar utilizando toda la base de datos. Depuramos el modelo eliminando aquellas variables que tengan un VIP significativamente menor que 0.8 y, finalmente, obtenemos un modelo que realiza la siguiente predicción para los datos del año 2015.

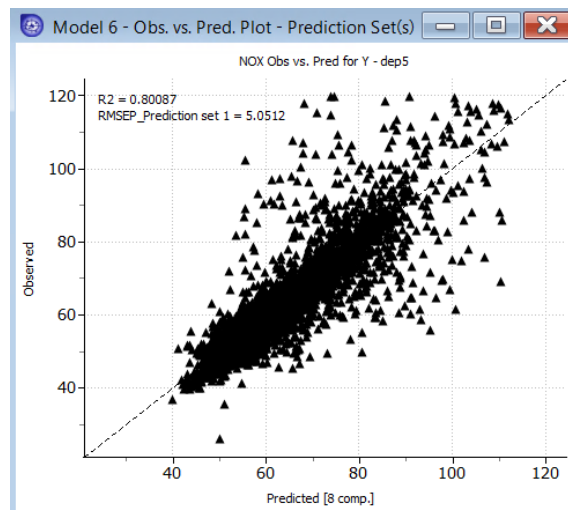


Figura 9: Predicción del NOX del conjunto de prueba con el modelo PLS-TS.

El modelo depurado tiene un RMSE = 5.0512 para todo el conjunto de prueba, una predicción bastante mejor que con otros modelos vistos durante el máster:

Modelo	RMSE
ARIMAX	9.8525
Regresión lineal	10.2713
Random Forest	10.9131

Tabla 1: Predicción de distintos modelos para el conjunto de prueba

MULTIVARIATE STATISTICAL PROCESS CONTROL (MSPC)

La técnica de análisis de componentes principales es de gran utilidad para la monitorización de las variables del proceso. Entrenando un modelo PCA con las observaciones que se consideren bajo control, se puede controlar si los nuevos datos que se recojan presentan problemas y, en caso afirmativo, se pueden estudiar las causas con gráficos de contribución del dato en cuestión.

En consecuencia, en esta parte del estudio se empleará el conjunto de datos bajo control para ajustar un modelo PCA sobre todas las variables, tanto las variables explicativas como las de interés. Con este enfoque se controlan las emisiones de gases contaminantes y con gráficos de contribución se identifican las causas de fallo del proceso en caso de que sean demasiado elevadas, a la

vez que se vigilan que las otras variables del proceso no estén fuera de control, algo que eventualmente puede provocar emisiones fuera de especificaciones.

Como conjunto de prueba se utilizarán las primeras 105 observaciones del año 2015, suficientes para visualizar cómo se puede utilizar el modelo PCA para la monitorización multivariante. Los resultados se pueden consultar en la Figura 10.

