

Informe Estudio para Ministerio del Medioambiente

Estudio de variables para modelo de predicción de SO_2

Alumno: Nicolás Cabello.
Andrés González F.
Consuelo Rojas N.
Curso: Proyecto de Ciencia de Datos
Profesores: Constanza Contreras P.
Francisco Forster B.
Fecha de entrega: 24 de diciembre de 2022
Santiago, Chile

Índice de Contenidos

| | |
|--|-----------|
| 1. Introducción | 1 |
| 2. Metodología | 2 |
| 3. Desarrollo | 3 |
| 3.1. Torres de Calidad de Aire | 3 |
| 3.2. CEMS Codelco | 8 |
| 3.3. Torre Meteorológica | 8 |
| 4. Conclusión | 11 |

Índice de Figuras

| | |
|--|---|
| 1. Matrices de correlación | 4 |
| 2. Series de tiempo normalizadas | 4 |
| 3. Conteo de peaks | 5 |
| 4. Gráficos de estacionalidad | 6 |
| 5. Cantidad de peaks según la hora | 7 |
| 6. Histograma de peaks según la hora | 7 |
| 7. Correlación SO_2 entre y otras variables, ordenadas de manera descendente | 8 |
| 8. Estabilidad Pasquill Gifford | 9 |

1. Introducción

En el siguiente informe se presenta el desarrollo del proyecto desarrollado en conjunto con el Ministerio del Medio Ambiente (MMA ¹ por sus siglas). El MMA es el órgano del Estado encargado de aplicar políticas, planes y programas en materia ambiental, así como en la protección y conservación de la diversidad biológica y de los recursos naturales renovables e hídricos, promoviendo el desarrollo sustentable, la integridad de la política ambiental y su regulación normativa. Entre sus intereses está el proteger y mitigar el impacto climático negativo en las distintas regiones del país.

Una de las zonas más conocidas por su contaminación a nivel nacional es la de Quintero - Puchuncaví. Esta zona, también llamada de sacrificio, presenta altos niveles de contaminantes nocivos para la salud de sus habitantes. En los últimos años se han producido eventos críticos de contaminación, con episodios de intoxicación en la población que afectan a la salud de las personas e impiden que se desenvuelvan con normalidad en su día a día. La contaminación es producida por el cordón industrial presente en la zona que emite diversos gases tóxicos, tales como el SO_2 , NO , NO_2 , entre otros. Abordar esta problemática es de gran interés por parte del Ministerio, ya que debe velar por el bienestar de la población en lo que concierne a temas climáticos.

El objetivo general de este proyecto es generar un sistema de pronóstico de concentración de dióxido de azufre (SO_2) en la zona de Quintero - Puchuncaví. Los objetivos específicos son caracterizar el comportamiento del contaminante, identificando las condiciones en las que se genera un peak. También relacionar datos meteorológicos de la zona e información de calidad del aire con los episodios del contaminante de interés.

Al ser este el primer acercamiento a este proyecto el alcance que se abarca es en la caracterización de los elementos que definen los peaks, verificando las condiciones con las cuales estos ocurren. Debido a lo complejo del problema y su carácter aparentemente aleatorio para la ocurrencia de peaks, se consideró que generar un modelo sin un profundo entendimiento previo de lo que ocurre en la zona resultaría en resultados poco útiles y sin la robustez necesaria para el gran impacto del problema. Entre los riesgos presentes se encuentra el no alcanzar una solución útil, llegar a conclusiones de que el sistema de pronóstico no es posible con las condiciones actuales, que los modelos generados no aporten a la solución del problema o que la solución no sea aceptada por lo ajena que es al área de estudio.

¹ Ministerio del Medio Ambiente <https://mma.gob.cl/estructura-organizacional/>

2. Metodología

La realización del estudio se hizo en dos partes. En la primera se utilizaron solo variables provenientes de las torres de observación de calidad del aire (SINCA ²). En específico, se utilizaron los datos correspondiente a los gases nocivos, como lo son SO_2 , NO , NO_2 , NO_x y O_3 . Estos datos corresponden a series de tiempo con una periodicidad de 1 hora y con una disponibilidad desde el año 2016, dependiendo de la estación.

