

Clasificación automatizada de imágenes coronográficas y análisis estadístico de interrupciones eléctricas

**Trabajo de Título para optar al título profesional de
Ingeniero Civil Matemático**

Sergio Nicolás Pincheira Tapia

Profesores Guía

Prof. Lisandro Fermín (Ingemat-UV)

Prof. Héctor Olivero Q. (Ingemat-UV)

Integrantes de la comisión evaluadora:

Cristian Meza (Ingemat-UV)

Afiliación

INSTITUTO DE INGENIERÍA MATEMÁTICA
VALPARAÍSO, FECHA.

Resumen

Las inspecciones coronográficas son un tema relevante en el ámbito de la electricidad, especialmente en la revisión del estado de los componentes en sistemas de distribución eléctrica. El enfoque normalmente usado para ejecutar este tipo de inspecciones es a través de cámaras especializadas para capturar imágenes coronográficas, las cuales posteriormente son analizadas y clasificadas. Sin embargo, al evaluar individualmente las imágenes se consume una gran cantidad de tiempo.

Dado este contexto, se realiza el presente trabajo en colaboración con Chilquinta el cual se centra en automatizar la clasificación de daños en las imágenes coronográficas y analizar un historial de interrupciones eléctricas mediante estimaciones en procesos de conteo con el fin de evaluar el posible impacto de las inspecciones.

Para esto, se realizan dos etapas en el desarrollo del proyecto. La primera es la automatización en la clasificación de imágenes coronográficas, se lleva a cabo realizando un preprocesamiento, aplicación de Reconocimiento Óptico de Caracteres (OCR) e implementación de un sistema automatizado de clasificación integrado en un desarrollo web. Por otro lado, tenemos la segunda etapa que consiste en analizar las interrupciones eléctricas y realizar estimaciones en base a procesos de Poisson homogéneos y no homogéneos con la finalidad de evaluar el posible impacto de las inspecciones coronográficas.

En cada etapa de desarrollo se muestran los resultados obtenidos. Se destaca que al emplear el sistema de clasificación automatizada en las imágenes de prueba se logran excelentes resultados utilizando el OCR apropiado, esto posibilita una clasificación precisa de todas las imágenes coronográficas disponibles. Por otro lado, en el análisis mediante estimaciones de procesos de conteo se obtiene información relevante en cuanto a las intensidades de las interrupciones eléctricas por zonas y sectores, además de evaluar el posible impacto que tienen las inspecciones coronográficas.

Agradecimientos

En primer lugar, quiero agradecerles a mis padres Maricarmen y Sergio por su comprensión y apoyo incondicional durante toda mi vida. Siempre dándome consejos y ejemplos vivos de que las adversidades pueden superarse.

Agradecerles a mis hermanas, partiendo por Camila, a la cual le debo un reconocimiento especial por su constante apoyo y cariño durante todo el tiempo que hemos estado juntos. También a mi hermanita Maite, quien es una de mis principales fuentes de felicidad y amor. Ambas han sido una inspiración para alcanzar mis metas.

A mi pareja Giselle, por su incondicional apoyo, comprensión y fe en mí durante todo este proceso. Estoy profundamente orgulloso de contar con tu presencia a mi lado, alentándome a perseverar en cada paso que decido dar.

También quiero expresar mi sincero agradecimiento a mis profesores de universidad, especialmente a mis profesores guía Héctor y Lisandro, quienes desempeñaron roles fundamentales en mi trayectoria académica.

Por último, quiero agradecer a mis compañeros de universidad con quienes tuve la oportunidad de compartir. En especial a Omar, quien fue un excelente compañero y amigo durante todo este proceso.

Índice general

Resumen	2
Algunas Palabras	4
1 Introducción	6
1.1 Contexto y formulación del problema	6
1.2 Objetivos	7
1.2.1 Objetivo general	7
1.2.2 Objetivos específicos	8
1.3 Hipótesis de Trabajo	8
2 Marco teórico	9
2.1 Coronografía	9
2.1.1 Factores atmosféricos	9
2.1.2 Factores eléctricos	10
2.1.3 Inspecciones coronográficas	11
2.1.4 Imágenes Coronográficas	11
2.1.5 Clasificación de nivel de daños	13
2.2 Reconocimiento óptico de caracteres	13
2.3 Procesos de conteo	15
2.3.1 Procesos de Poisson homogéneos	16
2.3.2 Procesos de Poisson no homogéneos	19
2.3.3 Test de homogeneidad para procesos de Poisson	25
3 Clasificación automática de imágenes coronográficas	29
3.1 Pruebas Iniciales	30
3.2 Preprocesamiento de Imágenes Coronográficas	32
3.3 Implementación de los OCR	34
3.3.1 Pytesseract	35
3.3.2 KerasOCR	35
3.3.3 EasyOCR	37

3.4	Automatización del OCR	37
3.5	Comparación de librerías de OCR	39
3.5.1	Análisis de resultados	40
3.6	Sistema automatizado e interfaz de usuario	41
3.7	Resultados de la clasificación de imágenes coronográficas	45
4	Análisis de Interrupciones eléctricas	47
4.1	Análisis Preliminares	48
4.2	Análisis de la Zona A	49
4.2.1	Sectores de la Zona A	54
4.2.2	Impacto de las coronografías en sectores homogéneos	58
4.2.3	Impacto de las coronografías en sectores no homogéneos	60
4.3	Análisis de la Zona B	63
4.3.1	Sectores de la Zona B	67
4.3.2	Impacto de las coronografías en sectores homogéneos	70
4.3.3	Impacto de las coronografías en sectores no homogéneos	72
4.4	Análisis de los resultados encontrados	73
5	Conclusiones generales	75
5.1	Clasificación automatizada de imágenes	75
5.2	Análisis de interrupciones eléctricas	76

Capítulo 1

Introducción

1.1. Contexto y formulación del problema

Este trabajo se desarrolló en colaboración con Chilquinta, una empresa de distribución de electricidad que opera mayormente en la región de Valparaíso. Su función principal es garantizar un suministro eléctrico confiable y seguro para los clientes de la región. Para lograr esto, Chilquinta actualmente realiza diferentes tipos de inspecciones periódicas a la infraestructura eléctrica. Entre estas inspecciones se encuentran las del tipo visual, termográfica y coronográfica con el objetivo de evaluar y mantener en óptimas condiciones todas sus componentes, asegurando así un servicio eficiente y de alta calidad.

La implementación de inspecciones coronográficas en la empresa es relativamente nueva, actualmente se llevan a cabo de manera bastante manual, ya que solo se cuenta con cámaras coronográficas para la captura de imágenes en los componentes del sistema. Para realizar una inspección coronográfica completa, se deben tratar tres diferentes etapas: la captura de imágenes de los componentes en el terreno, el análisis y clasificación de dichas imágenes, y la elaboración de informes detallados sobre la inspección realizada. Para la segunda etapa que corresponde al análisis y clasificación, los ingenieros deben analizar las imágenes y clasificarlas una por una basándose en la cantidad de descargas eléctricas medidas por la cámara utilizando el criterio establecido por la empresa. Por lo tanto, es razonable pensar en cómo minimizar el tiempo en cuanto a la clasificación y evaluar qué tan efectivas son las inspecciones de coronografía.

Para abordar la automatización y evaluación sobre qué tan efectivas son las inspec-

ciones coronográficas, el presente trabajo se realiza en dos etapas. La primera busca automatizar la clasificación de imágenes coronográficas utilizando el lenguaje de programación Python. En esta etapa se debe realizar un preprocesamiento a todas las imágenes para posteriormente aplicar Reconocimiento Óptimo de Caracteres (OCR) con el fin de extraer información numérica de las imágenes, con esta información se clasifican las imágenes utilizando el criterio establecido por la empresa. Además, se crea una interfaz de usuario sencilla y fácil de utilizar, la cual integra todo el algoritmo de clasificación automatizada.

La segunda etapa consiste en estimar mediante procesos de conteo la cantidad de interrupciones eléctricas por sector estudiado y poder visualizar si existe algún impacto de las inspecciones coronográficas sobre las interrupciones. Para realizar esto, primero se verifica la estacionalidad de los procesos para evitar tomar intervalos de tiempos no representativos antes y después de las inspecciones. Posteriormente, se verifica la homogeneidad mediante test de hipótesis para aplicar un enfoque adecuado a cada sector. Finalmente, para evaluar el impacto de las coronografías se utiliza un test de hipótesis en los procesos homogéneos y una estimación no paramétrica de las funciones de intensidad acumuladas junto con sus intervalos de confianza en los procesos no homogéneos.

El presente texto se estructura de la siguiente manera: en primer lugar, se presentan los conocimientos necesarios para comprender de manera técnica las inspecciones coronográficas, así como la base matemática para las estimaciones de las interrupciones eléctricas. Posteriormente, en el capítulo 3 se aborda la clasificación automática de imágenes coronográficas, mientras que en el capítulo 4 se desarrolla las estimaciones para los procesos de conteo, junto con un análisis del posible impacto de las inspecciones coronográficas. Finalmente en el capítulo 5 se exponen las conclusiones derivadas de los resultados obtenidos.

1.2. Objetivos

1.2.1. Objetivo general

El objetivo general de este trabajo es automatizar el proceso de clasificación de imágenes coronográficas y llevar a cabo un análisis del historial de interrupciones eléctricas para evaluar el potencial impacto de estas inspecciones preventivas.

1.2.2. Objetivos específicos

1. Procesar las imágenes coronográficas para resaltar y aislar la región de interés.
2. Evaluar y comparar los OCR utilizados en la región de interés, con el propósito de seleccionar el más efectivo y adecuado como base para el desarrollo de un sistema de clasificación automatizada.
3. Implementar una maqueta funcional del sistema automatizado en cuanto al procesamiento y clasificación de imágenes coronográficas.
4. Procesamiento sobre datos del historial de interrupciones eléctricas en los componentes eléctricos.
5. Analizar los procesos de interrupciones eléctricas por zonas, evaluando su homogeneidad y estacionalidad.
6. Analizar la estacionalidad del proceso de interrupciones eléctricas en todos los sectores, ya sean homogéneos o no homogéneos.
7. Evaluar el impacto de las coronografías en todos los sectores.

1.3. Hipótesis de Trabajo

En este proyecto se procede bajo las siguientes hipótesis:

- (H1) La toma de datos realizada por Chilquinta es sistemática y no tiene sesgos operacionales.
- (H2) Se dispone de una cantidad suficiente de datos para aplicar y validar las técnicas estadísticas que se utilizarán.
- (H3) La estructura de dependencia espacio temporal de los datos es compatible con las técnicas estadísticas que se utilizarán.

Capítulo 2

Marco teórico

2.1. Coronografía

El efecto corona son descargas eléctricas asociadas a la ionización del aire alrededor de conductores eléctricos deteriorados o dañados [13]. Las descargas son permanentes y acumulativas, tienen el potencial de deteriorar los aisladores y causar agrietamientos [9]. La Figura 2.1 exhibe visualmente las descargas eléctricas asociadas al efecto corona, identificadas por formas de color rojo.

En las líneas de transmisión eléctricas, el efecto corona se ha convertido en uno de los principales problemas asociados con las fallas en componentes eléctricos que afectan la continuidad del suministro [11]. Este fenómeno puede ser influenciado por varios factores que pueden potenciarlo y provocar fallas en los componentes eléctricos. Estos factores pueden agruparse en atmosféricos y eléctricos, como se detalla a continuación.

2.1.1. Factores atmosféricos

En el artículo de G. Sibilant [13] se explica que los factores atmosféricos desempeñan un papel importante en la manifestación y gravedad del efecto corona en líneas de transmisión de energía. Estos factores están relacionados con las condiciones atmosféricas del aire, como la temperatura, la presión, la humedad, las ionizaciones y el viento. En el siguiente listado se ven los diferentes impactos al efecto corona:

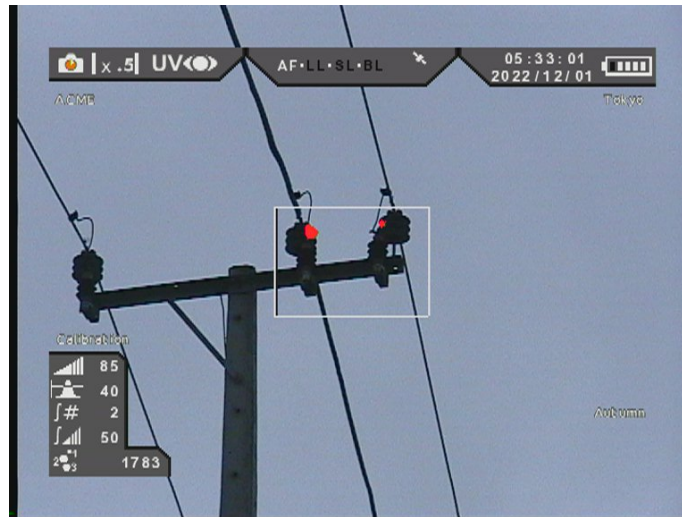


Figura 2.1: Efecto corona visible en imagen coronográfica.

- **Altitud:** La altura afecta debido a la densidad del aire en alturas, disminuyendo la tensión y aumentando la posibilidad del efecto corona.
- **Temperatura y presión:** La temperatura y la presión afectan la densidad del aire. Esta densidad influye en la facilidad con la que se forman descargas eléctricas alrededor de los conductores.
- **Humedad:** La humedad en el aire facilita la formación del efecto corona en líneas eléctricas, ya que las moléculas de agua ayudan en la ionización del aire.
- **Viento:** El viento dispersa electrones e iones, reduciendo las descargas eléctricas.

2.1.2. Factores eléctricos

Existen otro tipo de parámetros que dependen del proceso de diseño e intervención humana. Un ejemplo son los conductores, que se refiere a un componente utilizado para establecer una conexión eléctrica. A continuación se detallan algunos ejemplos:

- **Corriente eléctrica:** La circulación de corriente crea un campo magnético alrededor de los conductores, lo que puede provocar deterioro en los componentes aislantes. Este deterioro tiende a ser más pronunciado a tensiones altas.

- **Frecuencia:** La velocidad a la que la electricidad se propaga no suele ser un problema a menos que sea excesivamente rápida; esto rara vez ocurre en situaciones normales.
- **Tamaño del conductor:** Conductores más grandes distribuyen el campo eléctrico de manera más uniforme, reduciendo la probabilidad de descargas eléctricas, mientras que a conductores más pequeños ocurre lo contrario.

2.1.3. Inspecciones coronográficas

La coronografía es una técnica de inspección en ingeniería eléctrica diseñada para detectar y evaluar el efecto corona en equipos de alta y media tensión. Esta técnica permite examinar las líneas de transmisión de energía mediante imágenes que ofrecen una visualización de la actividad eléctrica en los componentes. Las imágenes son útiles para el mantenimiento predictivo, ya que facilitan la identificación temprana de áreas con anomalías eléctricas, agilizando así la implementación de medidas correctivas y preventivas para garantizar la continuidad del suministro eléctrico.

2.1.4. Imágenes Coronográficas

Las imágenes coronográficas se obtienen mediante uso de cámaras especializadas capaces de evidenciar la radiación ultravioleta. En este caso particular se utiliza la cámara CoroCam 6D. El proceso de captura se enfoca en detectar el efecto corona, proporcionando una representación visual de la actividad eléctrica y un conteo del número de descargas en el área que se está registrando.

En el centro de la Figura 2.2 se pueden apreciar pequeñas formas de color rojo, las cuales nos indican actividad eléctrica en la superficie de los componentes eléctricos. La intensidad del color rojo indica la magnitud de la actividad, permitiendo una primera evaluación visual de los equipos.

El análisis de estas imágenes se convierte en una tarea esencial del mantenimiento predictivo en sistemas eléctricos de media y alta tensión. La identificación temprana de áreas con actividad eléctrica anómalas, como el efecto corona, permite tomar medidas correctivas antes de que se produzcan fallos y cortes de suministro.

Las imágenes obtenidas a través de la cámara coronográfica CoroCam 6D presentan

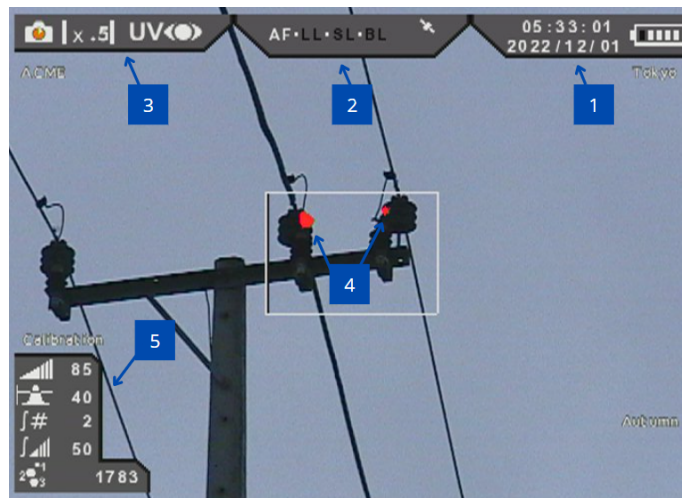


Figura 2.2: Elementos en imágenes coronográficas.

una serie de características que son fundamentales para el análisis y clasificación [14]. En la Figura 2.2 se ven números indicando los diferentes elementos que están presentes en las imágenes coronográficas, estos se detallan a continuación.

1. Fecha, hora y estado de la batería.
2. Enfoque automático (AF), la luz baja, entre otros.
3. El modo de cámara, nivel de zoom y canales activos.
4. Una superposición de color rojo que indica dónde los componentes están presentando descargas eléctricas.
5. Se aprecian cinco símbolos, cada uno con sus respectivos valores, que proporcionan información sobre ajustes y variables específicas: Ganancia=85, Umbral=40, Integración=2, Integración de la ganancia=50 y Recuento=1783

Es importante destacar que dentro de la región de interés, identificada con el número 5 y que se aprecia más claramente en la Figura 2.3, se encuentra una variable específica conocida como recuento, esta variable es la cantidad de descargas parciales. Dicha información nos permite clasificar la imagen coronográfica en función del nivel de daño, como se explica a continuación.

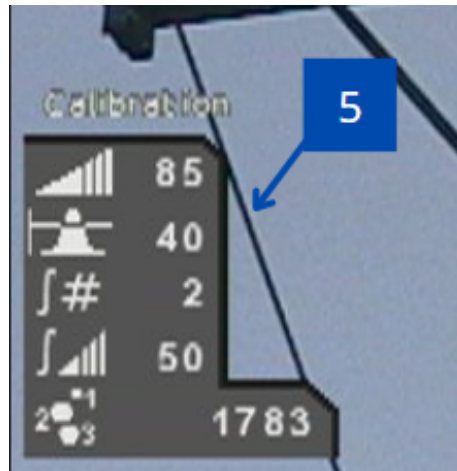


Figura 2.3: Región de interés en imágenes coronográficas.