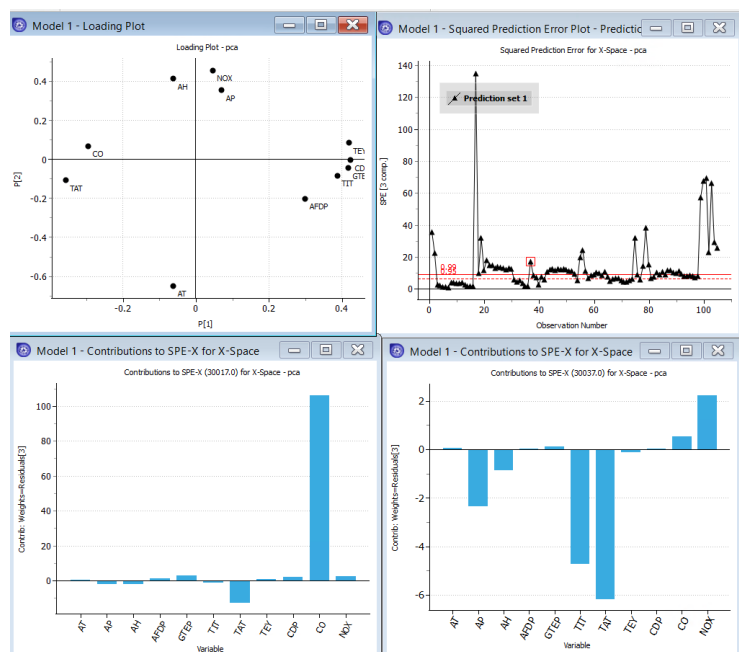


Figura 10: Resultados de la monitorización del proceso con la técnica PCA.

En la T^2 de Hotelling y en la Suma de Cuadrados Residual de la Figura 10 se detectan observaciones fuera de los límites de control, como la 17 o la 37. Con los gráficos de contribuciones se puede ver que para la primera se ha producido una emisión grande de monóxido de carbono (12.659) provocada por una combustión incompleta, con una temperatura de salida de los gases (TAT) baja (535.40). Para la observación 37 se temperatura de entrada (1034.3) y de salida (536.35) bajas, con emisiones elevadas de CO y de NOX.

Podemos ver que con el modelo PCA se detectan las salidas de control del proceso de forma adecuada y se encuentran las causas que las generan, luego es útil para llevar a cabo una monitorización multivariante. Además, estos gráficos podrían alertar de una posible salida de control de los gases antes de que sucediera, en el caso de detectar valores demasiado bajos de la temperatura de entrada o de salida de la turbina, que pueden ser síntoma de que una combustión incompleta esté cerca de producirse.

PROCESO POR LOTES

Con el fin de comprobar las relaciones vistas hasta ahora y determinar en qué condiciones se debería trabajar para que no se produzcan fallos se aplicará la metodología de procesos por lotes a nuestros datos. En concreto, como

sabemos que este proceso no es estacionario y las relaciones entre las variables presentan una dinámica cambiante, utilizaremos la técnica de procesos por lotes batch-wise.

La estrategia que se seguirá para llevar a cabo esta metodología será entrenar el modelo con el conjunto bajo control de los primeros 4 años y emplear el último año como conjunto de prueba para analizar los fallos que se producen en cada lote, que en nuestro caso serán meses (48 de entrenamiento, 12 de test). Como en todos los años no se tiene el mismo número de observaciones, el tamaño de todos los lotes no es el mismo. Con el programa Aspen ProMV se ha aplicado el método Dynamic Time Warping para llevar a cabo un alineado, pues con un Linear Alignment, que trunca los datos en el tamaño de lote más pequeño, se podría perder información interesante de emisiones al final de cada lote. Con el DTW Alignment se reduce el tamaño de los lotes a 50 puntos.

Construimos un modelo PCA con todas las variables, tanto explicativas como respuesta, con los lotes (meses) del conjunto bajo control. Los resultados la proyección del conjunto de prueba se pueden observar en la Figura 11.

