

# Examen - Procesos Gaussianos y Puesta en Marcha

MA6202 Laboratorio de Ciencia de Datos

Otoño 2020

## 1 Introducción

La siguiente evaluación corresponde al examen del curso de laboratorio de ciencia de datos. A modo de contexto, el conjunto de datos a trabajar consiste en el reporte estadístico oficial de calidad del aire en Beijing. Este conjunto esta compuesto por registros de frecuencia horaria en 11 estaciones de monitoreo ubicadas en distintos puntos de la capital, la distribución de las estaciones puede apreciar en la figura (1). Los datos a utilizar están disponibles en este [link](#) [1].

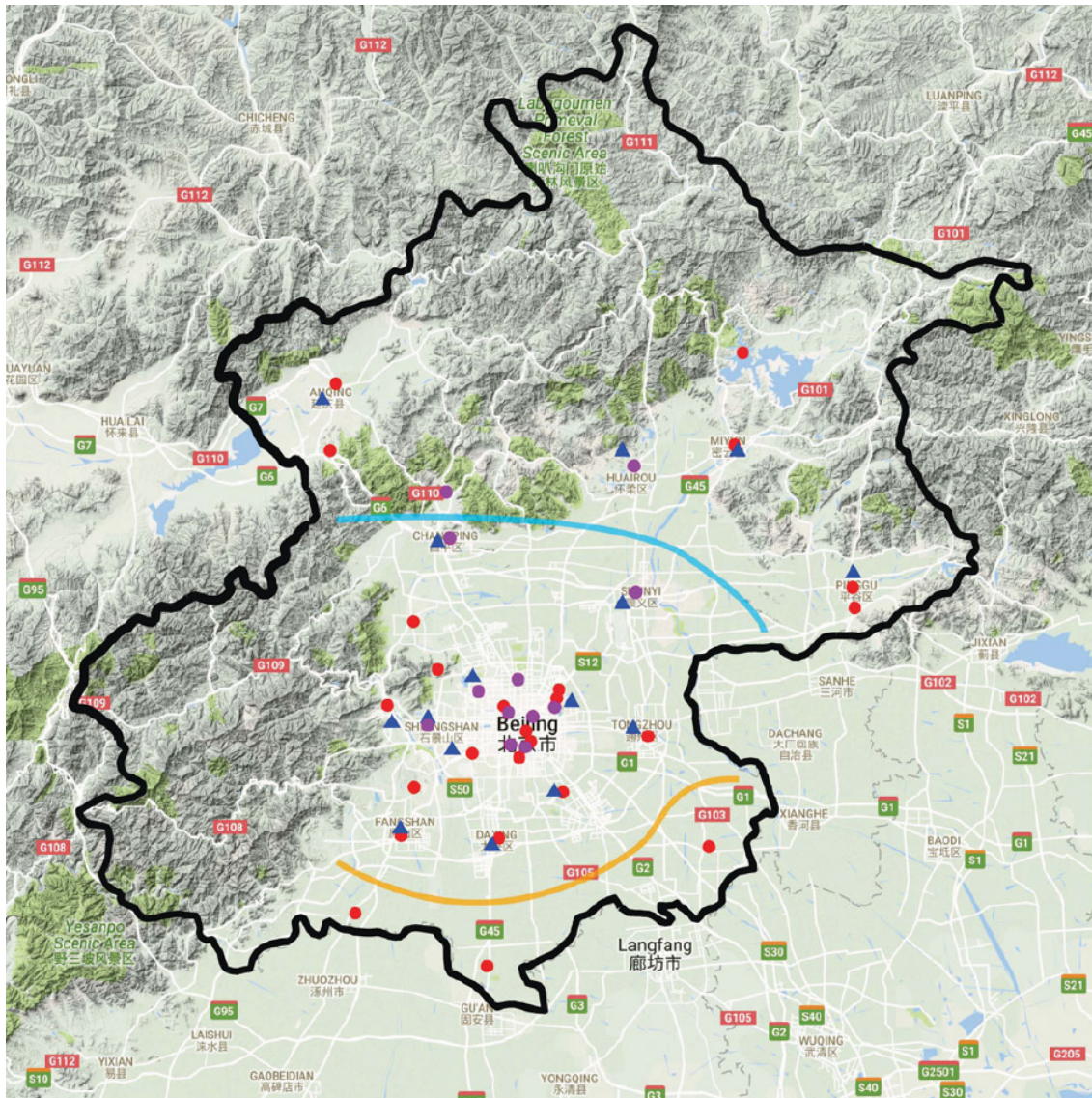


Figura 1: Estaciones de monitoreo en Beijing. los puntos violetas muestran su ubicación, los triángulos son estaciones meteorológicas y los puntos rojos son otras estaciones no incluidas en el conjunto de datos. Las líneas azul y naranjas dividen a la metrópolis en zona norte, centro y sur.

Se busca hacer una análisis de estos datos utilizando procesos gaussianos para finalmente generar una API con los modelos obtenidos. Esta evaluación es de carácter **individual**, se evaluará la presentación de sus resultados en dos modalidades. La primera por medio de un informe, cuyas condiciones de entrega son:

- La extensión máxima del informe es de 6 planas a las que puede añadir 2 para demostraciones.
- Debe adjuntar un repositorio `git` donde se incluya todo su código.
- A lo menos 1 `commit` por cada pregunta de la tarea
- Por lo menos 1 `merge` a través de su trabajo.
- Incluya un documento `jupyter notebook` llamado `examen.ipynb` en el cual se exponga todo el procedimiento realizado.
- Por último es necesario también entregar un archivo `.zip` con su modelo final.

La segunda opción de entrega es una presentación de sus resultados de manera remota a coordinar con el profesor. En esta modalidad no se requiere preparar un informe pero generar una presentación y mostrar sus resultados de investigación.

## P1. Carga y Exploración

En la presente sección se realizan los pasos de carga y limpieza de datos que permitirán realizar las secciones posteriores con un `DataFrame` consolidado que presente los tipos de datos adecuados para la información contenida en sus columnas.