Para el estudio, se hizo uso principalmente de las estaciones de Los Maitenes y Quintero, haciendo uso de datos auxiliares tale como las estaciones de Ventanas, Puchuncaví y las emisiones de las chimeneas de CODELCO.

El segundo estudio incorpora datos meteorológicos y de vientos de la zona. El fin de esta inclusión es ahondar en las relaciones existentes entre los gases nocivos junto a las condiciones del clima presentes en el sector. Para esto, se añadieron los datos provenientes de la torre meteorológica de CODELCO, que contiene datos de presión, temperatura, humedad y vientos a 10, 20 y 40 metros de altura.

En cuanto al análisis de los datos, se utilizaron las librerías de “Pandas”³ para el manejo general de los conjuntos de datos, “Matplotlib”⁴ para realizar visualizaciones y “Numpy”⁵ para el manejo puntual de algunos datos.

² Sistema de Información Nacional de Calidad del Aire <https://sinca.mma.gob.cl/>

³ Documentación Pandas: <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.html>

⁴ Documentación Matplotlib: <https://matplotlib.org/stable/index.html>

⁵ Documentación Numpy: <https://numpy.org/doc/stable/index.html>

3. Desarrollo

En la presente sección se describe el trabajo realizado, con sus resultados y correspondiente análisis. Para una mejor comprensión se dividirá en 3 secciones. La primera donde solo se utilizan los datos de las torres de calidad del aire, la segunda que incluye las chimeneas y luego la tercera con las variables meteorológicas.

Una parte importante del proyecto fue la obtención de la información. En general, esta estaba dispuesta de forma que mostraba tres tipos de registros distintos para cada gas: Registro validado, registro preliminar y registro no validado. De estos, se utilizó el registro no validado, dado que era la única columna que contaba con datos no nulos y por consulta a la contraparte, se indicó que para efectos prácticos estos datos se podía considerar como validados.

3.1. Torres de Calidad de Aire

Dentro de los primeros estudios realizados, se analizó correlación, temporalidades y formas de las series de datos. A continuación se presentan los resultados más importantes, mientras que las demás imágenes se encuentran en los Jupyter Notebooks adjuntos al proyecto.

En la imagen 1, se pueden ver las matrices de correlación de las variables de cada estación. En particular, se evidencia en ambas que existe una fuerte correlación entre los distintos óxidos de nitrógeno (NO , NO_2 y NO_X) con el dióxido de azufre (SO_2). Este comportamiento se ve claramente al momento de ver las series de tiempo de ambos gases.