2.1.5. Clasificación de nivel de daños

Chilquinta utiliza la cantidad de descargas parciales detectadas en las imágenes coronográficas para clasificar el nivel de daño presente. Esta clasificación es gestionada mediante un sistema de semáforo que clasifica los niveles de daños a través de colores y letras, como se observa en la Figura 2.4. El nivel más bajo de prioridad se representa con el color blanco (A), seguido por verde (B), amarillo (C) y finalmente, rojo (D) que indica el nivel de daño más alto posible.

2.2. Reconocimiento óptico de caracteres

La tecnología de Reconocimiento Óptico de Caracteres (abreviado en inglés como OCR) es una herramienta que reconoce automáticamente texto presente en imágenes. Admite diferentes tipos de formatos de imagen como JPG, PNG y archivos PDF. El OCR implica el análisis de los archivos y luego los traduce a formato de texto, de modo que la información se pueda almacenar y utilizarse en procesos automáticos [15]. Este último aspecto es especialmente relevante en el contexto de este proyecto.

El OCR tiene muchas aplicaciones, como extraer números específicos en imágenes, el reconocimiento de escritura a mano, extraer caracteres en patentes de autos, etc. Al aprovechar las capacidades automatizadas de extracción y almacenamiento de datos, la

Clasificación de nivel de daño	
A	Conteos menores que 500 eventos por segundo. Actividad de descarga benigna. No requiere acción.
B	Conteos entre 500 y 1000 eventos por segundo. La actividad de descarga puede estar disminuyendo la vida del componente. La actividad de descarga pueden generar un ruido audible (AN) o interferencias de radio (RI) alertar. Monitorear por mayor daño o dirección de la alerta AN o RI
C	Conteos entre 1000 y 5000 eventos por segundo. Dañado o componente degradado. Cambie o repare en la próxima oportunidad. El daño no afectará la seguridad o fiabilidad en un corto plazo.
D	Conteos mayores de 5000 eventos por segundo. El componente plantea un riesgo para la seguridad o fiabilidad del sistema en el corto plazo. Retirar del servicio lo antes posible.

Figura 2.4: Clasificación del nivel de daño presentes en las imágenes coronográficas.

tecnología OCR se convierte en un proceso eficaz que ahorra tiempo y otros recursos [4].

El proceso típico de OCR se puede explicar de manera sencilla en tres pasos, estos son: subir la imagen o archivo a analizar, aplicar el algoritmo de OCR específico y finalmente el resultado es la extracción de los caracteres reconocidos en forma de texto. Este proceso se puede adaptar a imágenes coronográficas obteniendo la Figura 2.5.

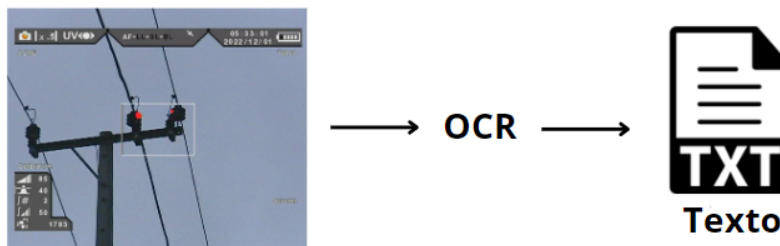


Figura 2.5: Proceso de OCR aplicado a imagen coronográfica.

En cuanto a herramientas de OCR se dispone de una gran variedad y lenguajes de programación para su implementación. En este trabajo se utilizará Python, el cual ofrece una serie de bibliotecas especializadas, entre estas, se encuentran: EasyOCR, KerasOCR y Pytesseract.

2.3. Procesos de conteo

A continuación, se busca describir el fenómeno de las interrupciones eléctricas a través de procesos estocásticos de conteo, los cuales pueden ser homogéneos o no. Para evaluar la homogeneidad de los procesos, se presentan tres test con sus respectivos estadísticos. Para una comprensión más detallada, se recomienda consultar [12, Capítulo 5] y [3].

Definición 1. Un proceso estocástico $(N(t), t \geq 0)$ es llamado proceso de conteo si $N(t)$ representa el número total de “eventos” que ocurren desde el tiempo 0 hasta t . Así, $N(t)$ satisface:

- (i) $N(t) \geq 0$.
- (ii) $N(t)$ es un valor entero.
- (iii) Si $s < t$, entonces $N(s) \leq N(t)$.
- (iv) Para $s < t$, $N(t) - N(s)$ es igual al número de eventos que ocurren en el intervalo de tiempo $(s, t]$.

Típicamente, se denota por $T_1, T_2, T_3, \dots, T_n$ a la secuencia de instantes aleatorios donde ocurren los eventos. En la Figura 2.6 se muestra como ejemplo una trayectoria de un proceso de conteo $N(t)$ simulada con sus tiempos de ocurrencia.

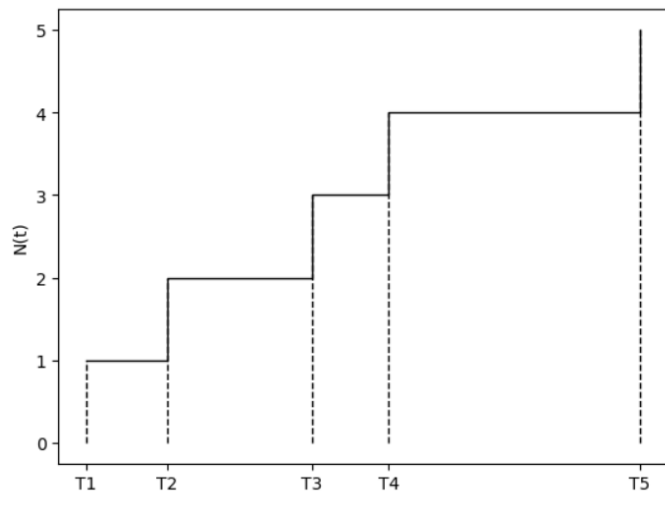


Figura 2.6: Trayectoria simulada de un proceso de conteo $N(t)$.

Un proceso de conteo con incrementos independientes es tal que el número de eventos que suceden en intervalos de tiempo disjuntos son independientes entre sí. Por ejemplo, esto implica que la cantidad de eventos que ocurren en el intervalo $(0, 10]$, es decir $N(10)$, no afecta ni está influenciado por la cantidad de eventos que ocurren en un intervalo posterior, digamos en el intervalo $(10, 15]$, es decir $N(15) - N(10)$.

En algunas situaciones de eventos aleatorios, asumir incrementos independientes podría resultar poco realista. Por ejemplo, si ocurren múltiples eventos de interrupciones eléctricas debido a la falla de un componente eléctrico crucial, es posible que estos eventos estén relacionados entre sí y no sean independientes. Sin embargo, si se considera que tales eventos son poco frecuentes, podría ser válido tratarlos como independientes para simplificar el análisis.

Un proceso de conteo se dice que posee incrementos estacionarios si la distribución del número de eventos que ocurren en cualquier intervalo de tiempo depende únicamente de la longitud del intervalo de tiempo. En otras palabras, el proceso tiene incrementos estacionarios si el número de eventos en el intervalo $(s, s + t]$, es decir $N(t + s) - N(s)$, tiene la misma distribución para todos los valores de s . En este caso se tiene en particular que $N(t + s) - N(s)$ tiene la misma distribución de probabilidad que $N(t) - N(0)$.

La consideración de incrementos estacionarios suele ser útil en situaciones donde los eventos ocurren con una intensidad constante a lo largo del tiempo. Esto quiere decir que la probabilidad de que ocurra un evento en cualquier intervalo de tiempo dado es constante.

Un caso común en el que se asumen incrementos independientes y estacionarios es en procesos de Poisson homogéneos, también abreviados como PPH. Estos procesos son utilizados para modelar situaciones donde los eventos ocurren de manera aleatoria en el tiempo, pero con una tasa constante promedio, como se explica a continuación.

2.3.1. Procesos de Poisson homogéneos

Definición 2. El proceso de conteo $\{N(t), t \geq 0\}$ se dice que es un proceso de Poisson homogéneo con tasa $\lambda, \lambda > 0$, si:

- (i) $N(0) = 0$.
- (ii) El proceso tiene incrementos independientes.

(iii) El número de eventos en cualquier intervalo de longitud t sigue una distribución de Poisson con media λt . Es decir, para todo $s, t \geq 0$

$$P\{N(t+s) - N(s) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, \dots$$

Se sigue de la condición (iii) que un PPH tiene incrementos estacionarios y también que

$$E[N(t)] = \lambda t$$

lo cual explica por qué λ es llamada la tasa del proceso.

Para determinar si un proceso de conteo es un PPH, debemos demostrar que se cumplen las condiciones (i), (ii) y (iii). La condición (i), simplemente establece que el conteo de eventos comienza en el tiempo ($t = 0$), y la condición (ii) generalmente se puede verificar directamente a partir de nuestro conocimiento del proceso. Sin embargo, no está del todo claro cómo determinar si se cumple la condición (iii), y por esta razón es útil una definición equivalente de un proceso de Poisson.

Antes de presentar una segunda definición de un PPH, se recuerda el concepto de “o pequeña de h ”, denotado por $o(h)$.

Definición 3. La función $f : \mathbb{R} \rightarrow \mathbb{R}$ se dice un $o(h)$ en 0 si:

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0$$

Teniendo esto en conocimiento, se puede presentar una definición alternativa a un PPH.

Definición 4. El proceso de conteo $\{N(t), t \geq 0\}$ se dice que es un proceso de Poisson homogéneo con tasa $\lambda, \lambda > 0$, si:

(i) $N(0) = 0$.

(ii) El proceso tiene incrementos estacionarios e independientes.

(iii) $P\{N(h) = 1\} = \lambda h + o(h)$.

(vi) $P\{N(h) \geq 2\} = o(h)$.

Es relevante destacar que la equivalencia entre las definiciones de PPH presentadas anteriormente se demuestra en el Teorema 5.1 de [12, Capítulo 5, Sección 3].

La demostración se basa en utilizar la propiedad de incrementos independientes y estacionarios del proceso de Poisson. Esto permite derivar una ecuación diferencial que relaciona una función calculada con la tasa de eventos λ .

Si consideramos un PPH y denotamos como τ_1 el tiempo entre el primer evento y 0. Además, para $n > 1$, sea τ_n el tiempo transcurrido entre el $(n - 1)$ -ésimo y el n -ésimo evento. La secuencia $\{\tau_n, n = 1, 2, \dots\}$ se denomina secuencia de tiempos entre llegadas. Para estos eventos se cumple la siguiente propiedad.

Proposición 1. *Para $n = 1, 2, \dots$, los τ_n son variables aleatorias exponenciales independientes e idénticamente distribuidas con media $\frac{1}{\lambda}$.*

La suposición de incrementos estacionarios e independientes es básicamente equivalente a afirmar que, en cualquier momento dado, el proceso se reinicia probabilísticamente. Es decir, el proceso a partir de cualquier punto es independiente de todo lo que ha ocurrido previamente (por incrementos independientes) y también tiene la misma distribución que el proceso original (por incrementos estacionarios). En otras palabras, el proceso no tiene memoria y, por lo tanto, es de esperar que los tiempos de llegada tengan distribución exponencial.

Por otro lado, la cantidad de interrupciones que suceden en líneas de distribución eléctrica, naturalmente, se puede modelar con un proceso de conteo. En la Figura 2.7 se ve la trayectoria $N(t)$ durante 300 días asociada al proceso de conteo en un sector de Valparaíso, el cual denominaremos Sector 1A. Sin embargo, si observamos la trayectoria del proceso $N(t)$ es evidente que hay días con periodos de pocos eventos, como en el intervalo aproximado entre los días 0 y 100, mientras que en otros periodos como entre los días 150 y 200 se observan muchas más ocurrencias. Utilizar un proceso de Poisson para modelar esta trayectoria no sería ideal, ya que la tasa de ocurrencias parece no ser constante a lo largo del tiempo.

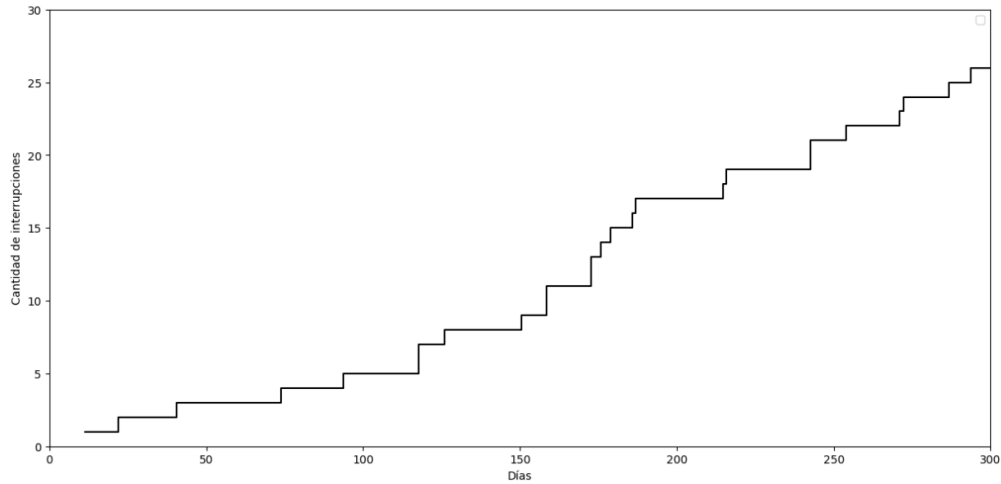


Figura 2.7: Conteo empírico de interrupciones eléctricas en el Sector 1A en 300 días.

Dada la variabilidad en la tasa de ocurrencia a lo largo del tiempo que observamos en el proceso $N(t)$ surge la necesidad de considerar un modelo más flexible. Una opción son los procesos de Poisson no homogéneos abreviados como PPNH, donde la tasa de ocurrencia varía en función del tiempo. Esto permite adaptar el modelo a la naturaleza cambiante de los datos y capturar de manera más precisa las tendencias en la ocurrencia de eventos a lo largo del tiempo.

2.3.2. Procesos de Poisson no homogéneos

Como se vio anteriormente, en ocasiones no es realista suponer que la intensidad de ocurrencia de eventos en un proceso de conteo sea constante. En la práctica, esta intensidad suele depender de la variable t . Por ejemplo, la tasa promedio de interrupciones eléctricas en una línea de distribución puede no ser la misma durante todo un año. Para abordar la variabilidad de la intensidad se presenta la definición de un PPNH.

Definición 5. El proceso de conteo $\{N(t), t \geq 0\}$ se dice que es un proceso de Poisson no homogéneo con función de intensidad $\lambda(\cdot)$, $\lambda > 0$, si:

- (i) $N(0) = 0$.
- (ii) El proceso tiene incrementos independientes.

$$(iii) P[N(t+h) - N(t) = 1] = \lambda(t)h + o(h),$$

$$(iv) P[N(t+h) - N(t) \geq 2] = o(h).$$

Observación. La condición (iii) implica que el proceso $\{N(t), t \geq 0\}$ no tiene incrementos estacionarios a menos que $\lambda(t) \equiv \lambda > 0$. En este caso, $\{N(t), t \geq 0\}$ se convierte en un proceso de Poisson homogéneo, con tasa λ .

Como en el caso particular cuando la tasa media de llegada de eventos es constante, encontramos que el número de eventos que ocurren en un intervalo dado tiene una distribución de Poisson.

Definición 6. Asociada a la función de intensidad de un PPNH se define la función de intensidad acumulada como

$$\Lambda = (\Lambda(t), t \geq 0) \quad (2.1)$$

donde

$$\Lambda(t) = \int_0^t \lambda(s) ds \quad \forall t \geq 0 \quad (2.2)$$

A continuación se presenta una proposición enunciada por Lefevbre en [8, Capítulo 5]

Proposición 2. Sea $\{N(t), t \geq 0\}$ un proceso de Poisson no homogéneo con función de intensidad $\lambda(t)$. Tenemos

$$N(s+t) - N(s) \sim \text{Poisson}(\Lambda(s+t) - \Lambda(s)) \quad \forall s, t \geq 0$$

donde

$$\Lambda(s) := \int_0^s \lambda(\tau) d\tau$$

Para estimar la variabilidad de λ se puede utilizar un enfoque basado en un PPNH con intensidad constante por tramos. En este enfoque se divide el tiempo en varios intervalos y consideramos una distribución de Poisson con parámetro λ_i en cada uno, donde λ_i representa la tasa de ocurrencia en el i -ésimo intervalo de tiempo.

Ahora, introduzcamos y desarrollemos el cálculo del estimador de máxima verosimilitud (MLE) de λ con k trayectorias para esta situación. Dado un PPNH con intensidad constante en tramos, consideremos k trayectorias del proceso observadas durante un

intervalo de tiempo específico $[a, b]$. Sea N_i el número de eventos observados en la trayectoria i , donde $i = 1, 2, \dots, k$.

La función de verosimilitud conjunta para todas las trayectorias estaría dada por el producto de las funciones de verosimilitud individuales, que son distribuciones de Poisson con parámetros correspondientes a las tasas de llegada en cada tramo:

$$L(\lambda) = \prod_{i=1}^k \frac{e^{-\lambda(b-a)} (\lambda(b-a))^{N_i}}{N_i!}$$

Tomando el logaritmo natural de esta función de verosimilitud conjunta:

$$\ln L(\lambda) = -k(b-a) + \sum_{i=1}^k N_i \ln(\lambda(b-a)) - \sum_{i=1}^k \ln(N_i!)$$

Para encontrar el estimador de máxima verosimilitud $\hat{\lambda}$, derivamos $\ln L(\lambda)$ con respecto a λ y la igualamos a cero:

$$\frac{d}{d\lambda} (\ln L(\lambda)) = -k + \frac{\sum_{i=1}^k N_i}{\lambda(b-a)} = 0$$

Resolviendo para λ , obtenemos:

$$\hat{\lambda} = \frac{\sum_{i=1}^k N_i}{k(b-a)}$$

Este es el estimador de máxima verosimilitud de la tasa de llegada λ en el intervalo $[a, b]$ considerando k trayectorias del PPNH. Es simplemente el promedio del número total de eventos observados en todas las trayectorias durante el intervalo de tiempo dividido por la duración del intervalo.

Teniendo esto en cuenta, se presenta una estimación no paramétrica para la función de intensidad e intensidad acumulada de un PPNH.

Estimación para la función de intensidad

Para las estimaciones de las intensidades de un PPNH, se sigue el enfoque propuesto por Leemis debido a la semejanza con el problema que trata en su artículo [7] y la naturaleza de los datos que se analizan en este trabajo.