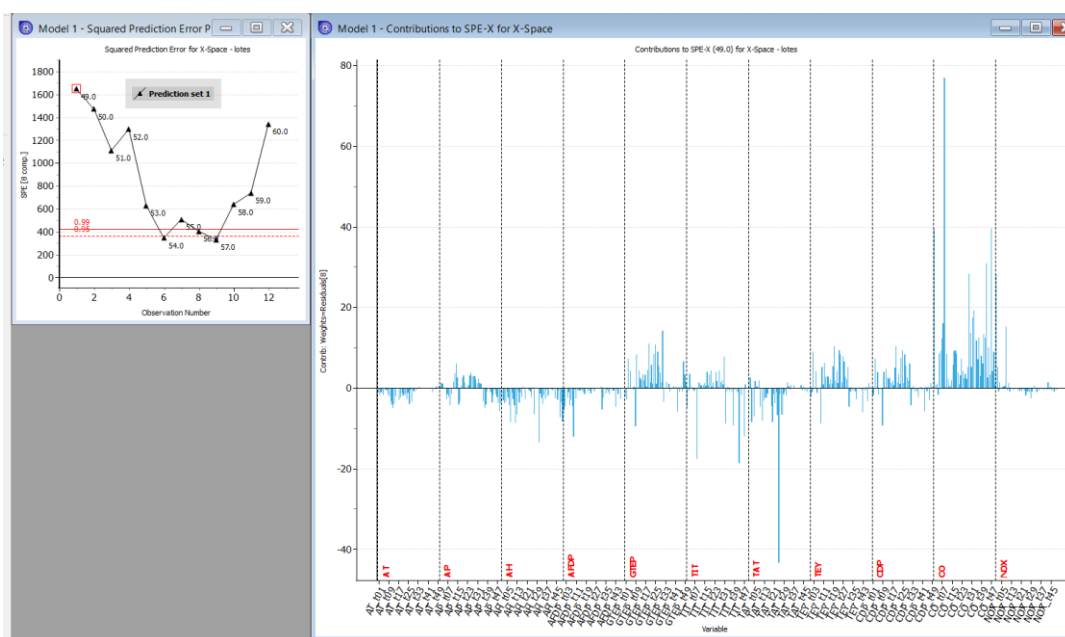


Figura 11: Proyección de los nuevos lotes en el modelo.

En el gráfico de la Suma de Cuadrados Residual se observa que casi todos los lotes están fuera de los límites de control, y haciendo un gráfico de contribución para el lote 49 se observan contribuciones muy altas de la variable CO y altas de NOX, que coinciden con contribuciones bajas de la variable TIT. Para observar con más detalle este fenómeno se hacen gráficos Raw Data para TIT y CO en este lote, en la Figura 12, y para TAT y NOX, en la Figura 13.

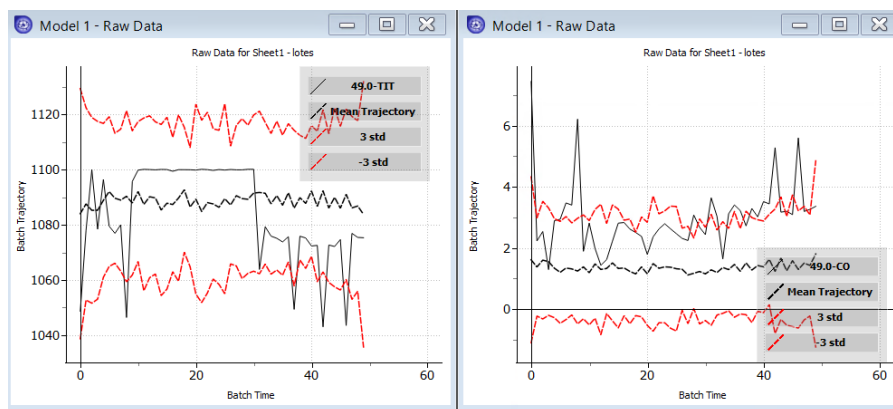


Figura 12: Valores de las variables TIT y CO en el lote 49.

Se puede comprobar que las mayores emisiones de CO coinciden con las temperaturas de entrada a la cámara más bajas del lote, las que salen del intervalo de confianza al 95%. Por otra parte, para valores de TIT cercanos o superiores a la trayectoria media de los lotes del conjunto de entrenamiento, las emisiones de CO están dentro de las especificaciones. Esta media está sobre los 1090°C.