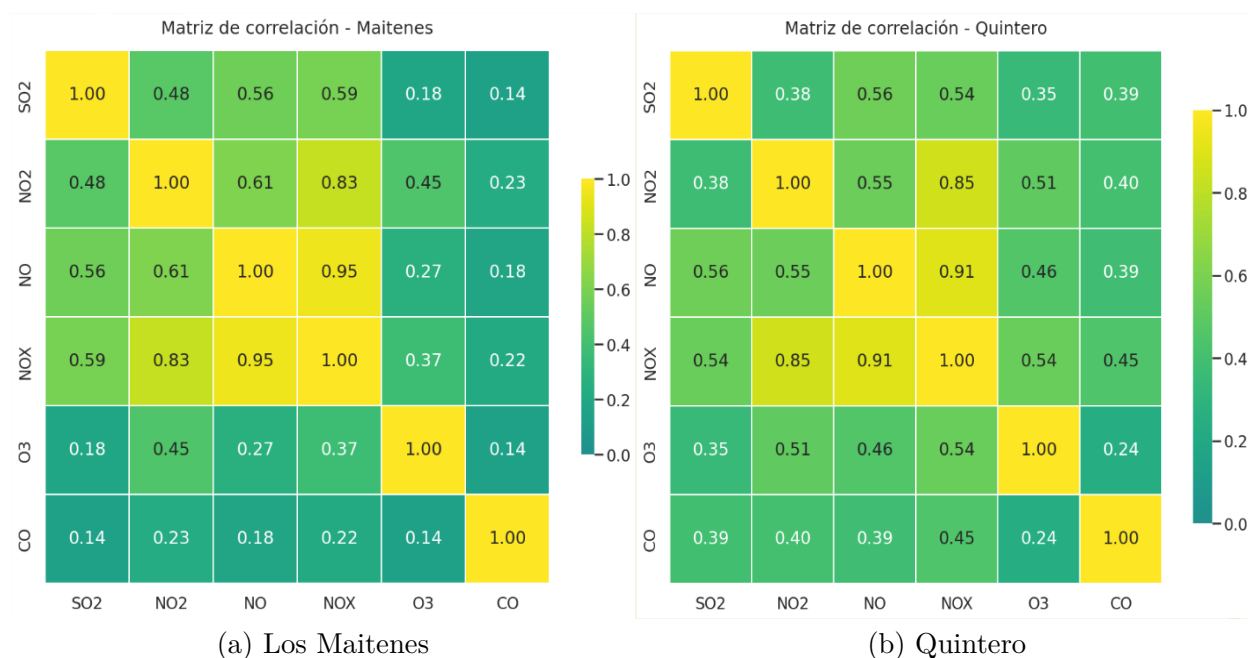


Figura 1: Matrices de correlación

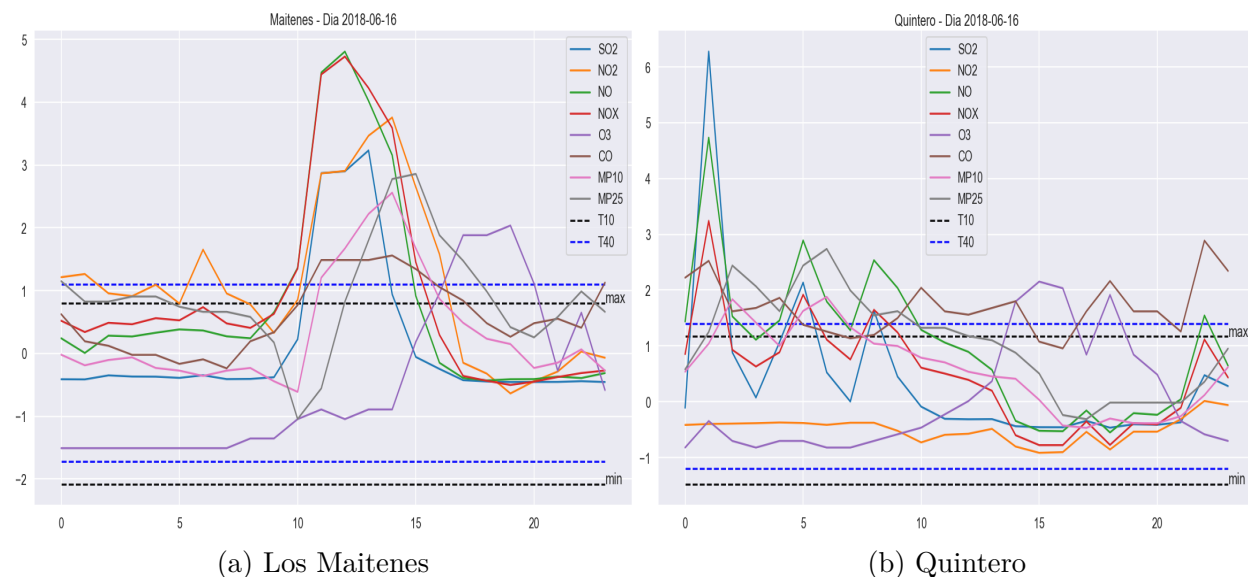


Figura 2: Series de tiempo normalizadas

En estos gráficos (figura 2), se ve que los peaks de SO_2 ocurren junto con los peaks de NO_X . Esto se puede deber a que ambas comparten origen, siendo ambas residuos de los mismos procesos químicos. Para este último punto, es necesario un estudio completo de la actividad industrial de la zona, teniendo en cuenta las emisiones de

cada proceso existente y así encontrar una relación útil entre componentes.