Para comenzar, dividimos el horizonte temporal de interés $[0, T]$ en subintervalos de longitud uniforme. Es importante destacar que aunque los estimadores propuestos por Leemis no requieren subintervalos equiespaciados, en este trabajo optamos por utilizar subintervalos de igual longitud. Esta elección es útil, ya que permite comparar trayectorias de manera más directa y obtener una estimación global por zonas del proceso de manera más sencilla.

Sea $n_1, n_2, \dots, n_m$ el número total de eventos ocurridos en cada uno de los m subintervalos para k realizaciones del proceso. Suponemos que el proceso de Poisson no homogéneo (PPNH) tiene una función de intensidad $\lambda(t)$ que es constante a tramos en cada subintervalo de la partición $(a_0, a_1], (a_1, a_2], \dots, (a_{m-1}, a_m]$.

Dado que la función de intensidad promedio en el intervalo $(a_{i-1}, a_i]$ representa la tasa de eventos por unidad de tiempo en ese intervalo, el estimador de máxima verosimilitud es el número promedio de eventos que ocurren en dicho intervalo, dividido por la longitud de ese intervalo, tal como se calculó anteriormente. Entonces, el estimador de máxima verosimilitud para la función de intensidad en el intervalo $(a_{i-1}, a_i]$, para k realizaciones del proceso, se expresa como:

$$\hat{\lambda}(t) = \frac{n_i}{k(a_i - a_{i-1})}, \quad a_{i-1} < t \leq a_i \quad (2.3)$$

para $i = 1, 2, \dots, m$

Estimación para la función de intensidad acumulada

La función de intensidad acumulada $\Lambda(t)$ representa la cantidad acumulada de eventos hasta el tiempo t . Para obtener esta función, necesitamos sumar el número de eventos ocurridos hasta el final de cada subintervalo y luego agregar la contribución de los eventos en el subintervalo actual, ponderada por el tiempo transcurrido dentro de ese subintervalo.

El número acumulado de eventos hasta el final del subintervalo $(a_i, a_{i+1}]$ con $i = 1, 2, \dots, m$ se calcula sumando el número de eventos ocurridos en todos los subintervalos anteriores. Por lo tanto, para obtener la función de intensidad acumulada $\Lambda_k(t)$, para las k realizaciones del proceso, en cada subintervalo $(a_{i-1}, a_i]$, se deben realizar los siguientes pasos.

1. Sumar los eventos acumulados hasta el final del subintervalo anterior:

$$\sum_{j=1}^{i-1} \frac{n_j}{k}$$

donde $\frac{n_j}{k}$ es el número promedio de eventos por unidad de tiempo en el subintervalo j .

2. Agregar la contribución de los eventos en el subintervalo actual:

$$\frac{n_i (t - a_{i-1})}{k (a_i - a_{i-1})}$$

- n_i es el número de eventos ocurridos en el subintervalo actual.
- $(t - a_{i-1})$ es la longitud del tiempo transcurrido dentro del subintervalo actual.
- $k(a_i - a_{i-1})$ es la longitud total del subintervalo por la cantidad de trayectorias del proceso.

Por lo tanto, la función estimada de intensidad acumulada para k realizaciones es

$$\hat{\Lambda}(t) = \left(\sum_{j=1}^{i-1} \frac{n_j}{k} \right) + \frac{n_i (t - a_{i-1})}{k (a_i - a_{i-1})}, \quad a_{i-1} < t \leq a_i \quad (2.4)$$

para $i = 1, 2, \dots, m$. Esta función de intensidad es lineal a tramos y continua en el intervalo $(0, T]$.

Notemos que si no existen eventos en un intervalo i ($n_i = 0$) entonces la función de intensidad estimada en ese intervalo es 0 y la función acumulada estimada permanece constante.

Intervalos de confianza

En el artículo [7] Leemis propone un intervalo de confianza para el estimador $\hat{\Lambda}(t)$ el cual es asintóticamente exacto en los extremos de cada subintervalo.

La función de intensidad poblacional $\lambda(t)$ en la mayoría de las aplicaciones no suele ser constante a tramos en cada subintervalo $(a_{i-1}, a_i]$ de la partición arbitraria de $(0, T]$. Como plantea Leemis, para el caso general, el estimador $\hat{\Lambda}(t)$ solo es consistente en los puntos extremos de cada intervalo a medida que $k \rightarrow \infty$, siendo k la cantidad de realizaciones observadas. Esto es,

$$\lim_{k \rightarrow \infty} \hat{\Lambda}(a_i) = \Lambda(a_i),$$

para $i = 1, 2, \dots, m$.

Por otro lado, Leemis prueba que el estimador $\hat{\Lambda}(t)$ converge en distribución a una distribución normal en los puntos a_i , $i = 1, \dots, m$. Específicamente, se tiene

$$\lim_{k \rightarrow \infty} P \left(\hat{\Lambda}(a_i) - z_{\alpha/2} \sqrt{\frac{\hat{\Lambda}(a_i)}{k}} < \Lambda(a_i) < \hat{\Lambda}(a_i) + z_{\alpha/2} \sqrt{\frac{\hat{\Lambda}(a_i)}{k}} \right) = 1 - \alpha \quad (2.5)$$

para $i = 0, 1, \dots, m$.

De esta manera, tenemos que un intervalo de confianza estimado al $(1 - \alpha) \times 100\%$ para la cantidad $\Lambda(t)$ es

$$\hat{\Lambda}(t) - z_{\alpha/2} \sqrt{\frac{\hat{\Lambda}(t)}{k}} < \Lambda(t) < \hat{\Lambda}(t) + z_{\alpha/2} \sqrt{\frac{\hat{\Lambda}(t)}{k}} \quad (2.6)$$

para $0 < t \leq T$, donde $z_{\alpha/2}$ es el percentil $1 - \alpha/2$ de la distribución normal estándar. El resultado es asintóticamente exacto en los extremos a_i de los intervalos de la partición de $(0, T]$.

2.3.3. Test de homogeneidad para procesos de Poisson

Las pruebas de hipótesis son herramientas estadísticas utilizadas para tomar decisiones sobre las características de una población basadas en muestras de datos. Se basan en la formulación de dos hipótesis: la hipótesis nula (H_0) y la hipótesis alternativa (H_a). La hipótesis nula representa la afirmación predeterminada y es la que se prueba. La hipótesis alternativa es una afirmación complementaria a la hipótesis nula y se acepta si se rechaza la hipótesis nula.

Por lo general, una prueba de hipótesis implica definir un estadístico de prueba, que es una función de los datos muestrales. Se compara este estadístico con un valor crítico o se calcula un valor p que representa la probabilidad de obtener resultados al menos tan extremos como los observados, suponiendo que la hipótesis nula es verdadera.

El valor p se calcula utilizando métodos estadísticos específicos dependiendo del diseño de la prueba y la distribución de los datos. Se interpreta como la probabilidad de observar los resultados muestrales si la hipótesis nula fuera cierta. Si el valor p es menor que un nivel de significancia predefinido (generalmente denotado como α), se rechaza la hipótesis nula en favor de la hipótesis alternativa.

Un ejemplo común de una prueba de hipótesis es el test t de Student, utilizado para determinar si hay una diferencia significativa entre las medias de dos muestras independientes.

Test t de Student para diferencias de medias

El test t de Student se utiliza para determinar si hay una diferencia significativa entre las medias de dos muestras independientes. Supongamos que tenemos dos muestras independientes, $X_1, X_2, \dots, X_{n_1}$ y $Y_1, Y_2, \dots, Y_{n_2}$, con medias μ_X y μ_Y respectivamente. La hipótesis nula y alternativa serían:

$$\begin{aligned} H_0 : \mu_X &= \mu_Y \\ H_a : \mu_X &\neq \mu_Y \end{aligned} \tag{2.7}$$

El estadístico del test t se calcula como:

$$t = \frac{\mu_X - \mu_Y}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \quad (2.8)$$

Donde:

- μ_X y μ_Y son las medias de las dos muestras.
- s_X^2 y s_Y^2 son las varianzas muestrales de las dos muestras.
- n_1 y n_2 son los tamaños de las dos muestras.

Para calcular el valor p , primero calculamos el grado de libertad df utilizando la fórmula:

$$df = n_1 + n_2 - 2 \quad (2.9)$$

Luego, utilizamos el valor absoluto de t y df para obtener el valor p a partir de la distribución t de Student. Si $p < \alpha$, donde α es el nivel de significancia predefinido, rechazamos la hipótesis nula y concluimos que hay una diferencia significativa entre las medias de los dos grupos.

Test de homogeneidad

Seguindo a Brown y Zhao en [3], se presenta la base de los test de hipótesis de homogeneidad y tres estadísticos para el proceso de Poisson, comúnmente utilizados para verificar la homogeneidad.

Para comprender estas pruebas, es fundamental recordar que en un proceso de Poisson los incrementos generan muestras independientes de variables aleatorias de Poisson. Por lo tanto, estas pruebas están diseñadas para evaluar la hipótesis nula de homogeneidad, la cual implica que los procesos subyacentes son distribuciones de Poisson con parámetros λ_i iguales:

$$H_0 : X_i \sim \text{Poiss}(\lambda_i), \quad \lambda_1 = \dots = \lambda_n. \quad (2.10)$$

En un contexto como el nuestro, la hipótesis alternativa no siempre se delinea con precisión. En general, uno suele desear enfocarse en alternativas que usen la sobredispersión en sentido de que la variabilidad observada entre las muestras es mayor de lo que se esperaría bajo la hipótesis nula de homogeneidad.

Para medir esta sobredispersión, comparamos la varianza muestral S^2 con la media muestral $\bar{X}$.

$$\frac{E(S^2)}{E(\bar{X})} > 1 \quad (2.11)$$

donde

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Bajo la hipótesis nula, se supone que todas las muestras provienen de una distribución de Poisson con el mismo parámetro λ . En este caso, la varianza de cada muestra S^2 y la media de cada muestra $\bar{X}$ deberían ser aproximadamente iguales.

Si la relación es mayor que uno, indica que la dispersión entre las muestras es mayor que la esperada bajo la hipótesis nula, lo que sugiere una posible sobredispersión. Por esta razón, considerará la siguiente hipótesis alternativa.

$$H_a : X_i \sim \text{Poisson}(\lambda_i), \sum (\lambda_i - \bar{\lambda})^2 > 0. \quad (2.12)$$

Teniendo esto en cuenta, a continuación se describen brevemente tres estadísticos comúnmente utilizadas para evaluar la hipótesis nula H_0 .

Estadístico razón de verosimilitudes

El estadístico del cociente de verosimilitudes para probar H_0 versus H_a es

$$T_{LR} = 2 \sum_{i=1}^n X_i \ln \left(\frac{X_i}{\bar{X}} \right). \quad (2.13)$$

Bajo la hipótesis nula, este estadístico se distribuye asintóticamente como una variable Chi-cuadrado con $n - 1$ grados de libertad (asintóticamente conforme $n \rightarrow \infty$ para λ fijo). Por lo tanto, este test rechaza H_0 cuando $T_{LR} > \chi_{n-1;1-\alpha}^2$.

Estadístico Chi-cuadrado Condicional

Bajo la hipótesis H_0 se tiene $\lambda_i = \bar{X}$ y bajo un estado general $\lambda_i = X_i$, así que, la diferencia relativa entre estos estadísticos puede ser un criterio para probar la hipótesis H_0 . El estadístico es el siguiente:

$$T_{cc} = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{\bar{X}} = \frac{(n-1)}{\bar{X}} S^2 \quad (2.14)$$

Bajo la hipótesis nula H_0 , este estadístico sigue una distribución asintótica χ_{n-1}^2 . Se rechaza la hipótesis H_0 si $T_{cc} > \chi_{n-1;1-\alpha}^2$. Esta prueba también es conocida como “test de dispersión de Poisson.” “test de varianza”.

Estadístico de Neyman-Scott

Esta estadístico está directamente motivada por la expresión 2.11. El estadístico de prueba es:

$$T_{NS} = \sqrt{\frac{n-1}{2}} \left(\frac{S^2}{\bar{X}} - 1 \right). \quad (2.15)$$

Este estadístico está normalizada de modo que asintóticamente $T_{NS} \sim N\left(\frac{\psi^2}{\sqrt{2n}}, 1\right)$. La afirmación asintótica aquí es válida cuando $n \rightarrow \infty$, con $\lambda_1, \dots, \lambda_n$ elegidos de manera que $\psi^2 = O(\sqrt{n})$ y $\liminf \lambda > 0$.

Para realizar una prueba de hipótesis utilizando este estadístico, se rechaza la hipótesis nula si $T_{NS} > \Phi^{-1}(1 - \alpha)$, donde α representa el nivel de significancia previamente establecido.

Capítulo 3

Clasificación automática de imágenes coronográficas

En este capítulo, se lleva a cabo una automatización en la clasificación de imágenes coronográficas tomadas en sectores localizados en dos amplias áreas dentro de la región de Valparaíso. Para preservar la confidencialidad de los datos, nos referiremos a estas áreas como Zona A y Zona B. Los sectores dentro de cada una de estas zonas se nombrarán como Sector 1A hasta Sector 9A para Zona A, y Sector 1B hasta Sector 5B para Zona B.

Para la automatización se trabaja la aplicación de herramientas OCR sobre las imágenes coronográficas, esto se implementa en el lenguaje de programación Python. Se evalúa el rendimiento de diferentes paquetes de OCR: Pytesseract, KerasOCR y EasyOCR, inspirado en artículos que enfrentaban problemas similares. Uno de los problemas abordados involucra la extracción de caracteres en patentes [5], mientras que el otro se centra en la identificación de números en billetes [1].

Las pruebas iniciales de la aplicación de los OCR revelaron desafíos relacionados con la variabilidad de la precisión en la identificación de números. Luego de las pruebas iniciales, se detalla el proceso de procesamiento de imágenes coronográficas, incluyendo estrategias para resaltar y aislar la región de interés en la imagen original. Posteriormente, se evalúa la aplicación de OCR en estas imágenes procesadas, comparando el desempeño de Pytesseract, KerasOCR y EasyOCR.

Finalmente, se implementa la automatización en la clasificación de imágenes coro-

nográficas utilizando el OCR con mejores resultados. Este algoritmo se integra en una interfaz de usuario desarrollada con Streamlit en Python. Esta interfaz proporciona una forma sencilla y eficiente de clasificar las imágenes de las inspecciones de coronografía, agilizando así el proceso de análisis.

3.1. Pruebas Iniciales

Durante la fase inicial de desarrollo, se llevaron a cabo pruebas utilizando imágenes coronográficas completas (consulte la Figura 3.1) para evaluar el rendimiento de los paquetes Pytesseract, KerasOCR y EasyOCR. Estas implementaciones se realizan en Python y con las configuraciones predeterminadas de las herramientas. Como era de esperar, los resultados fueron insatisfactorios, mostrando una baja precisión e inconsistencia en el reconocimiento de caracteres debido a todos los elementos presentes en las imágenes.



Figura 3.1: Imagen coronográfica de ejemplo.

Debido a los malos resultados obtenidos en las primeras pruebas, se piensa en recortar la región de interés donde se encuentra la información relevante. Posterior al recorte se aplican los tres diferentes OCR considerados. Para ilustrar esto se toma como ejemplo la Figura 3.1. Obteniendo la imagen recortada y los resultados de los OCR en la Figura 3.2.

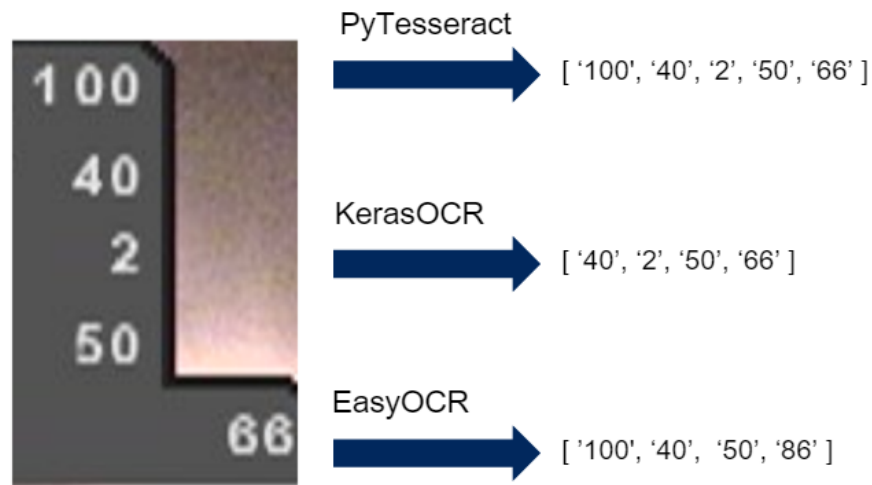


Figura 3.2: Recorte en la zona de interés y aplicación de los OCR.

Los resultados de los OCR aplicados a la imagen recortada muestran que PyTesseract logra extraer correctamente los dígitos. Sin embargo, se observa que KerasOCR y EasyOCR presentan errores al identificar los números y además no logran capturarlos todos.

Después de realizar pruebas adicionales, se pudo determinar que el problema estaba relacionado con el ruido generado al recortar la imagen de manera no óptima. Al aplicar una binarización a la imagen recortada, se observa que resalta tanto el contorno exterior de la zona de interés como los números, esto se puede observar en la Figura 3.3.

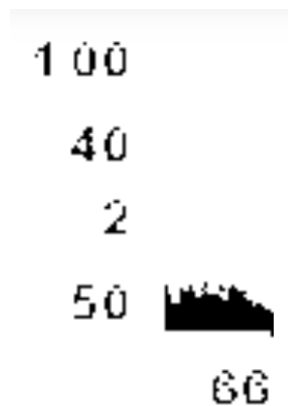


Figura 3.3: Umbralización a la zona de interés recortada.

Por lo tanto, se propone llevar a cabo un preprocesamiento de la imagen antes de aplicar el OCR, con el objetivo de eliminar el ruido y mejorar la calidad de los resultados. Las técnicas utilizadas se detallan a continuación.

3.2. Preprocesamiento de Imágenes Coronográficas

En esta sección, se presenta el código desarrollado para el preprocesamiento de imágenes coronográficas utilizando las bibliotecas OpenCV y NumPy de Python. El objetivo es aislar la región de interés lo máximo posible, eliminando así el ruido de las imágenes.