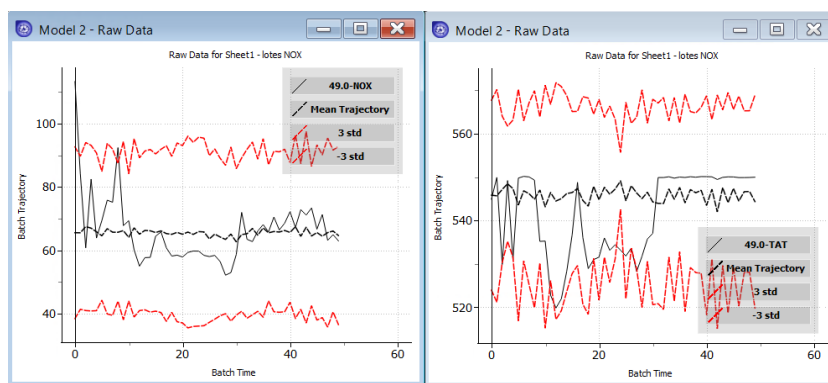


Figura 13: Valores de las variables NOX y TIT del lote 49.

Como podemos observar en la Figura 13, la variable NOX también sale de los límites en algunas observaciones para las que se produce combustión incompleta, aunque ambos hechos no están directamente relacionados. La trayectoria media la variable TAT en los lotes del conjunto bajo control tiene valores entre 545°C y 550°C, luego tomando estos valores de TAT y los observados para la trayectoria media de la variable TIT se reduciría en gran medida la posibilidad de que ocurriera una combustión incompleta.

CONCLUSIONES

En primer lugar, en este estudio se ha descubierto que el problema de la combustión incompleta no es un problema puntual, sino que afecta a muchas observaciones de la base de datos y cambia la estructura de las relaciones entre las variables. Según el artículo de los donantes de la base de datos, la combustión incompleta, causante de altas emisiones de CO, se genera cuando

la temperatura de entrada de los gases a la turbina (TIT) es demasiado baja. También se menciona que la variable NOX está moderadamente relacionada de forma negativa con la temperatura ambiental y que, en consecuencia, trabajar con mayores temperaturas ambientales reduce la emisión de estos gases contaminantes. No se realiza ningún comentario más sobre el proceso.

La relación entre las emisiones del NOX y la temperatura ambiental de entrada a la turbina no parece tener fundamento causal, es decir, no hay ninguna evidencia de que en el proceso químico de la formación de estos gases contaminantes en una turbina de gas influya el hecho de que el aire de entrada sea más frío o caliente. Las emisiones de NOX son más elevadas en los meses de invierno y más reducidas en los de verano, es por ello por lo que se encuentra una gran relación con la temperatura ambiental, pero dicha relación no es causa-efecto, sino que es más probable que tenga que ver con diferencias en el régimen de operación de la turbina entre estaciones. Habría que consultar con los técnicos responsables posibles causas para estos hechos (se utiliza una mezcla distinta, mayor producción en invierno, ...).

Por otro lado, respecto a la combustión incompleta, hemos visto que hay una posible relación entre temperatura ambiental (AT) baja y una temperatura de entrada (TIT) a la turbina baja. No obstante, también hemos visto que valores de TIT bajos se producen sin que ninguna otra variable anterior en el proceso tome valores anómalos. En general, se puede deducir que en muchas ocasiones el aire no se calienta lo suficiente en el compresor y que ello conlleva la combustión incompleta. Se debería investigar con los técnicos las causas por las que ocurre este hecho, ya que la compresión del aire (CDP) parece realizarse correctamente.

Por otra parte, hemos observado muchos datos en los que se produce combustión incompleta y altas emisiones de NOX, aunque la formación de este gas contaminante no está relacionada con dicho fallo. Además, valores de la salida de los gases demasiado bajos son indicativos de que la combustión no se ha realizado correctamente, aún con la variable TIT tomando valores altos. En ambos casos deducimos que, por un motivo que habría que aclarar, el proceso de combustión no se realiza correctamente. Con la metodología de Procesos por Lotes hemos visto que si la variable TIT toma valores iguales o superiores a los 1090°C y la variable TAT sobre los 545°C se reduciría en gran forma la posibilidad de que ocurriera un fallo por combustión incompleta.