Continuando con el análisis, se prosigue con la información de peaks de contaminación. Para considerar un día como peak, se utilizó la normativa horaria de 133 ppb como máximo permitido. Por ende, se puede ver que la estación de Los Maitenes tiene un total del 124 peaks, en los últimos 5 años. En el mismo periodo, la estación de Quintero registró un total de 146 episodios de contaminación (Figura 3), mientras que la coincidencia de los días peaks en ambas estaciones es de solo 24 días. Esto sugiere la presencia de algún fenómeno climatológico subyacente, puesto que a pesar de que las estaciones sean cercanas geográficamente, la coincidencia de días en los que ocurren peaks es baja, lo cual contradice la intuición de pensar que los episodios ocurrirían de manera simultánea.

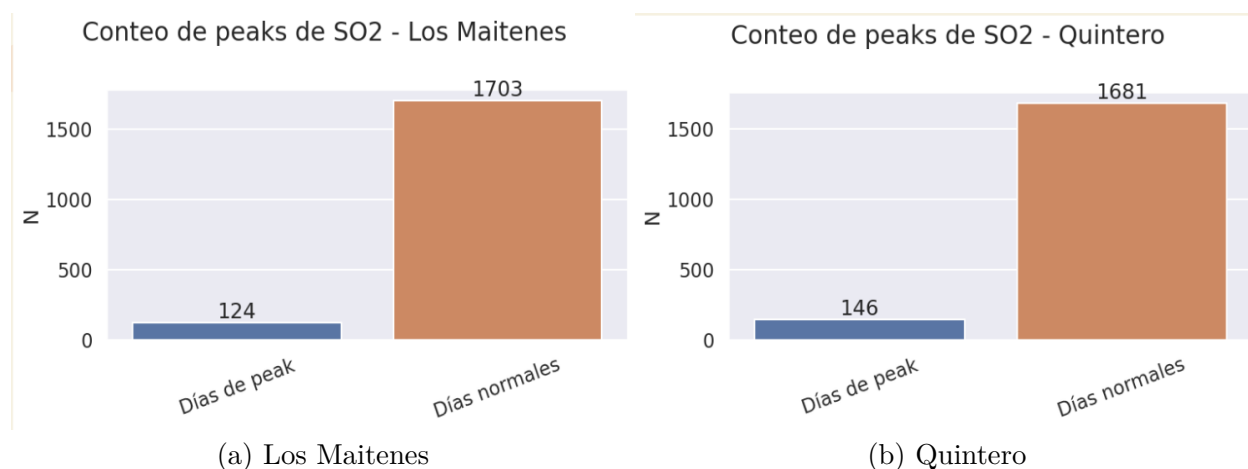


Figura 3: Conteo de peaks

Analizando las estacionalidad, se puede notar que Quintero tiene una fuerte tendencia a presentar episodios de contaminación durante las estaciones frías (Otoño - Invierno), mientras que en Los Maitenes esto no se aprecia de forma tan clara (Figura 4). Por otro lado, los días con mayores peaks de contaminación en Los Maitenes suceden los días martes y sábados, siendo esto no notorio para la estación de Quintero.