El código realizado se presenta a continuación.


```

img = cv2.imread('ruta')
polygon_vertices = np.array([
    [84, 390],    # Vértice 1
    [124, 390],   # Vértice 2
    [124, 500],   # Vértice 3
    [170, 500],   # Vértice 4
    [170, 525],   # Vértice 5
    [84, 525],    # Vértice 6
], np.int32)
mask = np.zeros_like(img)
cv2.fillPoly(mask, [polygon_vertices], (255, 255, 255))
result = cv2.bitwise_and(img, mask)
start_row = 390
start_col = 84
end_row = 528
end_col = 170
cropped_img = result[start_row:end_row, start_col:end_col, :]
cropped_img_rgb = cv2.cvtColor(cropped_img, cv2.COLOR_BGR2RGB)
gray_img = cv2.cvtColor(cropped_img_rgb, cv2.COLOR_BGR2GRAY)
_, thresholded = cv2.threshold(gray_img, 200, 255, cv2.THRESH_BINARY_INV)

```

Figura 3.4: Código de preprocesamiento de imágenes desarrollado en Python.

La explicación del código se puede resumir en los siguientes pasos:

1. **Cargar la imagen:** En este paso, se carga la imagen en formato de matriz utilizando la función “cv2.imread()”. Esto prepara la imagen para su posterior procesamiento.
2. **Definir la región de interés:** Se especifican los vértices de la forma geométrica donde se encuentra la región de interés. Esta tiene forma de “L” y se utilizará para seleccionar una región específica de la imagen. Estos vértices se definen como una matriz NumPy de puntos.
3. **Crear una máscara:** Se crea una máscara de la misma forma y tamaño que la imagen original pero de color negro. Luego, se utiliza la función “cv2.fillPoly()” para colorear la máscara negra con color blanco en las coordenadas de la figura en forma de “L” definida anteriormente con valores blancos.
4. **Aplicar la máscara a la imagen original:** La máscara se aplica a la imagen original utilizando la función “cv2.bitwise_and()”. Esto permite conservar solo la región de interés de la imagen original definida por la figura en forma de “L”.

5. **Recortar la región de interés:** Se recorta la región de interés de la imagen resultante utilizando las coordenadas especificadas. Esto permite centrarse en la parte de la imagen que contiene la información relevante.
6. **Convertir la imagen recortada a RGB:** La imagen recortada se convierte de formato BGR a formato RGB utilizando “cv2.cvtColor()”. Esto asegura que la imagen esté en el formato correcto para su visualización y procesamiento.
7. **Convertir la imagen recortada a escala de grises:** Se convierte la imagen recortada en formato RGB a escala de grises utilizando “cv2.cvtColor()”. Esto simplifica el procesamiento posterior y permite realizar operaciones de umbralización.
8. **Aplicar umbralización a la imagen en escala de grises:** La umbralización se aplica a la imagen en escala de grises utilizando “cv2.threshold()”. Esto permite establecer un umbral para distinguir solamente los píxeles blancos que corresponden a los números.

Al aplicar este procesamiento a las imágenes coronográficas se logra eliminar el ruido y conservar únicamente la región de interés. Para ilustrar el funcionamiento se considera la imagen de la Figura 3.1, obteniendo como resultado la Figura 3.5

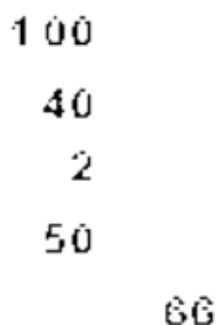


Figura 3.5: Imagen coronográfica procesada.

3.3. Implementación de los OCR

El proceso de OCR se aplica a las imágenes coronográficas preprocesadas (Figura 3.5) mediante el uso de las bibliotecas PyTesseract, KerasOCR y EasyOCR. Este paso

tiene como objetivo principal presentar y configurar estas herramientas.

3.3.1. Pytesseract

PyTesseract es una herramienta de OCR para Python que utiliza el motor Tesseract-OCR de Google. Esta herramienta es conocida por su robustez en la detección de texto en diversas condiciones, convirtiéndola en una opción confiable para nuestra aplicación en imágenes de coronografía. [6]

Para la implementación de Pytesseract se tiene el siguiente código.

```
# Inicializar PyTesseract
pytesseract.pytesseract.tesseract_cmd=r"C:\\Program Files\\Tesseract-OCR\\tesseract.exe"

# *Código de preprocesamiento de la imagen*

# Configuración de PyTesseract
custom_config = r'--oem 3 --psm 6 outputbase digits -c tesseract_char_whitelist=0123456789'
# Reconocer texto (números) en la imagen preprocesada
text = pytesseract.image_to_string(Image.fromarray(thresholded), config=custom_config)
# Imprimir los números extraídos
print(text)
```

Figura 3.6: Configuración de OCR Pytesseract.

En general, todos los OCR siguen la misma lógica: inicialización, configuración, reconocimiento de texto y finalmente la impresión del resultado. Sin embargo, lo que diferencia a Pytesseract es la gran cantidad de opciones de configuración que ofrece. En la figura 3.6, en el parámetro “custom_config” se especifica al OCR que los números están en bloque y que solo se deben reconocer dígitos del 0 al 9.

3.3.2. KerasOCR

KerasOCR es una biblioteca basada en redes neuronales, la configuración se ajusta automáticamente para adaptarse a las particularidades de las imágenes [10]. En comparación con Pytesseract, KerasOCR no se puede configurar de tal manera que se enfoque solo en números del 0 al 9 por lo que aumenta la probabilidad de error al reconocer dígitos.

Para la implementación de KerasOCR se tiene el siguiente código.

```
# Inicializar el modelo OCR de Keras
pipeline = keras_ocr.pipeline.Pipeline()

# *Código de preprocesamiento de la imagen*

# Reconocer texto (números) en la imagen preprocesada
prediction_groups = pipeline.recognize(images)
recognized_text = [text for text, bbox in prediction_groups[0]]
# Imprimir los números extraídos
print(recognized_text)
```

Figura 3.7: Configuración de KerasOCR.

En el caso de esta biblioteca, se tuvieron complicaciones, ya que todos los números “0” que estaban aislados fueron reconocidos como letras “o” utilizando la implementación anterior. Debido a este problema, se decidió implementar KerasOCR con una configuración que permite cambiar las “o” reconocidas por “0”. A esta nueva implementación la llamaremos KerasOCR modificado.

Para la implementación de KerasOCR modificado se tiene el siguiente código.

```
# Inicializar el modelo OCR de Keras
pipeline = keras_ocr.pipeline.Pipeline()

# *Código de preprocesamiento de la imagen*

# Reconocer texto (números) en la imagen preprocesada
prediction_groups = pipeline.recognize(images)
# Reemplazar las letras 'o' detectadas por '0'
recognized_text = [text.replace('o', '0') for text, bbox in prediction_groups[0]]
# Imprimir los números extraídos
print(recognized_text)
```

Figura 3.8: Configuración de KerasOCR modificado.

3.3.3. EasyOCR

EasyOCR es una biblioteca de OCR que se destaca por su facilidad de uso y su capacidad para manejar una variedad de idiomas y tipos de texto. Utiliza modelos de aprendizaje profundo entrenados para realizar el reconocimiento de texto en imágenes de manera eficiente. [2]

EasyOCR no ofrece configuraciones como las bibliotecas Pytesseract o KerasOCR. Por lo que simplemente se aplicó con las configuraciones predefinidas.

Para la implementación de EasyOCR se tiene el siguiente código.

```
# *Código de preprocesamiento de la imagen*

# Configuración de EasyOCR
reader = easyocr.Reader(['en'], gpu=False)
# Reconocer texto (números) en la imagen preprocesada
result = reader.readtext(gray_img, detail=0)
# Imprimir los números extraídos
print("Números extraídos:", result)
```

Figura 3.9: Configuración de EasyOCR.

3.4. Automatización del OCR

Antes de comenzar se debe tener en consideración la estructura de las carpetas que se tratan. Se tienen dos carpetas llamadas Zona A y Zona B, dentro de cada carpeta hay subcarpetas con el nombre del sector donde se realizaron las inspecciones. Finalmente, en las subcarpetas se encuentran las imágenes correspondientes a las coronografías de cada sector. La estructura de las carpetas se puede apreciar en la Figura 3.10.

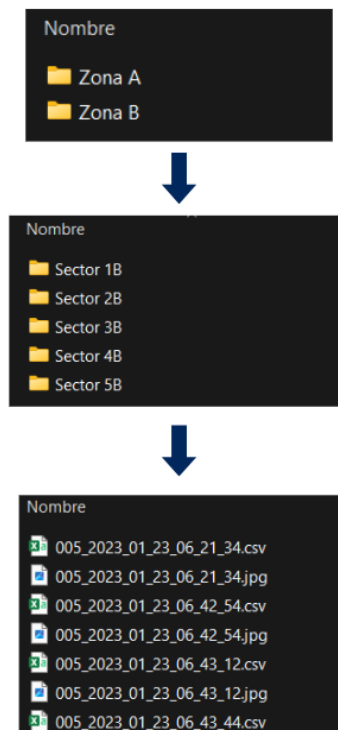


Figura 3.10: Estructuras de las carpetas y subcarpetas donde están ubicadas las imágenes coronográficas.

Para automatizar el preprocesamiento y la aplicación de OCR a las imágenes coronográficas se utiliza la biblioteca `Os` de Python. Esta herramienta facilita manejo y la navegación a través de las carpetas. Al recorrer cada carpeta y subcarpeta, se podrá acceder a todas las imágenes coronográficas.

El siguiente algoritmo permite recorrer cada carpeta y subcarpeta en una ruta específica.

```

# Ruta de la carpeta que contiene las carpetas con imágenes
folder_path = "Ruta de la carpeta a utilizar"

# Recorre todas las carpetas y archivos en la carpeta principal y sus subcarpetas
for root, dirs, files in os.walk(folder_path):
    for filename in files:
        # Verifica si el archivo es una imagen (puedes ajustar las extensiones según tus necesidades)
        if filename.lower().endswith(('.png', '.jpg', '.jpeg')):
            # Obtiene el nombre de la carpeta actual
            current_folder = os.path.basename(root)
            # Construye la ruta completa de la imagen
            image_path = os.path.join(root, filename)

# *Código de preprocesamiento y aplicación de OCR*

```

Figura 3.11: Algoritmo que permite identificar cada imagen coronográfica ubicada en subcarpetas o carpetas.

Con este algoritmo, es posible aplicar el preprocesamiento y el OCR a múltiples imágenes coronográficas ubicadas en carpetas o subcarpetas. Esta capacidad nos permite analizar un gran número de imágenes en poco tiempo. Por lo tanto, utilizando esta automatización, podemos comparar las tres librerías de OCR y determinar cuál se adapta mejor a este problema específico.

3.5. Comparación de librerías de OCR

Se cuenta con un total de 1167 imágenes coronográficas entre las Zonas A y Zona B. Para decantarnos por un OCR se testearán las herramientas Pytesseract, KerasOCR y EasyOCR en 140 imágenes tomadas de manera aleatoria, esta cantidad corresponde aproximadamente al 12 % del total.

Para evaluar el desempeño se considera solamente el número de errores o aciertos al extraer la variable de descargas parciales presentes en las imágenes coronográficas. La métrica a utilizar es la exactitud (accuracy) que nos indica el porcentaje de aciertos. Esta metodología de desempeño se puede observar en el artículo desarrollado por Idrose et al. [5].

Los resultados de las pruebas utilizando los cuatro algoritmos de OCR se presentan en Tabla 3.1. Se debe recordar que KerasOCR presenta dos posibles configuraciones.

Librería	Aciertos	Errores	% Aciertos
Pytesseract	140	0	100 %
KerasOCR modificado	115	25	81,14 %
EasyOCR	51	89	36,43 %
KerasOCR	40	100	28,57 %

Cuadro 3.1: Rendimiento de las librerías de OCR.

El cuadro anterior se puede ver de manera gráfica en la siguiente Figura.

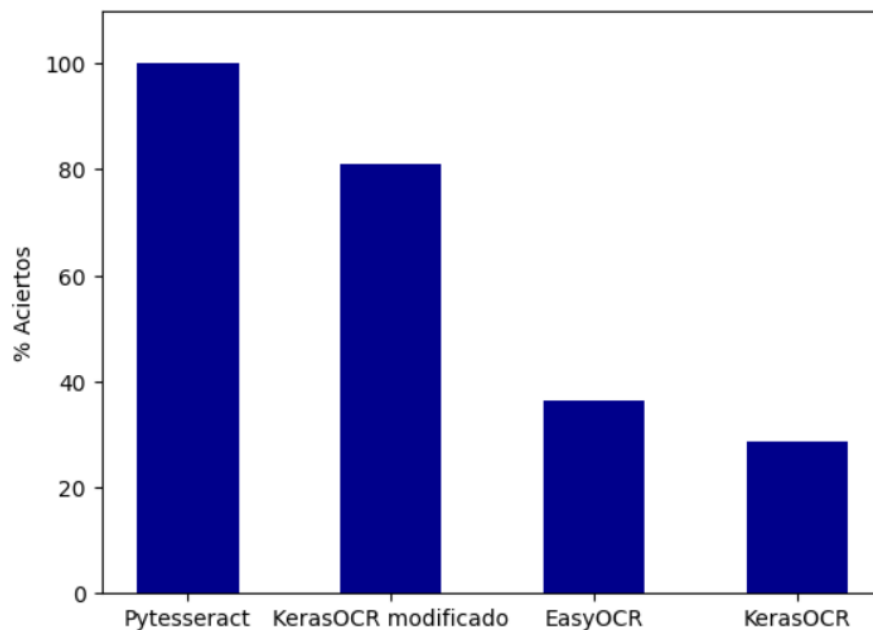


Figura 3.12: Gráfico de barras del rendimiento de las librerías de OCR.

3.5.1. Análisis de resultados

- PyTesseract ha demostrado un rendimiento excelente con un 100 % de aciertos en la muestra de 140 imágenes. Su robustez y versatilidad en la detección de texto son evidentes, lo que lo convierte en una buena elección a utilizar para la automatiza-

ción.

- La versión modificada de Keras OCR muestra un desempeño respetable con un 81,14% de aciertos. Aunque no alcanza la precisión de PyTesseract, haber modificado el OCR funciona mejor que el Keras OCR. Sin embargo, la modificación realizada deja que desear al tener que cambiar letras “o” reconocidas por “0” para obtener un mejor resultado.
- EasyOCR, aunque fue fácil de implementar y proporcionó resultados satisfactorios en algunas imágenes, mostró un desempeño significativamente inferior en comparación con PyTesseract y la versión modificada de Keras OCR. Con solo un 36,43 % de aciertos.
- La versión estándar de Keras OCR presenta un rendimiento muy bajo con un 28,57 % de aciertos. Esto sugiere que, sin ajustes específicos, Keras OCR puede tener dificultades para lidiar con las características particulares de las imágenes utilizadas en este trabajo.

Basándose en los resultados, PyTesseract es la mejor opción en términos de aciertos en la extracción de texto en imágenes coronográficas. Por lo tanto, para la clasificación automática se utilizará este OCR.

3.6. Sistema automatizado e interfaz de usuario

Para la clasificación automatizada de imágenes coronográficas se utiliza el algoritmo de extracción de dígitos mencionado previamente, el cual facilita la extracción de todas las variables presentes en las imágenes. Una vez que se han extraído los datos numéricos, se pueden organizar en un data frame utilizando la librería Pandas de Python.

En este data frame, cada columna representa una variable extraída de las imágenes, como el nombre de la imagen, ganancia, descargas, entre otras. Esto permite tener una columna específica para el recuento de descargas parciales en cada imagen, lo cual es fundamental para clasificar el nivel de daño según el criterio ilustrado en la Figura 2.4. Al aplicar este criterio a la columna de descargas parciales, podemos generar una nueva columna de clasificación que asigna a cada imagen una clase (A, B, C o D) según corresponda.

La clasificación automática se integra en una interfaz de usuario creada en Streamlit. Esta es una biblioteca Python de código abierto que facilita la creación de aplicaciones

web personalizadas de manera sencilla. A continuación, se detallan las diversas funciones implementadas en este desarrollo web para mostrar los resultados de la clasificación sobre las imágenes disponibles.

Al ejecutar el desarrollo web, se abre automáticamente en el navegador predeterminado, mostrando un título “Clasificación de imágenes coronográficas” y un área donde se puede cargar una carpeta comprimida, como se ilustra en la Figura 3.13. Es importante destacar que las carpetas deben estar en formato “.ZIP”, de acuerdo con las necesidades de la empresa, ya que es el formato en el que manejan las imágenes.



Figura 3.13: Inicio de interfaz creada con Streamlit.

La aplicación web tiene la capacidad de cargar una carpeta comprimida que contenga imágenes coronográficas directamente, o incluso una estructura de carpeta comprimida con subcarpetas que contengan estas imágenes.

Para seguir con las funciones del desarrollo web, se carga la carpeta que contiene las imágenes de coronografías correspondientes a al Sector 1B de la Zona B. Al cargar la carpeta comprimida, se muestra la data obtenida del proceso automatizado de clasificación. Esta data incluye el nombre de la imagen analizada, las cinco variables presentes en la región de interés y la clasificación correspondiente basada en la cantidad de descargas parciales, tal como se observa en la figura 3.14.

Data obtenida por coronografía

↑	Imagen	Ganancia	Umbral	Integración	Integración ganancia	Descargas	A
0	005_2023_01_23_06_21_34.jpg	85	40	2	50	0	c
1	005_2023_01_23_06_42_54.jpg	85	40	2	50	783	c
2	005_2023_01_23_06_43_12.jpg	85	40	2	50	683	c
3	005_2023_01_23_06_43_44.jpg	85	40	2	50	583	c
4	005_2023_01_23_06_44_08.jpg	85	40	2	50	2,450	c
5	005_2023_01_23_06_44_34.jpg	85	40	2	50	166	c
6	005_2023_01_23_06_44_53.jpg	85	40	2	50	183	c
7	005_2023_01_23_06_45_38.jpg	85	40	2	50	0	c
8	005_2023_01_23_06_45_53.jpg	85	40	2	50	133	c
9	005_2023_01_23_06_46_28.jpg	85	40	2	50	200	c

Figura 3.14: Data obtenida de las imágenes coronográficas.

La web local tiene tres opciones relevantes en la data, la primera permite descargar la data en formato “.csv”, la segunda incluye un buscador de caracteres y una aplicación de la data para verla de manera completa. Estas opciones se logran apreciar en la Figura 3.15.