Por último, con modelos PLS-TS hemos visto que las emisiones de CO y NOX están muy relacionadas con las emisiones en instantes anteriores, por lo que si se detectan emisiones fuera de control es importante parar el proceso para que no se sigan produciendo. Esta detección se puede conseguir con el modelo PCA creado para la monitorización del proceso. Además, el PLS-TS predice mejor la variable NOX que otros modelos vistos durante el máster.

BIBLIOGRAFÍA

- Ferrer Riquelme, A. J. (2025). *Análisis, monitorización y diagnóstico de procesos multivariantes* [Apuntes de clase]. Universidad Politécnica de Valencia.
- Kaya, H., Tüfekci, P., & Uzun, E. (2019). Predicting CO and NOx emissions from gas turbines: Novel data and a benchmark PEMS. *Turkish Journal of Electrical Engineering and Computer Sciences*: Vol. 27: No. 6, Article 53.
- Gas Turbine CO and NOx Emission Data Set [Dataset]. (2019). *UCI Machine Learning Repository*.
- R Core Team (2024); [<http://www.r-project.org/>]
- RStudio Team (2020); [<http://www.rstudio.com/>]
- Aspen Technology, Inc. (2025). *Aspen ProMV® [Software]*. AspenTech. <https://www.aspentech.com>

ANEXO

Empezaremos con gráficos de las emisiones de las dos variables de interés. En primer lugar, se mostrará un gráfico de la serie temporal de la variable CO.

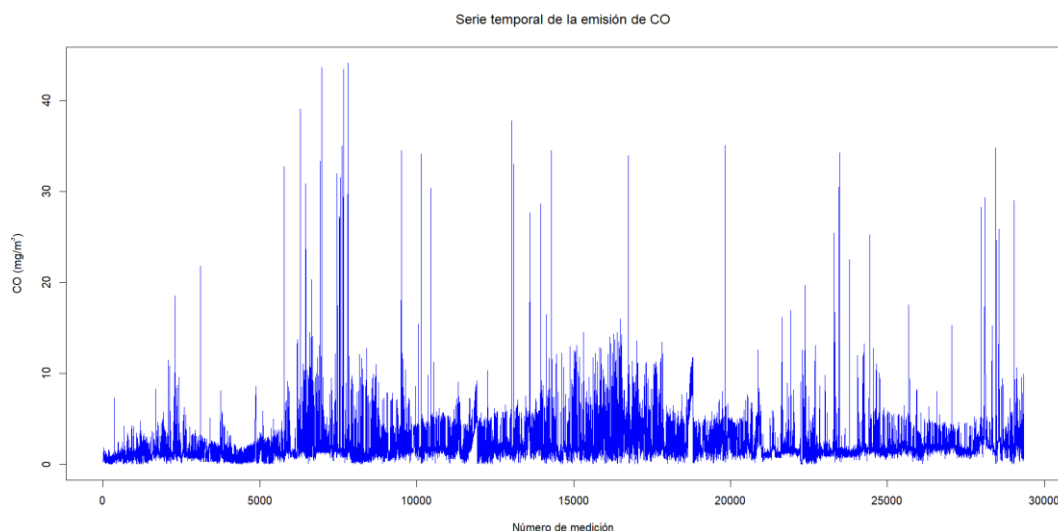


Figura 1A. Serie temporal de las emisiones de CO.

De la Figura 1A se puede deducir que las emisiones de monóxido de carbono por encima de los límites marcados por la Unión Europea no ocurren con demasiada frecuencia. Esto puede corroborarse con la media de 2.37 que presenta esta variable. Pasamos a observar ahora la serie temporal de las emisiones de NOX.