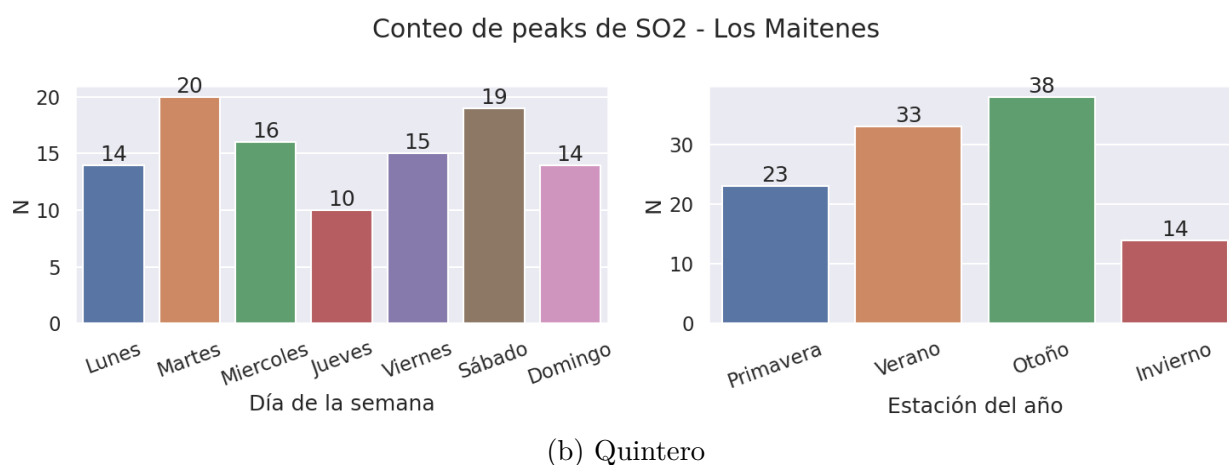
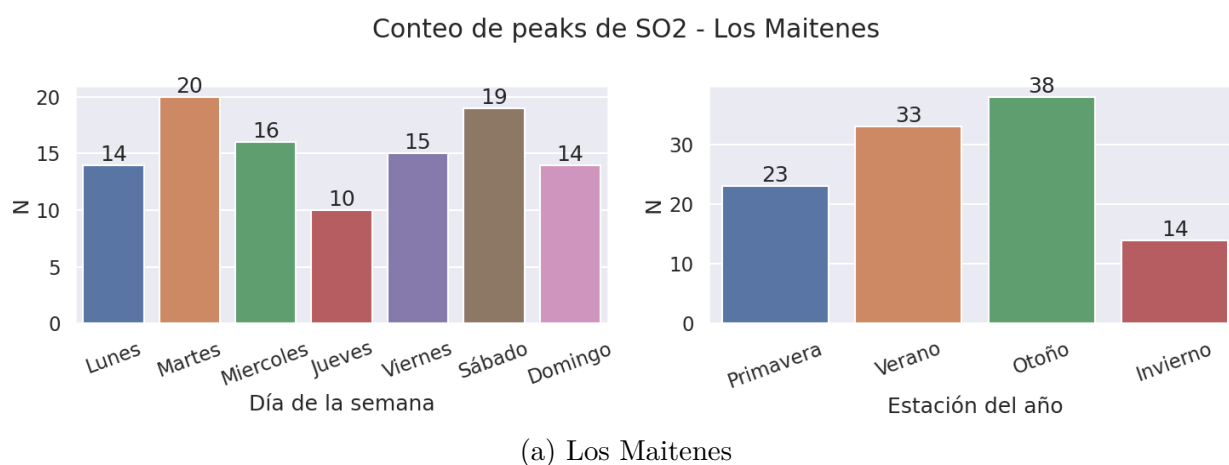


Figura 4: Gráficos de estacionalidad

Al momento de ver el comportamiento de estos momentos críticos, la gran mayoría de ellos tienen una duración de una hora que es la mínima duración dado que las observaciones disponibles están en escala horaria. Esto muestra que los episodios se “disuelven” rápido, esto tiene relación con la re-actividad de los gases emitidos, siendo estos fuertemente inestables y reaccionan fácilmente con el agua. Además, otras variables ambientales tales como el viento pueden hacer que estos gases se difundan más rápido.