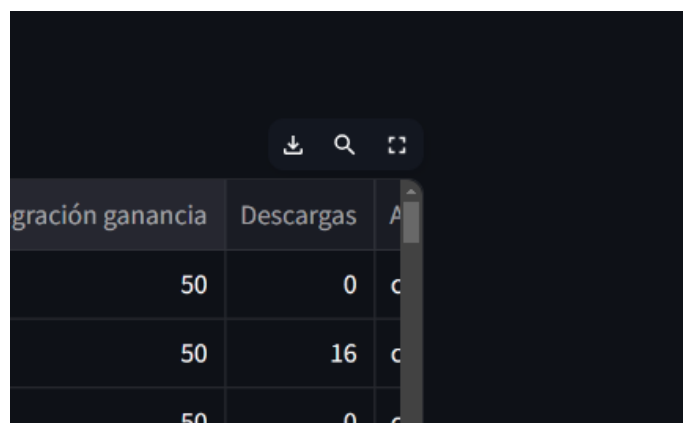


Figura 3.15: Opciones para la data obtenida a través del desarrollo web.

Posteriormente, la interfaz entrega un gráfico de barras de las clasificaciones de las imágenes analizadas. Además, si se identifican imágenes con clasificación D se presenta la observación junto con la imagen correspondiente para que el técnico las verifique, dado

que estas son las más urgentes a tratar. Lo anteriormente mencionado se puede observar en la figura 3.16.

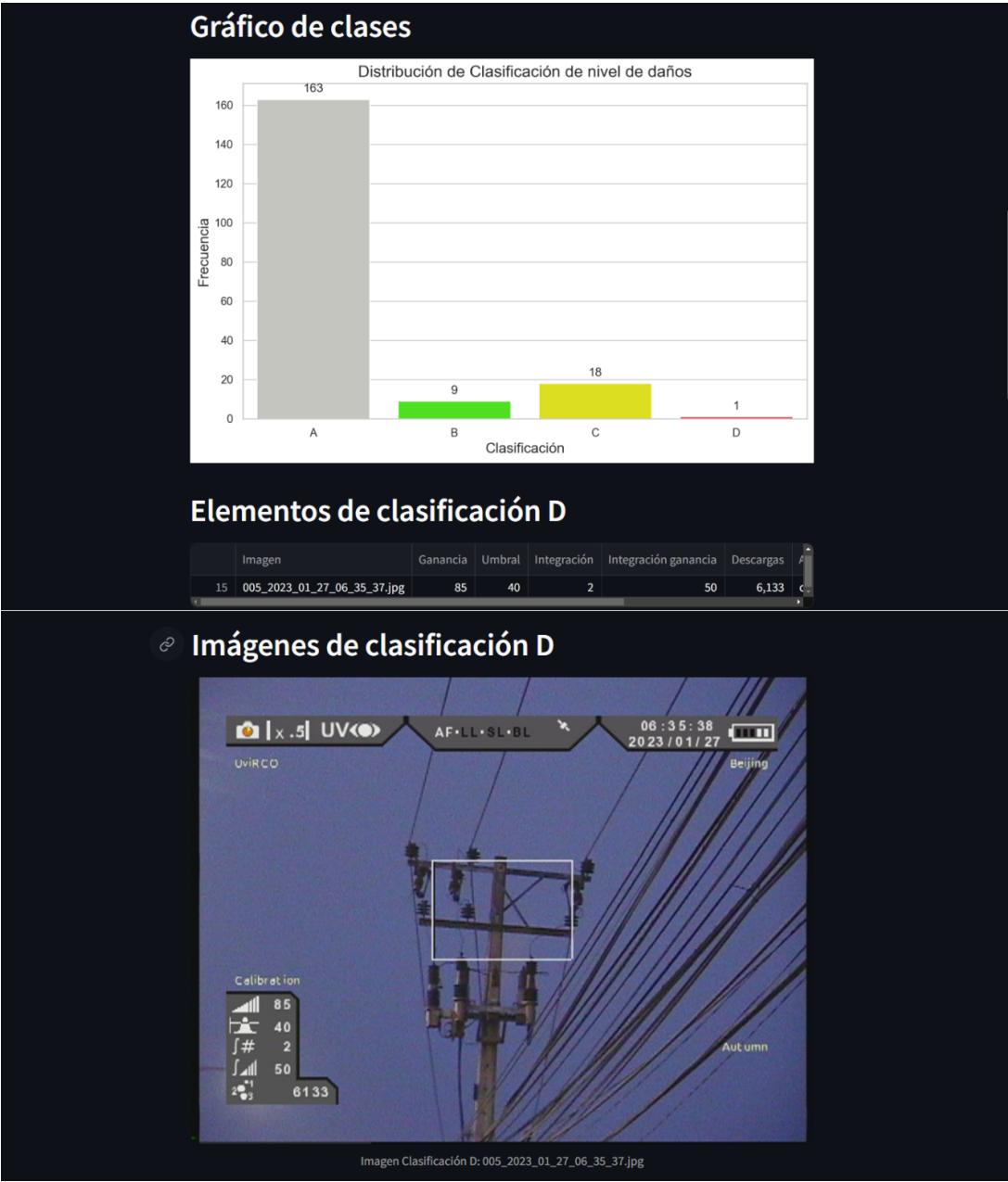


Figura 3.16: Gráfico de barras sobre la frecuencia de clases del nivel de daño e imágenes detectadas como clasificación D.

3.7. Resultados de la clasificación de imágenes coronográficas

El sistema automatizado cuenta con la capacidad de clasificar de manera individual cada imagen coronográfica que se encuentra en las subcarpetas asociadas a sectores específicos, ya sea en la Zona A o en la Zona B.

En cuanto a la clasificación de imágenes coronográficas de los sectores en Zona A, los resultados pueden ser resumidos en la siguiente tabla:

Sectores de Zona A	A	B	C	D
Sector 1A	110	0	1	0
Sector 2A	30	0	0	0
Sector 3A	30	0	0	0
Sector 4A	29	0	1	0
Sector 5A	30	0	0	0
Sector 6A	30	0	0	0
Sector 7A	30	0	0	0
Sector 8A	64	0	0	0
Sector 9A	23	5	2	0
Total	376	5	4	0

Cuadro 3.2: Resultados de clasificación de daños en imágenes coronográficas de sectores en la Zona A.

Se destaca que la mayoría de las imágenes clasificadas corresponden al tipo A, lo que indica un estado general satisfactorio de los equipos eléctricos inspeccionados. Sin embargo, se identificaron algunas imágenes clasificadas como tipo B y C, indicando la presencia de problemas menores.

Para la clasificación de imágenes coronográficas de los sectores en Zona B son los siguientes:

Sectores de Zona B	A	B	C	D
Sector 1B	49	11	11	0
Sector 2B	163	9	18	1
Sector 3B	93	4	1	1
Sector 4B	62	4	4	0
Sector 5B	290	23	37	1
Total	657	51	71	3

Cuadro 3.3: Resultados de clasificación de daños en imágenes coronográficas de sectores en la Zona B.

Para la Zona B, la mayoría de las imágenes también se clasificaron como tipo A. Sin embargo, se identificaron más casos de imágenes tipo B y C en comparación con la Zona A. Además, la presencia de imágenes clasificadas como tipo D destaca la existencia de situaciones críticas que requieren atención inmediata.

Es importante considerar que mientras mayor sea la cantidad de imágenes coronográficas permitiría obtener una perspectiva más completa y precisa del estado de la infraestructura eléctrica en estas zonas. Esto facilitaría la identificación y el abordaje más efectivo de cualquier problema potencial.

Capítulo 4

Análisis de Interrupciones eléctricas

En este capítulo, se lleva a cabo un análisis estadístico sobre las interrupciones eléctricas en sectores localizados en dos amplias áreas dentro de la región de Valparaíso. Al igual que en el capítulo anterior, para preservar la confidencialidad de los datos, nos referiremos a estas áreas como Zona A y Zona B. Los sectores dentro de cada una de estas zonas se nombrarán como Sector 1A hasta Sector 9A para Zona A, y Sector 1B hasta Sector 5B para Zona B.

La data utilizada es un historial de interrupciones eléctricas de un periodo de tiempo que comienza el 1 de enero del 2021 y termina el 18 de septiembre del 2023 (990 días). Cada registro incluye información como el momento exacto cuando ocurre la interrupción, la causa identificada y el nombre del alimentador asociado donde ocurrió la falla. En este trabajo llamaremos sector a los diferentes alimentadores.

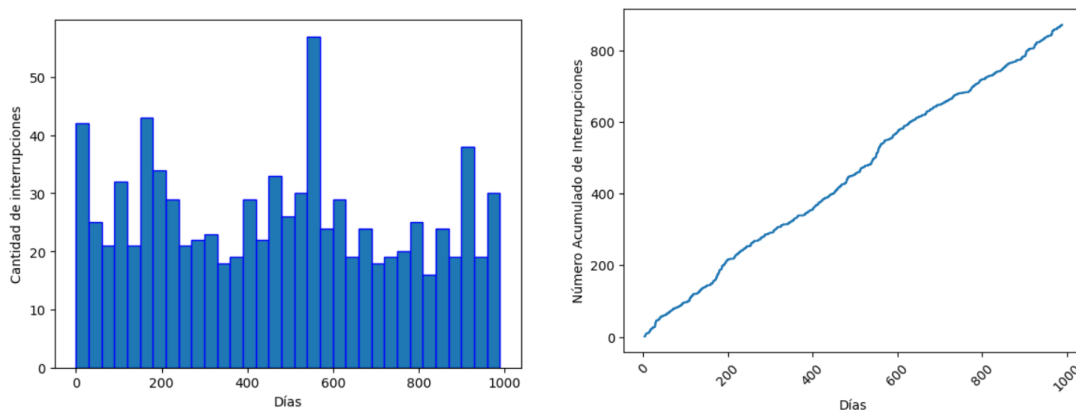
En la data se encuentran una gran variedad de causas que pueden provocar las interrupciones eléctricas, ya sean detectables a través de inspecciones coronográficas o no. Dado que el interés principal es analizar el impacto de estas inspecciones en las interrupciones eléctricas, se realiza un proceso de filtrado, limitando las causas únicamente a aquellas identificables mediante imágenes de coronografía. Esto incluye causas como: daño de aisladores, desequilibrio de cargas, envejecimiento de materiales, etc.

El propósito de este capítulo es comprender y describir las interrupciones eléctricas y examinar el potencial impacto de las coronografías sobre estas interrupciones. Las ins-

pecciones se realizan sector por sector, por lo que deben abordarse de manera individual. Para evaluar el impacto en sectores considerados homogéneos se utilizan estimaciones basadas en PPH, mientras que para los sectores no homogéneos se recurre a estimaciones no paramétricas basadas en PPNH.

4.1. Análisis Preliminares

Para iniciar el análisis, se examinan todas las observaciones de interrupciones eléctricas en dos zonas de Valparaíso, las cuales se denominarán como Zona A y Zona B. En la Figura 4.1 se presenta un gráfico de frecuencia de las interrupciones considerando intervalos de 30 días y también se presenta el proceso conteo de las interrupciones en tiempo continuo. Se opta por agrupar las interrupciones en periodos de 30 días para tener la interpretación de los resultados de manera mensual, además de que estos intervalos contienen datos suficientes para realizar las estimaciones estadísticas.



(a) Frecuencia de interrupciones en Valparaíso (b) Conteo de interrupciones en Valparaíso

Figura 4.1: Frecuencia de interrupciones eléctricas agrupada por meses y su proceso de conteo asociado en Valparaíso.

Se observa claramente un pico pronunciado cerca del día 550, así como en los extremos del histograma. En el proceso de conteo de interrupciones, estos picos se reflejan como cambios en las pendientes del proceso, la inclinación aumenta en ciertos períodos debido a la acumulación de eventos.

Como complemento, a continuación presentamos una tabla que detalla las principales causas identificadas detrás de las interrupciones eléctricas, junto con su frecuencia de ocurrencia correspondiente.

Causa de la Interrupción	Frecuencia
Envejecimiento de materiales	388
Corrosión	252
Desequilibrio de cargas	152
Falla en artefacto eléctrico	49
Falta de mantenimiento	11
Falla de material	11
Falla por contaminación industrial en el área	5
Construcción deficiente	3

Cuadro 4.1: Causas de las interrupciones eléctricas en todos los sectores y su frecuencia asociada.

Se puede observar en la Tabla 4.1 que las causas que presentan mayor número de frecuencias son el envejecimiento de materiales y la corrosión, mientras que hay muy pocas interrupciones asociadas a falta de mantenimiento, falla de material, falla por contaminación industrial en el área o construcción deficiente. Estos datos resaltan la importancia de implementar medidas preventivas y correctivas dirigidas especialmente hacia las causas más recurrentes.

Para comprender el comportamiento de las interrupciones eléctricas a lo largo del tiempo y evaluar el impacto de las coronografías, es necesario llevar a cabo un análisis por sector. Por lo tanto, comenzaremos examinando las interrupciones eléctricas en la Zona A y sus sectores.

4.2. Análisis de la Zona A

Se observan las interrupciones en la Zona A de Valparaíso a través de un gráfico de interrupciones eléctricas agrupadas por meses y el proceso de conteo asociado presen-

tes en la Figura 4.2.

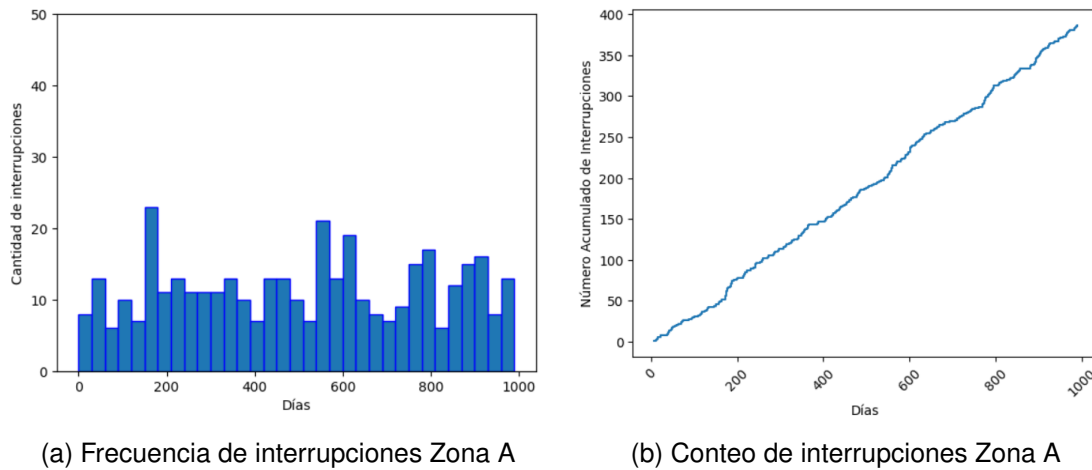


Figura 4.2: Frecuencia de interrupciones eléctricas agrupada por meses y su proceso de conteo asociado en la Zona A.

En cuanto al proceso de conteo se ve bastante homogéneo, con aumentos de intensidades leves en ciertos periodos, pero nada muy significativo. Estos picos de frecuencias ocurren mayormente en estaciones frías, por lo tanto, es interesante estudiar la estacionalidad.

Estacionalidad Zona A

Se realiza un análisis de la estacionalidad dado que la cantidad de interrupciones eléctricas puede verse influenciada por las condiciones climáticas y la temperatura. Con este fin, procederemos a dividir el período de observación en periodos: invierno y verano. En este contexto, el invierno engloba las estaciones frías, como otoño e invierno, mientras que el verano las estaciones más cálidas, como primavera y verano. Esto se puede observar en la Figura 4.3 donde los periodos de verano se visualiza con color naranja y los inviernos de color celeste.

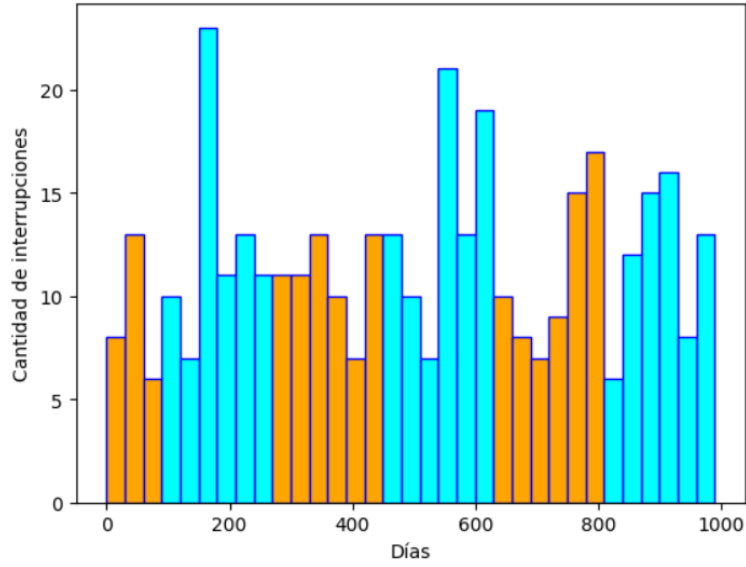


Figura 4.3: Frecuencia de interrupciones eléctricas en la Zona A separadas por periodos de veranos (naranja) e inviernos (celeste).

A primera vista, se observa una mayor cantidad de interrupciones en invierno. Sin embargo, para determinar la relevancia de estas variaciones, primero se estima el valor de la intensidad λ mensual para los veranos e inviernos. Luego, los tiempos entre ocurrencias de ambos periodos se consideran como dos muestras independientes y se les aplica el test t para evaluar si existe variación significativa.

El estimador de máxima verosimilitud utilizado para el cálculo de λ mensual es:

$$\hat{\lambda} = \frac{N}{\sum_{i=1}^N \tau_i} \quad (4.1)$$

donde N es la cantidad de eventos y τ_i son los tiempos entre ocurrencias. Este estimador se desprende de la Preposición 1.

Utilizando este estimador se puede graficar la distribución exponencial estimada y la distribución empírica de los tiempos entre ocurrencias por periodos, como se ve en la Figura 4.4.