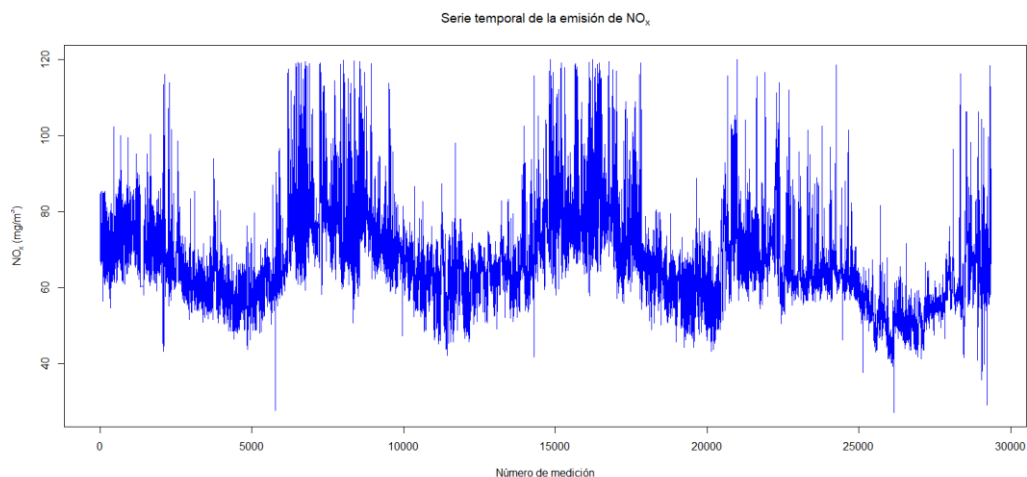


Figura 2A. Serie temporal de las emisiones de NOX.

Por otro lado, según la Figura 2A, las emisiones de NOX son más elevadas a lo largo del tiempo, y están por encima del límite de 41.8 que marca la Unión Europea. Además, se puede apreciar un patrón cíclico en las emisiones, que se observa con más detalle si se hacen gráficos de las emisiones por año.

Serie temporal de la emisión de NO_x en el 2013

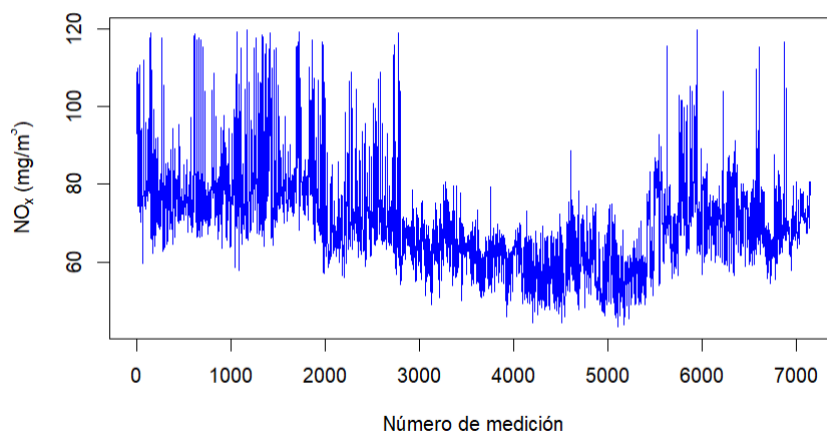


Figura 3A. Serie temporal de las emisiones de NOX en el año 2013.