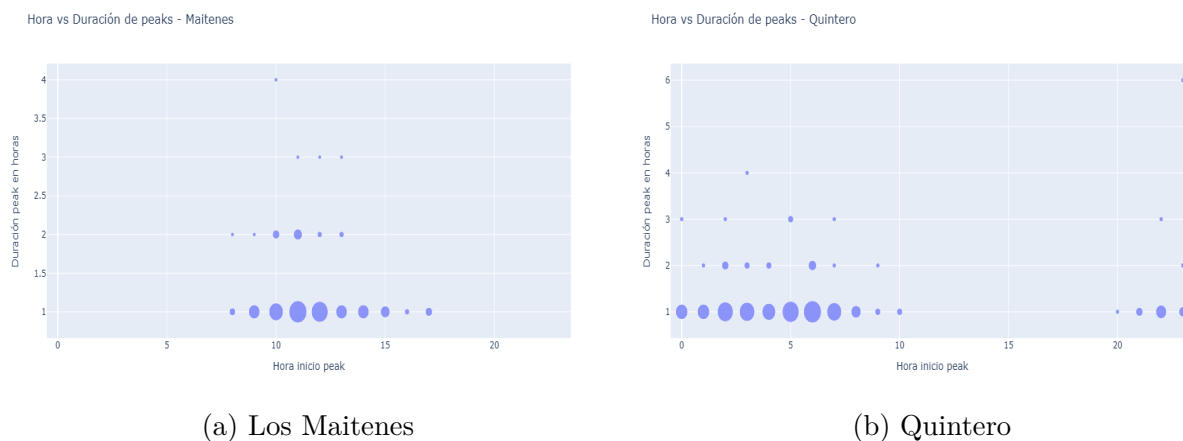


Figura 5: Cantidad de peaks según la hora

Para finalizar esta subsección, se analizaron las series de tiempo de las emisiones. En estas se puede ver que las horas más comunes para la ocurrencia de los episodios en Quintero son en la madrugada. Por parte de Los Maitenes, estos suelen suceder en horario laboral, es decir, entre las 10:00 hrs y las 20:00 hrs, aproximadamente. Una de las posibles razones de la ocurrencia de esto, en particular para Quintero, es la vaguada costera. Este fenómeno genera una disminución en la altura de la capa límite de la troposfera, lo que induce a la reducción del volumen en el cual los gases sean capaces de difundirse.

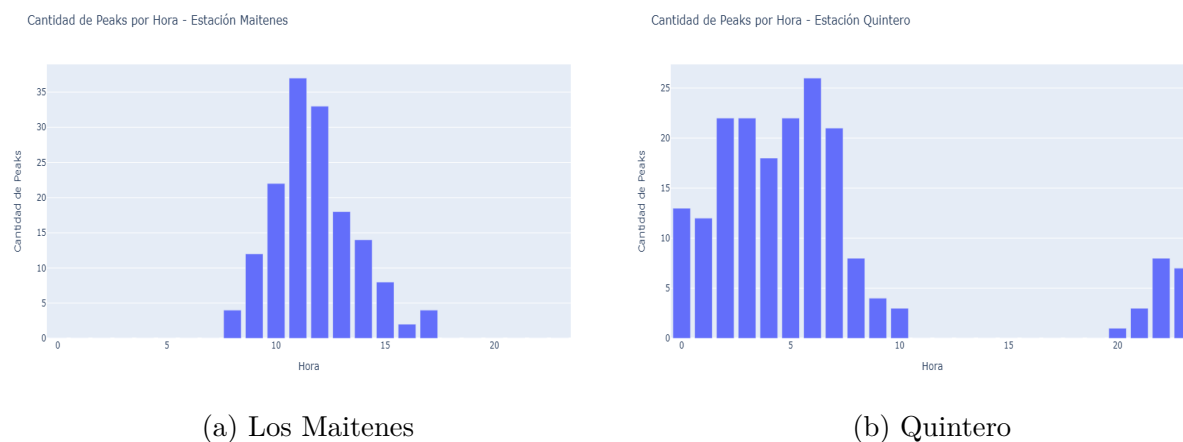


Figura 6: Histograma de peaks según la hora

Como los análisis anteriores no fueron realmente concluyentes, se hizo un estudio más profundo, incluyendo las variables meteorológicas. Este estudio se ve a continuación, luego de la adición de las variables provenientes de las chimeneas de CODELCO.

3.2. CEMS Codelco

Los datos provenientes de los CEMS se encuentran incompletos. Cuentan con una disponibilidad menor al 80 % con respecto a la ventana de tiempo analizada, por lo que las intersecciones de los datos eran menores y no superaban a los 6 meses de forma continua.

De los gráficos se puede concluir que las emisiones de las distintas chimeneas no tienen una relación particular con los días con peaks de contaminación. Esto se puede deber a que el problema se asocia más a condiciones meteorológicas y no en emisiones de las empresas.

3.3. Torre Meteorológica

El último paso fue la adición de las torres meteorológicas que se encuentra en CODELCO. Particularmente, esta tiene datos de temperaturas a distintas alturas, velocidades del viento a distintas alturas y dirección del mismo, junto con otras variables meteorológicas, como la presión y humedad relativa.

Se hizo nuevamente un estudio de correlación (Figura 7), en este se puede ver las variables más correlacionadas con el SO_2 , siguen siendo los óxidos de nitrógeno, sin embargo, ahora aparecen los vientos como variables importantes para el problema.