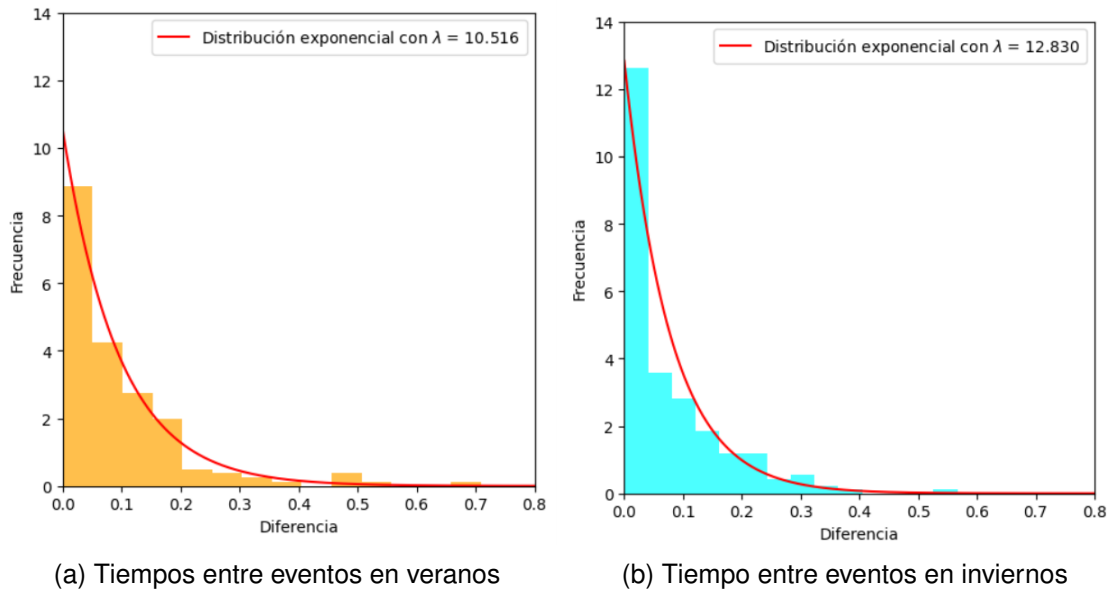


Figura 4.4: Distribución de tiempos entre ocurrencias para los periodos de verano (naranja) e invierno (celeste) en la Zona A junto con la distribución exponencial del lambda estimado.

Se puede ver que para los veranos se tiene un $\hat{\lambda} = 10,516$ y para el invierno se tiene que $\hat{\lambda} = 12,830$, por lo que se puede ver un aumento en cuanto las interrupciones de manera mensual en el invierno. El paso siguiente es ver si estos cambios en las estaciones es significativo para ver si se presenta un tipo de estacionalidad o no. A continuación se recuerda el test t para evaluar si es o no relevante la variación.

Recordemos que las hipótesis del test t utilizando en dos muestras independientes son:

$$\begin{aligned} H_0 : \mu_1 &= \mu_2 \\ H_a : \mu_1 &\neq \mu_2 \end{aligned} \quad (4.2)$$

Aplicando el test a los tiempos entre eventos de veranos e invierno como dos muestras independientes se tienen los siguientes resultados:

Sector	$\hat{\lambda}$ Veranos	$\hat{\lambda}$ Inviernos	Test t
Zona A	10.516	12.830	Acepta H_0

Cuadro 4.2: Lambdas mensuales de veranos e inviernos de la Zona A junto con resultados de test t al 95 % de confianza aplicado a las dos muestras de tiempos entre ocurrencias.

Dado que el test t acepta la hipótesis nula H_0 , no se dispone de evidencia que sugiera la presencia de estacionalidad en la Zona A.

En los casos posteriores donde se investigue la estacionalidad, se seguirá los mismos pasos empleados anteriormente, que consiste en dividir los meses en periodos, estimar la intensidad y evaluar la estacionalidad utilizando el test t con dos muestras: los tiempos entre ocurrencias durante el verano y el invierno.

Homogeneidad Zona A

Ahora se quiere evaluar estadísticamente si la Zona A se comporta de manera homogénea, para esto se utilizan los test de hipótesis enfocados a la homogeneidad presentados en el marco teórico: Chi-cuadrado condicional, Neyman-Scott y Razón de verosimilitud.

Recordemos que para los test de homogeneidad de los PPH se consideran las siguientes hipótesis:

$$\begin{aligned}
 H_0 : X_i &\sim \text{Poisson}(\lambda_i), \quad \lambda_1 = \dots = \lambda_n. \\
 H_a : X_i &\sim \text{Poisson}(\lambda_i), \quad \sum (\lambda_i - \bar{\lambda})^2 > 0.
 \end{aligned}
 \tag{4.3}$$

Aplicando los tres test vistos en el macro teórico obtenemos los siguientes resultados:

Sector	Chi-cuadrado C.	Neyman-Scott	Razón de verosimilitud
Zona A	Acepta H_0	Acepta H_0	Acepta H_0

Cuadro 4.3: Resultados de los test de homogeneidad al 95 % de confianza en la Zona A.

Por lo tanto, se considera que esta zona presenta un comportamiento homogéneo en su totalidad, lo cual implica que es válido asumir $\lambda_1 = \dots = \lambda_n$. En los casos siguientes donde se examine la homogeneidad, se emplearán los tests e hipótesis mencionados en este análisis.

4.2.1. Sectores de la Zona A

Ahora surge el interés de llevar a cabo un análisis más detallado en cada uno de los sectores de la Zona A. Esta decisión se basa en el hecho de que las coronografías están organizadas por sector. Por lo tanto, procedemos a diferenciar los datos por sector, lo que resulta en nueve procesos de conteo de interrupciones eléctricas. Estos procesos se pueden apreciar en la Figura 4.5.

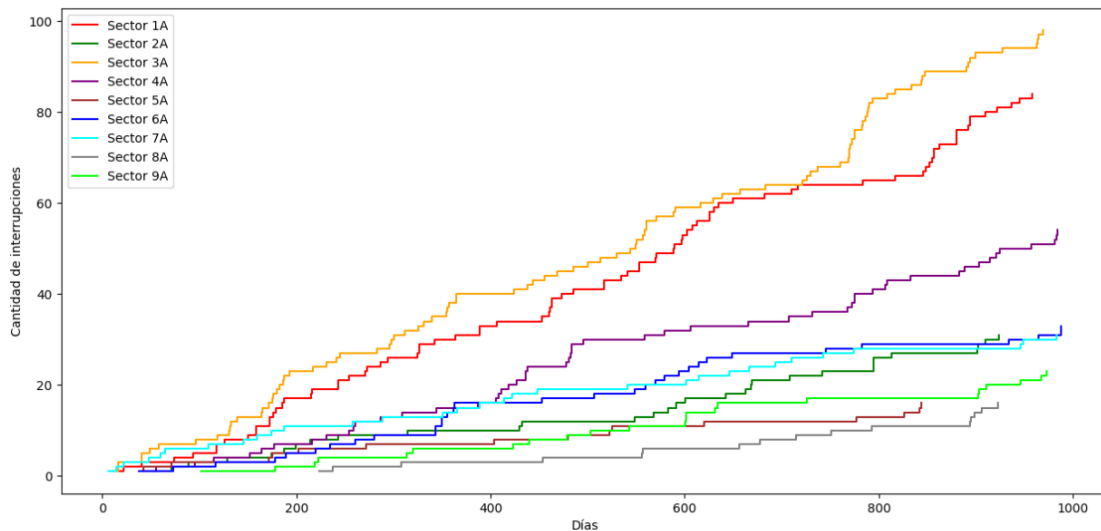


Figura 4.5: Procesos de conteo de interrupciones eléctricas por sectores en la Zona A.

Estacionalidad por sector

Considerar la estacionalidad por sector es fundamental en el análisis, ya que para evaluar el impacto de las inspecciones coronográficas se tiende a pensar en comparar periodos de tiempos antes y después de esta. Sin embargo, si existe estacionalidad en los periodos de verano e invierno, se pueden obtener resultados sesgados. Por ejemplo, si

una inspección coronográfica ocurre en enero, que es común en la mayoría de los casos, y existe estacionalidad, no sería apropiado comparar simplemente tiempos antes y después de la inspección. Esto se debe a que estaríamos pasando de un periodo de verano a uno de invierno (o viceversa) lo que puede influir significativamente en las variaciones de intensidad.

Por lo tanto, se realiza un análisis de estacionalidad para cada sector de la Zona A. Los resultados obtenidos son los presentes en la Tabla 4.4.

Sector	$\hat{\lambda}$ Veranos	$\hat{\lambda}$ Inviernos	Test t
Sector 1A	1.479	3.658	Rechaza H_0
Sector 2A	0.933	1.153	Acepta H_0
Sector 3A	3.187	2.930	Acepta H_0
Sector 4A	1.586	1.641	Acepta H_0
Sector 5A	0.465	0.651	Acepta H_0
Sector 6A	0.727	1.376	Acepta H_0
Sector 7A	0.871	1.025	Acepta H_0
Sector 8A	0.469	0.855	Acepta H_0
Sector 9A	0.482	1.034	Acepta H_0

Cuadro 4.4: Intensidad estimada mensualmente por cada sector en veranos e inviernos por sector junto con resultados del test t al 95 % de confianza aplicado a muestras de tiempos entre ocurrencias.

Los resultados muestran que las intensidades $\hat{\lambda}$ para los periodos de invierno tienden a ser mayores, con la excepción del Sector 3A. Además, solo el Sector 1A presenta una variación significativa en cuanto a los tiempos entre ocurrencias de veranos e inviernos. Esto indica existencia de estacionalidad en este sector.

Para observar la estacionalidad del Sector 1A se muestra a continuación las cantidades de interrupciones eléctricas agrupadas mensualmente, con los meses diferenciados por periodos de invierno y verano.

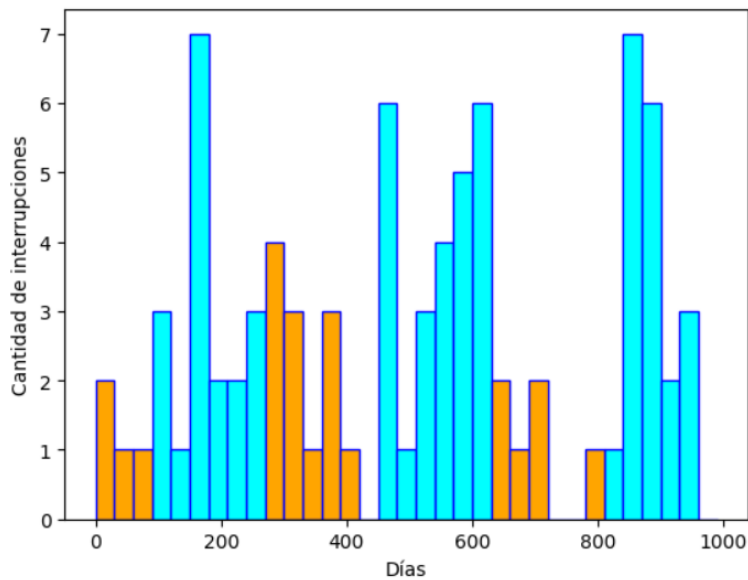


Figura 4.6: Frecuencia de interrupciones eléctricas en el Sector 1A separadas por periodos de veranos (naranja) e inviernos (celeste).

Periodos de estudio antes y después de coronografías

Según los resultados de la Tabla 4.4 solo para el Sector 1A se debe tener en consideración la estacionalidad al tomar los tiempos de estudio antes y después de las coronografía. Sin embargo, igualmente se presentan variaciones en las intensidades estimadas de inviernos y veranos para los demás sectores, por lo tanto, los periodos de estudio se tomarán igual para todos.

Los periodos de estudio se seleccionan de tal manera que sean comparables antes y después de las coronografías. Esto significa que se analizan los mismos meses del año tanto antes como después del procedimiento, asegurando así que las comparaciones sean lo más consistentes posible. Por ejemplo, si se realiza una coronografía en enero de 2022, se estudiarán los datos de febrero a diciembre de 2021 como periodo anterior, y de febrero a diciembre de 2022 como posterior. No se considera el mes en donde ocurre la inspección para considerar la misma cantidad de días en ambos periodos.

En casos donde no sea posible considerar 11 meses antes y después de la coronografía, se deben ajustar los intervalos de considerando la mayor cantidad de tiempo po-

sible. Por ejemplo, para una inspección realizada el 29 de noviembre de 2022, el período posterior será desde el 1 de enero de 2023 hasta el 18 de septiembre de 2023 (fecha límite de observaciones) y el período anterior sería desde el 1 de enero de 2022 hasta el 18 de septiembre de 2022.

Homogeneidad por sector

Ya que se dispone del estudio de estacionalidad por sector, es fundamental evaluar la homogeneidad en cada uno de ellos para aplicar el enfoque adecuado en la evaluación del impacto de la coronografía. Para esto, se aplican los tres tests de homogeneidad, los resultados de homogeneidad de cada sector se presentarán en la Tabla 4.5.

Sector	<i>Chi-cuadrado C.</i>	<i>Neyman-Scott</i>	Razón de verosimilitud
Sector 1A	Rechaza H_0	Rechaza H_0	Rechaza H_0
Sector 2A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 3A	Rechaza H_0	Rechaza H_0	Rechaza H_0
Sector 4A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 5A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 6A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 7A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 8A	Acepta H_0	Acepta H_0	Acepta H_0
Sector 9A	Acepta H_0	Acepta H_0	Acepta H_0

Cuadro 4.5: Resultados de los test de homogeneidad al 95 % de confianza en los sectores de la Zona A.

En los resultados de los tests, se observa que la mayoría de los sectores se pueden considerar homogéneos, excepto el Sector 1A y el Sector 3A. Comparando la información de la Tabla 4.5 con la representación visual de los procesos en la Figura 4.5, se puede notar que los sectores con mayor cantidad de interrupciones son precisamente aquellos clasificados como no homogéneos.

4.2.2. Impacto de las coronografías en sectores homogéneos

Para evaluar el impacto de las inspecciones coronográficas en los sectores con procesos de conteos homogéneos, primero se estima la intensidad antes y después de la inspección. Posteriormente, se evalúa si existen variaciones significativas entre las dos muestras independientes de tiempos entre ocurrencias.

A continuación, se presenta una tabla que detalla las fechas en las que se llevaron a cabo las inspecciones coronográficas en cada sector homogéneo de la Zona A.

Sector homogéneo	Fecha de Inspección Coronográfica
Sector 2A	2022/01/22
Sector 4A	2022/12/01
Sector 5A	2022/01/27
Sector 6A	2022/01/20
Sector 7A	2022/12/02
Sector 8A	2023/04/13
Sector 9A	2022/11/29

Cuadro 4.6: Fechas de inspecciones coronográficas en cada sector homogéneo de la Zona A.

Para comenzar se debe estimar la función de intensidad λ antes y después de la coronografía utilizando el estimador 4.1. A continuación se presenta la distribución de tiempos entre ocurrencias antes y después de la coronografía junto con la función de distribución exponencial utilizando la intensidad estimada para el Sector 2A. Se limita a mostrar solo este sector debido a que no se presentan casos atípicos o relevantes para un análisis detallado.

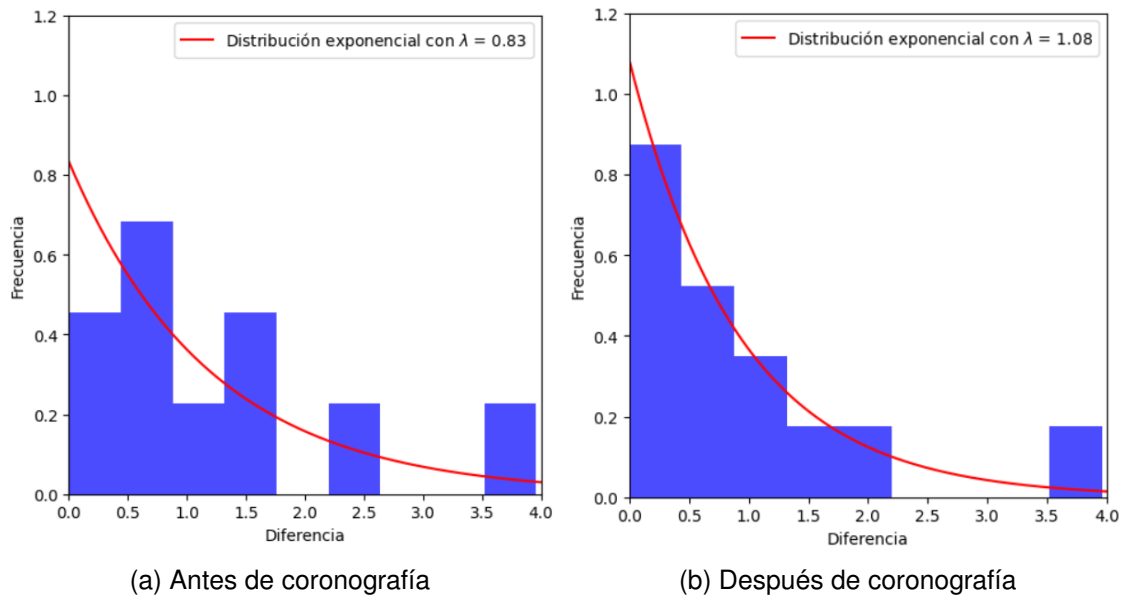


Figura 4.7: Distribución de tiempos entre ocurrencias antes y después de la coronografía junto con la función de distribución exponencial de la intensidad estimada.

Ahora se quiere saber si existe un impacto significativo de las coronografías sobre las interrupciones eléctricas, para esto se analizan los tiempos entre ocurrencias en los periodos anterior y posterior a las inspecciones en cada sector. Se aplica el test t utilizando las dos muestras de tiempos entre ocurrencias para determinar si existe una variación significativa.

En la siguiente tabla se muestran los valores estimados de $\hat{\lambda}$ antes y después de la coronografía en cada sector, así como los resultados del test t aplicado a las dos muestras de tiempos entre ocurrencias.

Sector	$\hat{\lambda}$ Antes coronografía	$\hat{\lambda}$ Después coronografía	Test t
Sector 2A	0.833	1.083	Acepta H_0
Sector 4A	1.780	2.168	Acepta H_0
Sector 5A	0.581	0.402	Acepta H_0
Sector 6A	1.156	1.091	Acepta H_0
Sector 7A	0.933	0.659	Acepta H_0
Sector 8A	0.502	0.992	Acepta H_0
Sector 9A	0.990	0.730	Acepta H_0

Cuadro 4.7: Lambdas mensuales antes y después de coronografías por sector junto con resultados del test t al 95 % de confianza aplicado a muestras de tiempos entre ocurrencias.

Se puede concluir que no hay evidencia suficiente para afirmar que las coronografías tienen un impacto significativo en las interrupciones eléctricas en ninguno de los sectores homogéneos. Esto se desprende de los resultados del test t, que indican que no hay una variación significativa en los tiempos entre ocurrencias antes y después de las inspecciones en cada sector.

4.2.3. Impacto de las coronografías en sectores no homogéneos

Para evaluar el impacto de las inspecciones coronográficas en los sectores no homogéneos, es necesario estimar la función de intensidad acumulada junto con sus intervalos de confianza en los periodos antes y después de las inspecciones coronográficas.

Antes de comenzar, se muestra la tabla que detalla las fechas en las que se llevaron a cabo las inspecciones coronográficas en cada sector no homogéneo.

Sector no homogéneo	Fecha de Inspección Coronográfica
Sector 1A	2022/01/28 y 2022/11/29
Sector 3A	2022/12/05

Cuadro 4.8: Fechas de inspecciones coronográficas por sector no homogéneo de la Zona A.