Serie temporal de la emisión de NO_x en el 2014

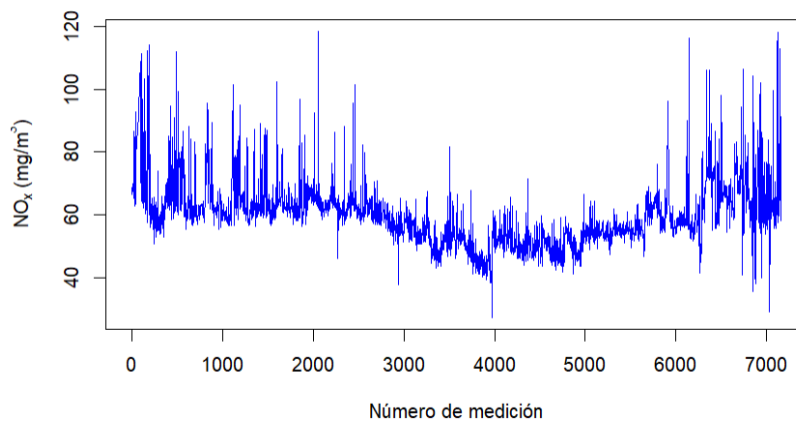


Figura 4A. Serie temporal de las emisiones de NOX en el año 2014.

Con carácter general, de las Figuras 3A y 4A se puede deducir que las emisiones de NOX presentan valores más elevados en las primeras y las últimas observaciones de cada año, correspondientes a la época de invierno.

A continuación, se mostrará el análisis de las relaciones entre variables que realizan los autores del artículo, que se basa en un análisis de coeficientes de correlación de Pearson de las variables respuesta con las variables explicativas, y entre éstas. En primer lugar, en la Figura 5A se muestra la matriz de correlaciones de las variables del proceso.

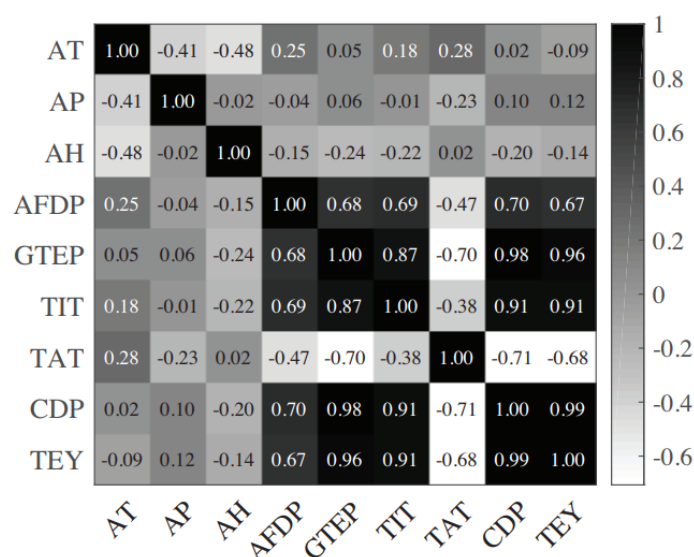


Figura 5A. Correlaciones de Pearson entre las variables explicativas.

A partir de la Figura 5A los autores del artículo destacan la alta correlación entre las variables CDP, GTEP y TEY. Además, indican que las variables de la turbina (AFDP, GTEP, TIT, TAT y CDP) están más correlacionadas con el rendimiento (TEY) que las variables ambientales (AT, AP, AH).

Por otro lado, se muestran los coeficientes de correlación de las emisiones de gases contaminantes con las variables del proceso.

Feature	CO	NO _x
AT	-0.174	-0.558
AP	0.067	0.192
AH	0.107	0.165
AFDP	-0.448	-0.188
GTEP	-0.519	-0.202
TIT	-0.706	-0.214
TAT	0.058	-0.093
CDP	-0.551	-0.171
TEY	-0.570	-0.116

Figura 6A. Correlaciones de Pearson entre las emisiones y las variables explicativas.

De esta Figura 6A los autores destacan la relación entre la variable TIT y la variable CO e indican que es conocido el hecho de que se generan mayores emisiones de CO cuando la temperatura de entrada a la turbina es baja y se produce una combustión incompleta. Por otro lado, se dice que la variable más correlacionada con la emisión de NOX es la temperatura ambiental, para la que se puede observar una correlación moderada (-0.558), lo que les hace indicar que trabajar con temperaturas más altas es más apropiado para reducir las emisiones de este gas contaminante.