Exponga en el reporte todas las decisiones que llevó en esta sección, además de reportar y discutir acerca de los aspectos específicos señalados en cada pregunta.

1. Convierta la información de las columnas `'year'`, `'month'`, `'day'` y `'hour'` en formato `datetime64`, mediante `pandas.to_datetime`. Añada una columna al `DataFrame` que contenga, para cada fila, el número de horas a partir de las 0 hrs del primero de marzo de 2013 (fecha inicial del conjunto de datos).

De ahora en adelante se entenderá por *columnas de contaminantes* a `'PM2.5'`, `'PM10'`, `'SO2'`, `'NO2'`, `'CO'` y `'O3'`.

2. Analice las correlaciones de presencia de valores faltantes para cada contaminantes en las diferentes estaciones. Genere visualizaciones y discuta lo observado.
3. En la columnas de contaminantes, rellene los valores faltantes mediante interpolación. Para ello utilice el método `pandas.DataFrame.interpolate`. Pruebe diferentes métodos de interpolación, seleccione uno y fundamente su elección. ¿Se puede justificar un esquema de llenado de valores faltantes con este método?
4. Analice las correlaciones para cada contaminantes en las diferentes estaciones. Genere visualizaciones, discuta lo observado y seleccione el contaminante que tenga el menor promedio de correlación entre estaciones. Para dicho contaminante, seleccione las tres estaciones que tengan el promedio de correlación más alto.
5. Mediante el función `statsmodels.tsa.seasonal.seasonal_decompose` descomponga en una tendencia, una señal periódica y su residuo; la señal del contaminante seleccionado en cada una de las 3 estaciones seleccionadas . Genere visualizaciones, discuta lo observado además de explicar el comportamiento de dicha función.

## P2. Modelación

## P3. Puesta en Marcha

## Referencias

- [1] Zhang, S., Guo, B., Dong, A., He, J., Xu, Z. and Chen, S.X. (2017) Cautionary Tales on Air-Quality Improvement in Beijing. Proceedings of the Royal Society A, Volume 473, No. 2205, Pages 20170457.