En la construcción del modelo no paramétrico propuesto por Leemis en [7] se logra estimar la función de intensidad acumulada $\Lambda(t)$ asociada a un proceso $N(t)$ para un intervalo $(0, T]$. Si consideramos el conteo de interrupciones eléctricas en los periodos de estudio antes y después de las coronografías se tienen dos procesos $N_1(t)$ y $N_2(t)$ para un intervalo $(0, T_c]$.

Si tomamos como ejemplo el Sector 1A, el periodo antes de la primera coronografía es desde 1 de febrero de 2021 hasta el 31 de diciembre de 2021, son aproximadamente 330 días. Por lo tanto, el intervalo donde se estima la función de intensidad acumulada es $(0, 330]$ donde se tienen 11 subintervalo de longitud 30. La función acumulada estimada quedaría de la siguiente manera.

$$\hat{\Lambda}(t) = \left(\sum_{j=1}^{i-1} n_j \right) + \frac{n_i (t - a_{i-1})}{(a_i - a_{i-1})}, \quad a_{i-1} < t \leq a_i \quad (4.4)$$

Además, se puede estimar su intervalo de confianza de la siguiente manera:

$$\hat{\Lambda}(t) - z_{\alpha/2} \sqrt{\hat{\Lambda}(t)} < \Lambda(t) < \hat{\Lambda}(t) + z_{\alpha/2} \sqrt{\hat{\Lambda}(t)} \quad (4.5)$$

El periodo posterior a la coronografía es desde 1 de febrero de 2022 hasta el 31 de diciembre de 2022, por lo tanto, se considera el mismo intervalo $(0, 330]$ para que sean comparables las trayectorias. Sin embargo, para estimar la función de intensidad acumulada igualmente se utiliza 4.4, solo varía la cantidad de interrupciones eléctricas por subintervalos. Para los intervalos de confianza es similar a 4.5, solo que ahora se tendrá una nueva función $\hat{\Lambda}$.

Utilizando esta metodología, nos permite tener trayectorias estimadas en periodos de estudios antes y después de las coronografías por cada sector. La finalidad es graficar

ambas trayectorias junto con sus intervalos de confianza para observar si existen variaciones relevantes.

Siguiendo con el análisis en el Sector 1A, a continuación se presentan las funciones acumuladas estimadas antes y después de la primera coronografía, junto con sus respectivos intervalos de confianza:

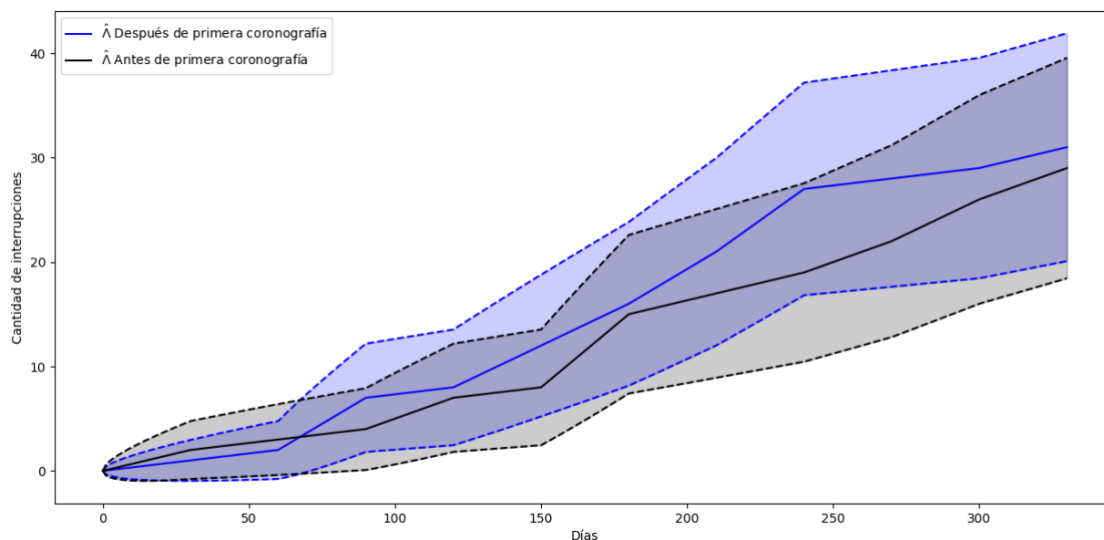


Figura 4.8: Funciones de intensidad acumulada de interrupciones eléctricas: Antes (azul) y después (negro) de primera coronografía. Las líneas punteadas representan las bandas de confianza al 95 %

Se observa que cada trayectoria está incluida en el intervalo de confianza de la otra, lo que sugiere que no hay una diferencia estadísticamente significativa entre ellas en el nivel de confianza dado. Por lo tanto, no se evidencia un impacto de la primera coronografía en las interrupciones eléctricas.

Para la segunda coronografía en el Sector 1A se realiza la misma metodología de estimación, obteniendo dos funciones de intensidad acumulada con sus respectivos intervalos de confianza. Se representa en el siguiente gráfico:

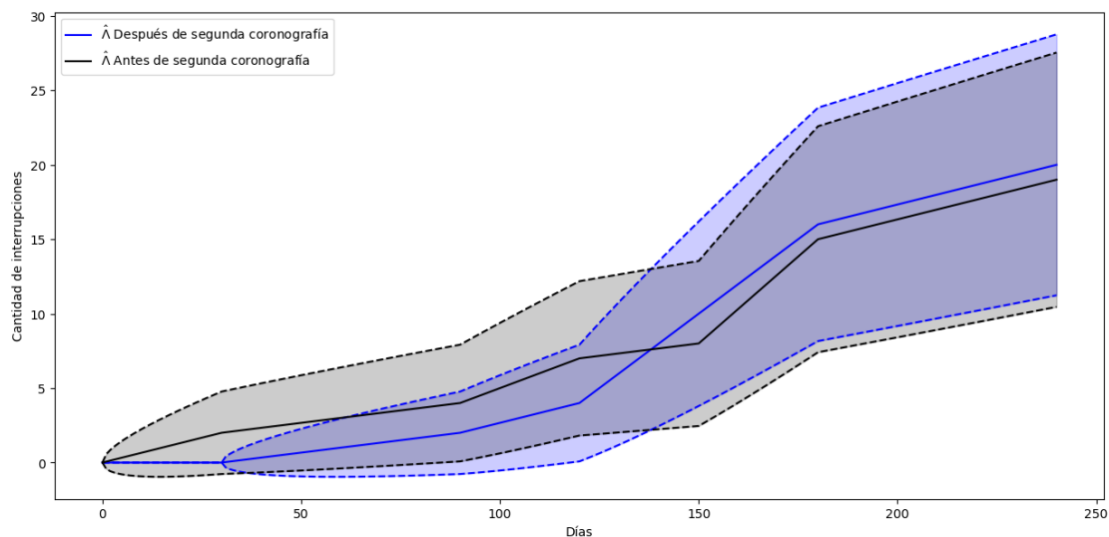


Figura 4.9: Funciones de intensidad acumulada de interrupciones eléctricas: Antes (azul) y después (negro) de segunda coronografía. Las líneas punteadas representan las bandas de confianza al 95 %

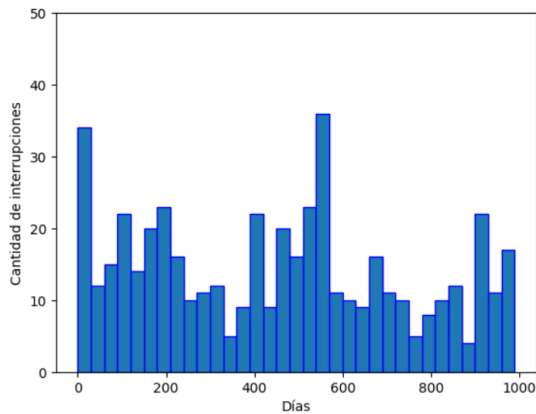
Al igual que en la primera coronografía, cada trayectoria está incluida en el intervalo de confianza de la otra. Por lo tanto tampoco se evidencia un impacto significativo en las interrupciones eléctricas.

Se limita a mostrar solo los resultados del Sector 1A debido a que el Sector 3A tiene un comportamiento similar al presentado en la Figura 4.8. Por lo tanto, en ningún sector no homogéneo de la Zona A se tiene un impacto relevante de las coronografías.

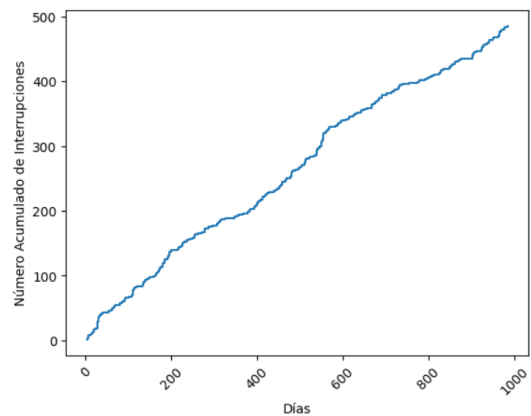
4.3. Análisis de la Zona B

El análisis en la Zona B se realiza siguiendo la misma metodología que en la Zona A, lo que implica el uso de las mismas herramientas para evaluar la homogeneidad, estacionalidad e impacto de las coronografías en las interrupciones eléctricas.

Se observan las interrupciones en la Zona B de Valparaíso a través de un gráfico de interrupciones eléctricas agrupadas por meses y el proceso de conteo asociados presentes en la Figura 4.2.



(a) Frecuencia de interrupciones Zona B



(b) Conteo de interrupciones Zona B

Figura 4.10: Frecuencia de interrupciones eléctricas agrupadas por meses y proceso de conteo en la Zona B.

En cuanto al proceso de conteo se ve con ciertos picos claros de intensidad, como por ejemplo en el periodo cercano al día 0 y en el día 600. Para verificar si esto ocurre debido a las estaciones del año se realizará un análisis.

Estacionalidad Zona B

La frecuencia de interrupciones eléctricas puede estar influenciada por las condiciones climáticas y las temperaturas como se vio anteriormente. Por lo tanto, se dividen los meses en dos periodos: verano e invierno. Cada periodo se identificará con un color correspondiente, como se muestra en la Figura 4.11.

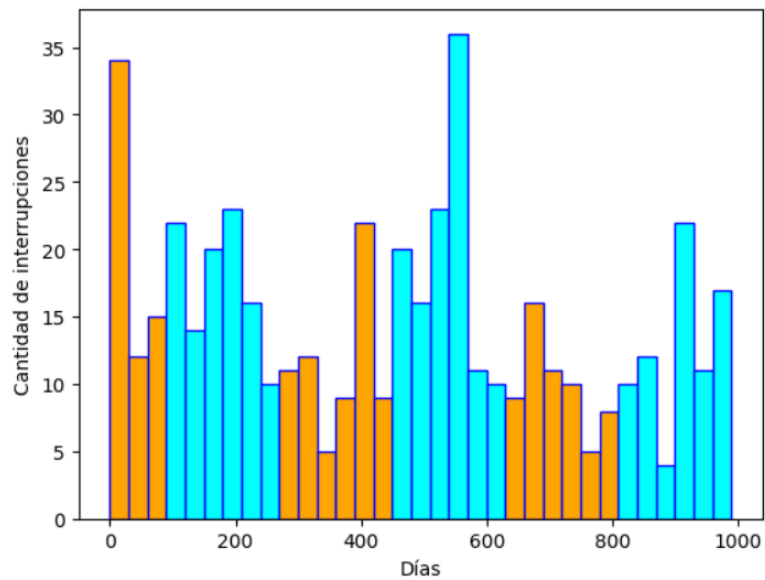


Figura 4.11: Frecuencia de interrupciones eléctricas en la Zona B separadas por periodos de veranos (naranja) e inviernos (celeste).

A primera vista, se observa una mayor cantidad de interrupciones en invierno. Sin embargo, para determinar la relevancia de estas variaciones, primero se estima el valor de la intensidad λ mensual para los veranos e inviernos. Luego, se consideran como dos muestras independientes los tiempos entre ocurrencias de ambos periodos y se les aplica el test t para evaluar si existe variación significativa.

Utilizando el estimador $\hat{\lambda}$ se puede graficar la distribución exponencial estimada y la distribución empírica de los tiempos entre ocurrencias como se ve en la Figura 4.12.

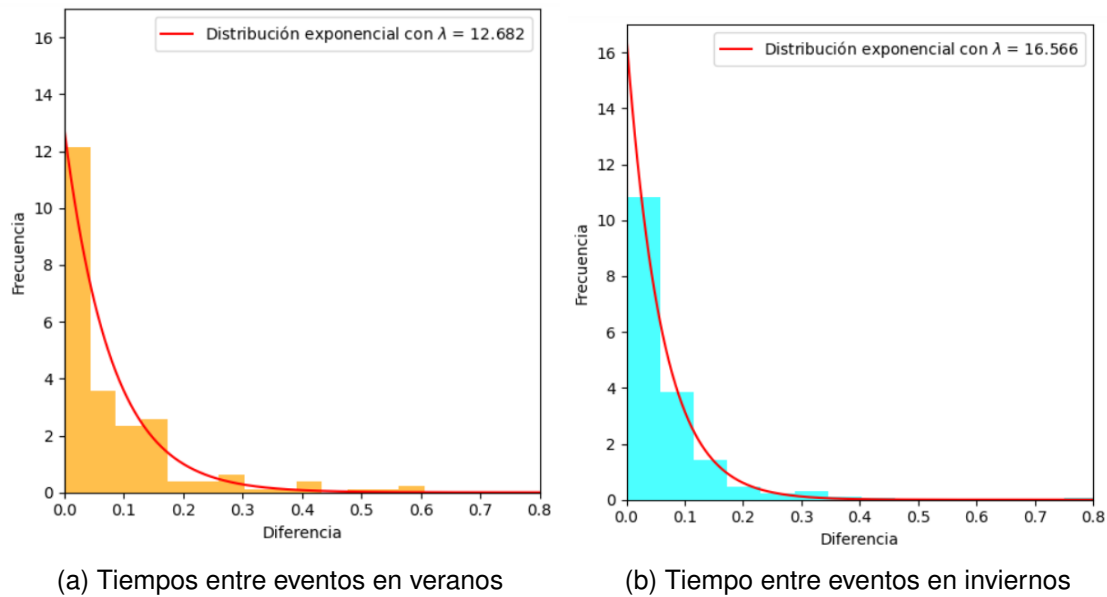


Figura 4.12: Distribución de tiempos entre ocurrencias para los periodos de verano (naranja) e invierno (celeste) en la Zona B junto con la distribución exponencial del lambda estimado.

Se puede ver que para los tiempos de ocurrencia del verano se tiene una intensidad estimada $\hat{\lambda} = 12,682$ y para el invierno se tiene que una intensidad estimada $\hat{\lambda} = 16,566$, existe un aumento en cuanto las interrupciones de manera mensual en el invierno.

A continuación, se presentan los resultados del test t sobre ambas muestras de tiempos entre ocurrencias y los $\hat{\lambda}$ de verano e invierno.

Zona	$\hat{\lambda}$ Veranos	$\hat{\lambda}$ Inviernos	Test t
Zona B	12.682	16.566	Rechaza H_0

Cuadro 4.9: Lambdas mensuales de veranos e inviernos de la Zona B junto con resultados de test t al 95 % de confianza aplicado a las muestras de los tiempos entre ocurrencias.

Dado que el test t rechaza la hipótesis nula H_0 , se asume existencia estacionalidad en la Zona B.

Homogeneidad Zona B

Ahora se quiere evaluar estadísticamente si la Zona B se comporta de manera homogénea, para esto se utilizan los test de hipótesis enfocados a la homogeneidad obteniendo los siguientes resultados.

Sector	<i>Chi-cuadrado C.</i>	<i>Neyman-Scott</i>	Razón de verosimilitud
Zona B	Rechaza H_0	Rechaza H_0	Rechaza H_0

Cuadro 4.10: Resultados de los test de homogeneidad al 95 % de confianza en la Zona B.

Por lo tanto, se considera que esta zona tiene un comportamiento no homogéneo, esto quiere decir que es correcto asumir $\lambda_1 \neq \dots \neq \lambda_n$.

4.3.1. Sectores de la Zona B

Las inspecciones coronográficas en la Zona B se dividen por sector. Por lo tanto, es importante separar las observaciones por cada sector de la Zona B para facilitar el análisis individual de cada uno. Al separar los datos y graficar el proceso de conteo asociado a cada sector se tienen las trayectorias observadas en la Figura 4.13.

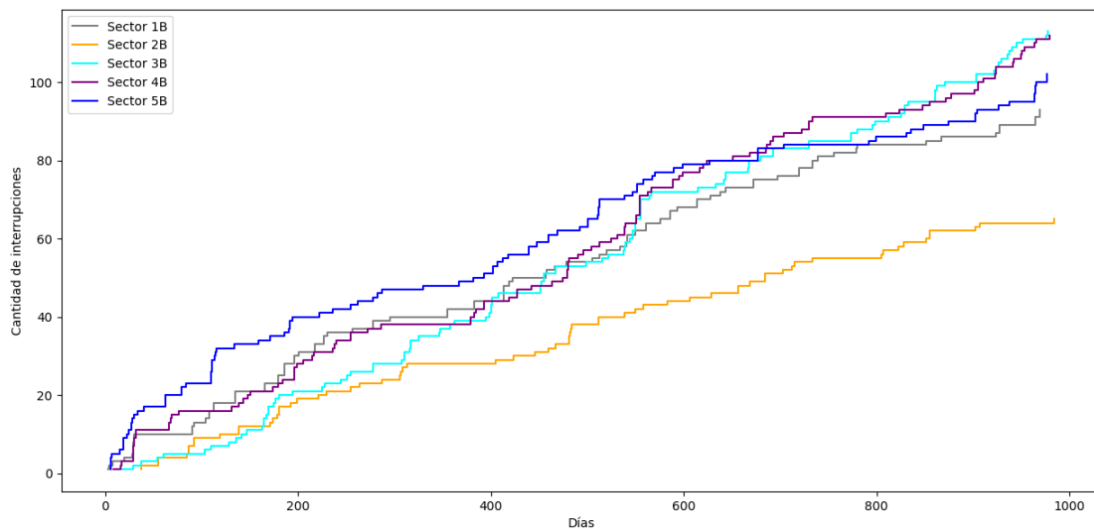


Figura 4.13: Procesos de conteo de interrupciones eléctricas por sectores en la Zona B.

Estacionalidad por sector

Es importante analizar la estacionalidad en cada sector, tal como se destacó anteriormente. Tomar intervalos no representativos antes y después de las coronografías podría generar errores en la evaluación del impacto de las inspecciones. Por lo tanto, se realiza un análisis de estacionalidad para cada sector de la Zona B, los resultados obtenidos son los presentes en la Tabla 4.4.