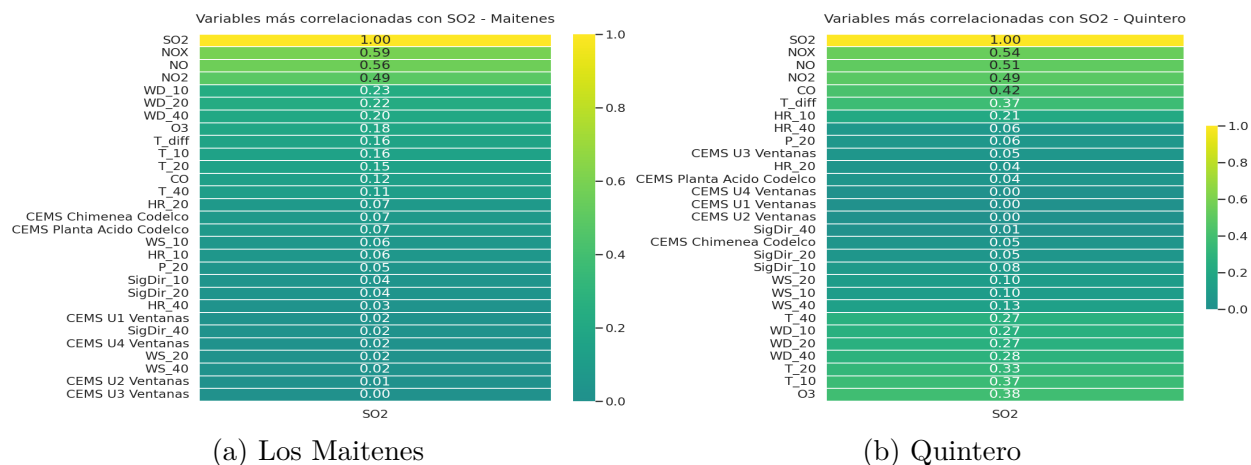
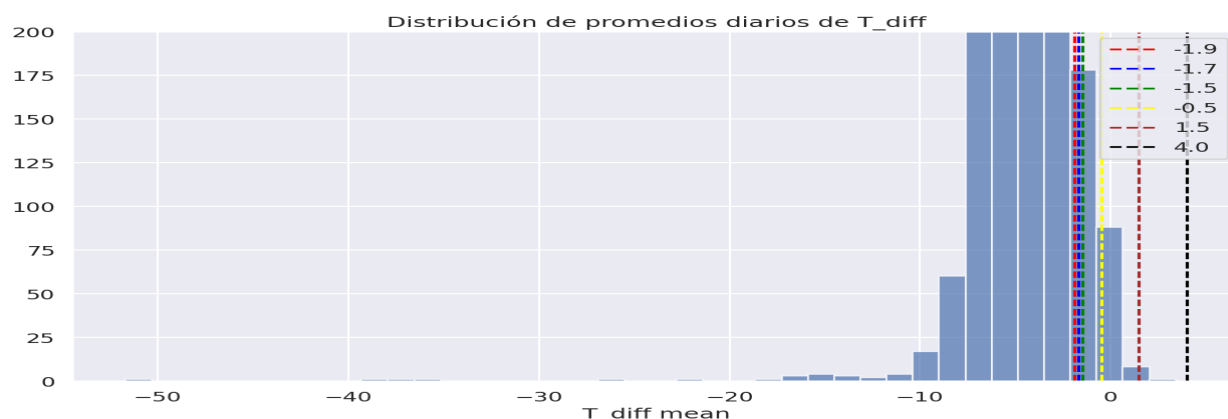


Figura 7: Correlación SO_2 entre y otras variables, ordenadas de manera descendente