Sector no homogéneo	$\hat{\lambda}$ Veranos	$\hat{\lambda}$ Inviernos	Test t
Sector 1B	2.251	3.455	Acepta H_0
Sector 2B	1.873	2.185	Acepta H_0
Sector 3B	2.849	4.021	Acepta H_0
Sector 4B	2.466	4.397	Rechaza H_0
Sector 5B	3.001	3.219	Acepta H_0

Cuadro 4.11: Intensidad estimada mensualmente por cada sector en veranos e inviernos por sector junto con resultados del test t al 95 % de confianza aplicado a muestras de tiempos entre ocurrencias.

Los resultados muestran que las intensidades $\hat{\lambda}$ para los periodos de invierno tienden a ser notablemente mayores. Además, solo el Sector 4B presenta una variación significativa en cuanto a los tiempos entre ocurrencias de veranos e inviernos. Esto indica clara la existencia de estacionalidad en este sector.

Para observar la estacionalidad del Sector 4B se muestra a continuación las cantidades de interrupciones eléctricas agrupadas mensualmente, con los meses diferenciados por periodos de invierno y verano.

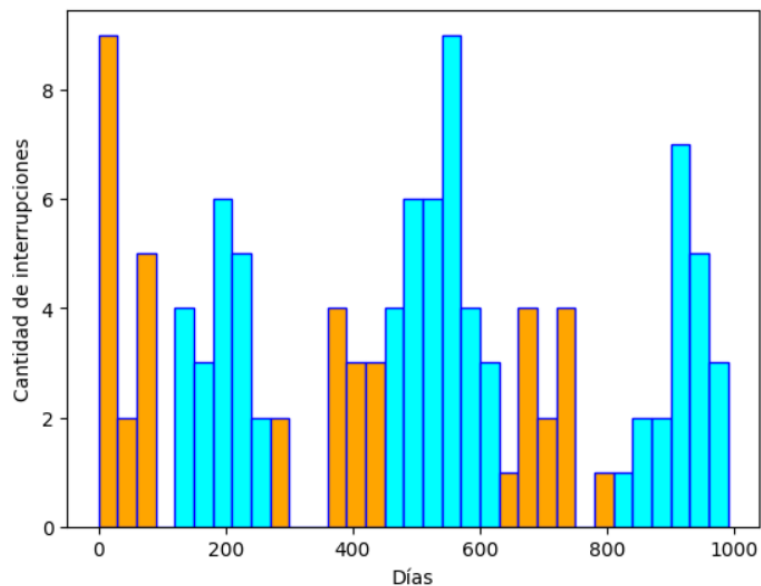


Figura 4.14: Frecuencia de interrupciones eléctricas en el Sector 4B separadas por periodos de veranos (naranja) e inviernos (celeste).

Periodos de estudio antes y después de coronografías

Según los resultados de la Tabla 4.11 solo para el Sector 4B se debe tener en consideración la estacionalidad al tomar los tiempos de estudio antes y después de las coronografía. Sin embargo, igualmente se presentan variaciones en las intensidades estimadas de inviernos y veranos para los demás sectores, por lo tanto, los periodos de estudio se tomarán para todos.

El cómo seleccionar los periodos de estudios antes y después de las inspecciones se

detallaron en la subsección 4.2.1.

Homogeneidad por sector

Ya que se dispone del estudio de estacionalidad por sector, es fundamental evaluar la homogeneidad en cada uno de ellos para aplicar el enfoque adecuado en la evaluación del impacto de la coronografía. Para esto, se aplican los tres tests de homogeneidad, los resultados de homogeneidad de cada sector se presentarán en la Tabla 4.12.

Sector	<i>Chi-cuadrado C.</i>	<i>Neyman-Scott</i>	Razón de verosimilitud
Sector 1B	Rechaza H_0	Rechaza H_0	Rechaza H_0
Sector 2B	Acepta H_0	Acepta H_0	Acepta H_0
Sector 3B	Rechaza H_0	Rechaza H_0	Rechaza H_0
Sector 4B	Rechaza H_0	Rechaza H_0	Rechaza H_0
Sector 5B	Rechaza H_0	Rechaza H_0	Rechaza H_0

Cuadro 4.12: Resultados de los test de homogeneidad al 95 % de confianza en los sectores de la Zona B.

Al utilizar los tests, se observa que la mayoría de los sectores se consideran no homogéneos, excepto el Sector 2B. Comparando la información de la Tabla 4.12 con la representación visual de los procesos en la Figura 4.13, se puede notar que los sectores con mayor cantidad de interrupciones son precisamente aquellos clasificados como no homogéneos.

4.3.2. Impacto de las coronografías en sectores homogéneos

Antes de comenzar el análisis, se presenta una tabla que detalla las fechas en las que se llevaron a cabo las inspecciones coronográficas en el sector homogéneo de la Zona B.

Sector homogéneo	Fecha de Inspección Coronográfica
Sector 2B	2023/01/27

Cuadro 4.13: Fechas de inspecciones coronográficas en cada sector homogéneo de la Zona B.

Para evaluar el impacto de las inspecciones coronográficas primero se debe estimar la función de intensidad λ antes y después de la coronografía utilizando el estimador 4.1. A continuación se presenta la distribución de tiempos entre ocurrencias antes y después de la coronografía junto con las funciones de distribución exponencial utilizando las intensidades estimadas para el Sector 2B.

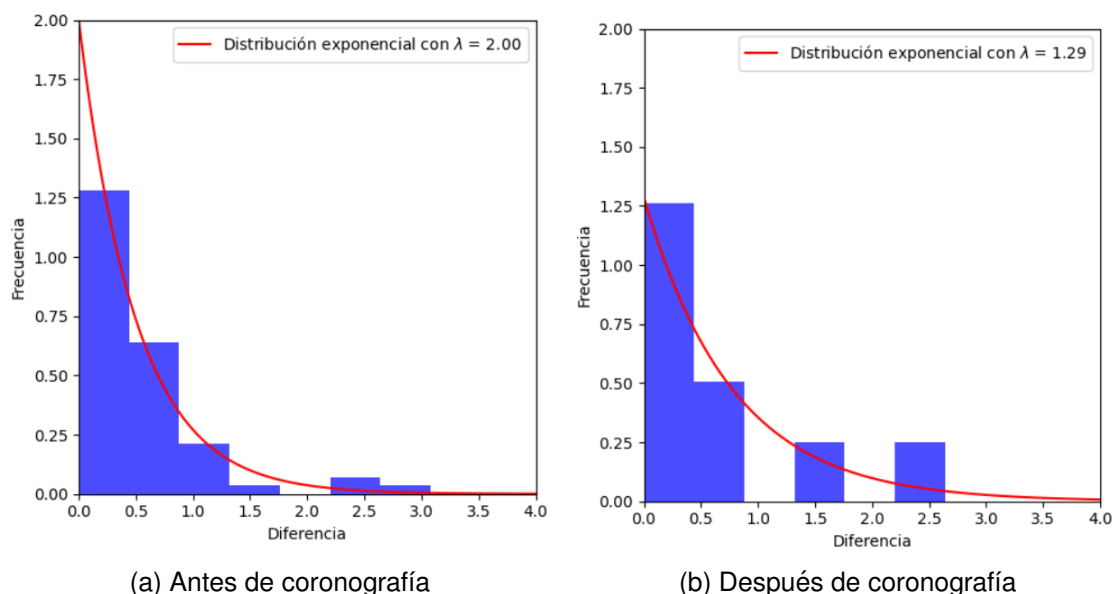


Figura 4.15: Distribución de tiempos entre ocurrencias antes y después de la coronografía junto con la función de distribución exponencial de la intensidad estimada.

Ahora se quiere saber si existe un impacto significativo de las coronografías sobre las interrupciones eléctricas, para esto se analizan los tiempos entre ocurrencias en los periodos anterior y posterior a la inspección. Se aplica el test t utilizando las dos muestras de tiempos entre ocurrencias para determinar si existe una variación significativa.

En la siguiente tabla se muestran los valores estimados de $\hat{\lambda}$ antes y después de la

coronografía en cada sector, así como los resultados del test t aplicado a las dos muestras de tiempos entre ocurrencias.

Sector	$\hat{\lambda}$ Antes coronografía	$\hat{\lambda}$ Después coronografía	Test t
Sector 2B	2.000	1.286	Acepta H_0

Cuadro 4.14: Lambdas mensuales antes y después de coronografías por sector junto con resultados del test t al 95 % de confianza aplicado a muestras de tiempos entre ocurrencias.

Se puede concluir que no hay evidencia suficiente para afirmar que las coronografías tienen un impacto significativo en las interrupciones eléctricas en el Sector 2B. Esto se desprende de los resultados del test t, que indican que no hay una variación significativa en los tiempos entre ocurrencias antes y después de la inspección.

4.3.3. Impacto de las coronografías en sectores no homogéneos

Antes de comenzar, se muestra la tabla que detalla las fechas en las que se llevaron a cabo las inspecciones coronográficas en cada sector no homogéneo.

Sector no homogéneo	Fecha de Inspección Coronográfica
Sector 1B	2023/01/23
Sector 3B	2023/01/19
Sector 4B	2023/02/10
Sector 5B	2023/02/09

Cuadro 4.15: Fechas de Inspecciones coronográficas por sector no homogéneo de la Zona B.

Para evaluar el impacto de las coronografía sobre las interrupciones eléctricas en los periodos antes y después de las coronografías se sigue la metodología y estimaciones vistas en 4.2.3.

Para mostrar el análisis realizado en el Sector 4B, a continuación se presentan las

funciones acumuladas estimadas antes y después de la coronografía, junto con sus respectivos intervalos de confianza.

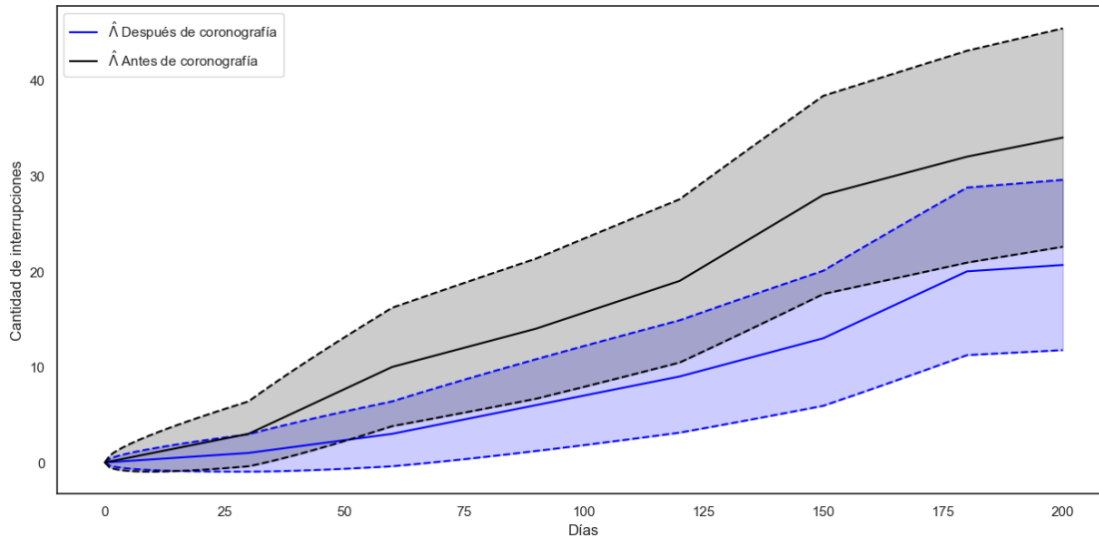


Figura 4.16: Funciones de intensidad acumulada de interrupciones eléctricas: Antes (azul) y después (negro) de la coronografía. Las líneas punteadas representan las bandas de confianza al 95 %

Se observa que cada trayectoria no está incluida en el intervalo de confianza de la otra, lo que sugiere que hay una diferencia estadísticamente significativa entre ellas en el nivel de confianza dado. Además, la trayectoria correspondiente al periodo después de la coronografía está más abajo. Por lo tanto, se puede decir que la inspección coronográfica pudo haber tenido un impacto positivo sobre las interrupciones eléctricas.

Se limita a mostrar solamente el resultado del Sector 4B debido a que los demás presentan un comportamiento similar, indicando un posible impacto positivo de las coronografías en los sectores no homogéneos de la Zona B.

4.4. Análisis de los resultados encontrados

Se abordó el fenómeno de interrupciones eléctricas en las Zonas A y B con la finalidad de describirlo y comprenderlo, para esto se analizó su estacionalidad y homogeneidad.

Al evaluar la frecuencia de eventos por meses se obtuvo que la Zona A tiene un comportamiento homogéneo, mientras que la Zona B no. En cuanto a la estacionalidad, existe un aumento de interrupciones eléctricas en invierno en ambas zonas, pero solo la Zona B presenta variaciones significativas.

Se procede con el estudio de los sectores ubicados en las Zonas A y B. En cuanto a la estacionalidad, considerando periodos de invierno y verano, se encontró que en todos los sectores las interrupciones eléctricas son más frecuentes en invierno. Sin embargo, solo los Sectores 1A y 4B muestran una variación estadísticamente significativa entre veranos e inviernos.

Al analizar la homogeneidad por sectores, se observó que más de la mitad poseen tasas homogéneas, mientras que otros presentan mayor variabilidad. Se empleó modelación basada en PPH para los procesos con tasas constantes en el tiempo, mientras que para aquellos con tasas variables se recurrió a una estimación no paramétrica utilizando PPNH.

Finalmente, al evaluar el impacto de las coronografías en las interrupciones eléctricas, solo se observan resultados significativos en los sectores no homogéneos de la Zona B. En estos sectores, la intensidad acumulada estimada antes de las inspecciones tiende a ser considerablemente mayor que la estimada después de su realización. Por lo tanto, se puede concluir que las inspecciones podrían haber tenido un impacto positivo en la reducción de las interrupciones eléctricas en dichos sectores.

Capítulo 5

Conclusiones generales

El objetivo general de este trabajo ha sido automatizar el proceso de clasificación de imágenes coronográficas y llevar a cabo un análisis del historial de interrupciones eléctricas para evaluar el impacto potencial de las inspecciones preventivas en el período de estudio comprendido entre el 1 de enero de 2021 y el 18 de septiembre de 2023.

A continuación, se presentan las conclusiones derivadas de los dos objetivos principales del estudio:

5.1. Clasificación automatizada de imágenes

En relación con el primer objetivo, se logró desarrollar y desplegar un algoritmo integrado en una plataforma web para la clasificación automática de imágenes coronográficas. La herramienta demostró un 100 % de precisión en las pruebas realizadas, basándose en el OCR PyTesseract. El análisis reveló que la Zona A presenta una menor cantidad de artefactos eléctricos dañados en comparación con la Zona B, lo que sugiere que las líneas de distribución en la Zona B están más afectadas y podrían requerir mayor atención.

5.2. Análisis de interrupciones eléctricas

En cuanto al segundo objetivo, se realizó un análisis detallado de las interrupciones eléctricas en las Zonas A y B. Se observó que la Zona A muestra un comportamiento homogéneo en la frecuencia de eventos a lo largo de los meses, mientras que la Zona B no presenta dicha homogeneidad. Ambas zonas experimentan un incremento en las interrupciones eléctricas durante el invierno, aunque solo en la Zona B estas variaciones son estadísticamente significativas.

Al desglosar el análisis por sectores, se identificó que las interrupciones eléctricas son más frecuentes en invierno en todos los sectores, pero solo los Sectores 1A y 4B muestran variaciones significativas entre estaciones. Asimismo, la mayoría de los sectores presentan tasas homogéneas, mientras que otros muestran mayor variabilidad, requiriendo modelos no paramétricos para una representación precisa.

Finalmente, se evaluó el impacto de las coronografías en las interrupciones eléctricas. Se encontró que en los sectores no homogéneos de la Zona B, la intensidad de las interrupciones antes de las inspecciones era significativamente mayor en comparación con después de las mismas. Esto sugiere que las inspecciones preventivas podrían haber contribuido a una reducción de las interrupciones eléctricas en estos sectores específicos.

Bibliografía

- [1] LB Adewole, EO Ogunleye, OE Faniyan, SA Adewale, and CY Daramola. Development of serial number extractor for nigerian currencies. University of Ibadan Journal of Science and Logics in ICT Research, 9(1), 2023. 29
- [2] Jaied Ai. Easyocr. Obtenido de <https://www.jaied.ai/easyocr/documentation/>, year=2024 . 37
- [3] Lawrence D Brown and Linda H Zhao. A test for the poisson distribution. Sankhyā: The Indian Journal of Statistics, Series A, pages 611–625, 2002. 15, 26
- [4] Akhil Chawla, Aarushi Gupta, and KS Shushrutha. Intelligent information retrieval: Techniques for character recognition and structured data extraction. Journal of Emerging Technologies and Innovative Research (JETIR), 9(7), 2022. 14
- [5] Haziq Idrose, Nouar AIDahoul, Hezerul Abdul Karim, Rehan Shahid, and Manish Kumar Mishra. An evaluation of various pre-trained optical character recognition models for complex license plates. In Multimedia University Engineering Conference (MECON 2022), pages 21–27. Atlantis Press, 2022. 29, 39
- [6] Meagan Lang. Pytesseract documentation. Obtenido de <https://pytesseract.readthedocs.io/en/latest/>, 2015. 35
- [7] Lawrence M Leemis. Nonparametric estimation and variate generation for a nonhomogeneous poisson process from event count data. IIE Transactions, 36(12):1155–1160, 2004. 22, 24, 61
- [8] Mario Lefebvre. Applied stochastic processes. Springer Science & Business Media, 2007. 20
- [9] Zhenyu Li, Licheng Li, Xingliang Jiang, Jhen-Jia Hu, Zhijin Zhang, and Wei Zhang. Effects of different factors on electrical equipment uv corona discharge detection. Energies, 9:369, 05 2016. 9

- [10] Fausto Morales. Kerasocr documentation. Obtenido de , 6:75–87, 2017. 9
- [12] Sheldon M Ross. Introduction to probability models. Academic press, 2014. 15, 17
- [13] G. Sibilant. High-Voltage Direct Current Corona Testing of Transmission Line Hardware and Insulator Assemblies: Development of Test Methodology. PhD thesis, Electric Power Research Institute, 2013. 9
- [14] UViRCO Technologies. CoroCam 6D: Manual de instrucciones. UViRCO Technologies, 2015. 12
- [15] S Vijayarani and A Sakila. Performance comparison of ocr tools. International Journal of UbiComp (IJU), 6(3):19–30, 2015. 13