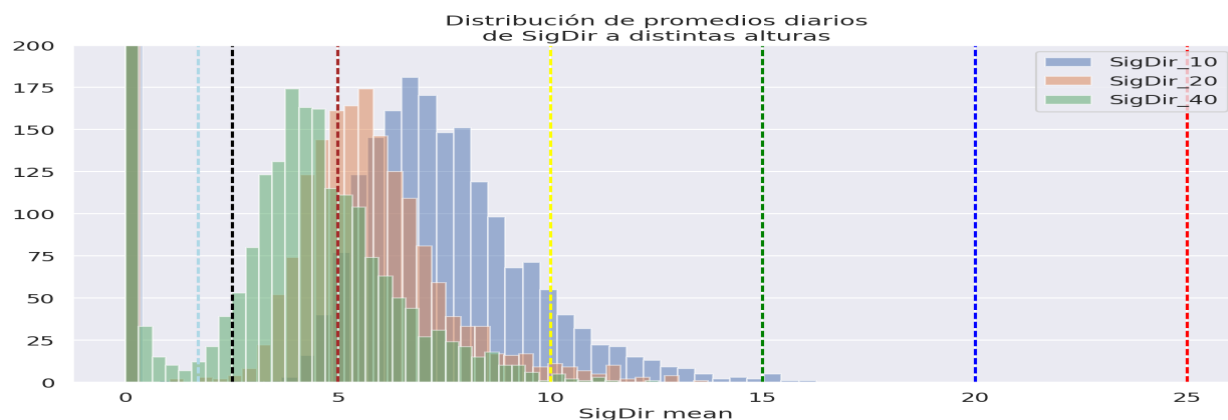
Con esto en mente, se hizo un análisis de estabilidad de la atmósfera en la capa

límite. Para esto se utilizó el modelo de Pasquill Gifford⁶, para ver que tipo de modelo es necesario aplicar.

De los datos obtenidos, se tiene que la gran mayoría de los días se concentran en una zona de inestabilidad atmosférica (Figura 8), lo que significa que los vientos tiene un papel protagonista en el rol de dispersión de contaminantes, esto se debe a que la zona geográfica es una zona de alta turbulencia, por lo que el movimiento de los vientos es caótico, y por ende, impredecible a grandes escalas de tiempo.



(a) Diferencia de temperatura



(b) Desviación estándar de la dirección del viento

Figura 8: Estabilidad Pasquill Gifford

Esta inestabilidad también se debe a que el gradiente ascendente de temperatura es mayor (Figura 8, a), lo que genera que el aire continúe subiendo en vez de bajar,

⁶ Modelo de dispersión de Pasquill Gifford https://www.upo.es/depa/webdex/quimfis/CA_old/php/apuntesCA0607_Tema2.pdf

como en un sistema adiabático. Esto genera un intercambio con gases de la zona superior de la troposfera, generando una capa límite inestable. Sin embargo, en los días de inversión térmica, se produce una estabilidad en las masas de aire. En específico, esta zona se puede relacionar con una inversión marina, que es la llegada de aire frío por el contacto con el agua a la costa. Esto produce un ambiente propicio para la mantención de los gases en la capa límite y por ende, aumentar su concentración.

Con esto se demuestra la necesidad de tener más información de vientos, para tener una idea a mayor escala de como se comportan las distintas masas de aire y sus interacciones, así se puede hacer una caracterización de los elementos claves en los días de contaminación.

4. Conclusión

Parte importante del trabajo realizado fue en la búsqueda de relaciones entre las distintas variables, esta búsqueda sugiere un acercamiento teórico al problema que complementa al estadístico. Por lo que se sugiere un mayor análisis de las variables meteorológicas incidentes.

Junto con lo anterior, se sugiere un monitoreo de nuevas variables, tales como la altura de la capa límite, para tener mayor información de la zona de difusión de los gases. Esto dado a que se cree que la mejor aproximación al sistema es una caracterización de los días con mayor probabilidad de peaks, considerando sus distintas variables, y no tan solo una predicción como se creía en primera instancia.

Además, es necesario considerar las distintas industrias de la zona, ver las emisiones de sus distintos procesos e identificarlos a través de sus emisiones. Esto con el fin de saber si las relaciones encontradas entre los óxidos y el dióxido de azufre sirven para el objetivo final, o son expulsados en conjunto.

Finalmente, se considera que el trabajo realizado aporta una visión estadística al problema, dejando claro que para encontrar una solución a este problema es necesario un trabajo en conjunto a expertos, dadas las condiciones meteorológicas difíciles de la zona.