



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

**GENERACIÓN DE INVENTARIO DE EMISIONES EN QUEMAS  
FORESTALES Y AGRÍCOLAS**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL ELÉCTRICA

**MARÍA JESÚS UGARTE CARABALL**

PROFESOR GUÍA:  
HANS ALBERT LÖBEL DÍAZ

MIEMBROS DE LA COMISIÓN:  
FELIPE ANTONIO SALAS BRAVO  
NICOLÁS JORGE HUNEEUS LAGOS

Este trabajo ha sido parcialmente financiado por: SPIKE SPA

SANTIAGO DE CHILE  
2022

*No se trata de lo que es.  
Se trata de lo que puede llegar a ser*

***Dr. Seuss***

# Agradecimientos

Me gustaría agradecer primero a mis papás, que siempre me apoyaron, me dieron todas las oportunidades y me motivaron a hacer todo lo que quise. Gracias por todo el cariño y por darme el mejor ejemplo.

Gracias a mis hermanos por su infinita ayuda. A la Caro, por todo su apoyo cuando entré a la universidad. A Cristián por su buena onda y disposición a ayudar. A Martín, por siempre estar disponible para ayudarme, enseñarme y guiarme a estar hacer las cosas bien.

Pancho, gracias por ser siempre la mejor compañía y darme todo tu apoyo. Gracias por siempre estar ahí para mí y estar dispuesto a ayudarme con lo que fuera.

También quiero agradecer a Hans por aceptar ser mi profesor guía y por enseñarme tanto en la primera práctica. Además, me gustaría agradecer a Nicolás Huneus por ayudarme a buscar que este trabajo tengo una utilidad real.

Gracias a todos los de Spike por toda su ayuda y buena onda. En especial quiero agradecer Pipe por guiarme en todo el proceso de memoria, además de estar siempre dispuesto enseñarme cosas nuevas y ayudarme. También, a Juan Pablo por darme la oportunidad de hacer la memoria en Spike.

A mis amigas del colegio, Rosario, Valeria, Feña, Maca, Coni y Cata, por su infinita buena onda y su disposición a juntarnos a hablar y reírnos. Por otra parte, quiero agradecer a todos mis amigos de la universidad, en especial a la Xavi, la Vale y Nico, que fueron un gran apoyo en estos 6 años.

# Tabla de Contenido

<b>1. Introducción</b>	<b>1</b>
<b>2. Marco Teórico y Estado del Arte</b>	<b>3</b>
2.1. Marco Teórico . . . . .	3
2.1.1. Machine Learning . . . . .	3
2.1.1.1. Redes Neuronales Convolucionales . . . . .	3
2.1.1.2. Redes Neuronales Residuales . . . . .	6
2.1.2. Red Neuronal UNet . . . . .	7
2.1.3. <i>Wildfires</i> y emisiones . . . . .	9
2.2. Estado del Arte . . . . .	9
2.2.1. <i>Fire Radiative Power</i> . . . . .	9
2.2.2. Área Quemada . . . . .	11
<b>3. Adquisición de Datos</b>	<b>12</b>
3.1. Imágenes Satelitales . . . . .	12
3.2. Datos de Vegetación . . . . .	15
<b>4. Metodología</b>	<b>18</b>
4.1. Estimación de Área Quemada . . . . .	18
4.1.1. Pre-Procesamiento . . . . .	18
4.1.2. Estimación del Modelo de Machine Learning . . . . .	19
4.2. Cálculo de Emisiones . . . . .	21
4.3. Registro Temporal de Emisiones . . . . .	22
4.3.1. Generación de Máscaras de Área Quemada . . . . .	23
4.3.2. Datos tabulares del Área Quemada . . . . .	23
4.3.3. Pre-procesamiento de Mapa de Uso de Suelo . . . . .	24
4.3.3.1. Proyección de coordenadas . . . . .	24
4.3.3.2. Transformación a Formato Tabular . . . . .	25
4.3.4. Asignación de Tipo de Uso de Suelo . . . . .	25
4.3.5. Cálculo de Emisiones . . . . .	26
4.4. Generación de Interfaz de Usuario . . . . .	27
4.4.1. Estimación de Área Quemada a Partir de Imagen Proporcionada por el Usuario . . . . .	27
4.4.2. Registro Temporal de Emisiones y Área Quemada . . . . .	28
<b>5. Resultados Obtenidos</b>	<b>29</b>
5.1. Estimación de Área Quemada . . . . .	29
5.1.1. Ejemplos de Desempeño . . . . .	29

5.1.1.1.	Falsos Positivos . . . . .	29
5.1.1.2.	Falsos Negativos . . . . .	32
5.1.1.3.	Verdaderos Positivos . . . . .	33
5.1.1.4.	Verdaderos negativos . . . . .	35
5.2.	Registro Temporal . . . . .	37
5.2.1.	Datos Tabulares de Área Quemada . . . . .	38
5.2.2.	Datos Tabulares de Tipo de Uso de Suelo . . . . .	38
5.2.3.	Registro temporal de emisiones . . . . .	39
5.3.	Interfaz de Usuario . . . . .	40
5.3.1.	Estimación de Emisiones por Imagen Ingresada . . . . .	41
5.3.2.	Registro temporal . . . . .	43
<b>6.</b>	<b>Discusión</b> . . . . .	<b>47</b>
6.1.	Modelo de Estimación de Emisiones . . . . .	47
6.2.	Registro Temporal de Emisiones . . . . .	48
6.3.	Interfaz de Usuario . . . . .	52
<b>7.</b>	<b>Conclusiones</b> . . . . .	<b>54</b>
	<b>Bibliografía</b> . . . . .	<b>56</b>
	<b>Anexos</b> . . . . .	<b>58</b>
	<b>Anexo A. Satélites</b> . . . . .	<b>58</b>
A.1.	Sentinel . . . . .	58
A.2.	MODIS . . . . .	59
	<b>Anexo B. Gases y Partículas</b> . . . . .	<b>60</b>
B.1.	Dióxido de Carbono (CO <sub>2</sub> ) . . . . .	60
B.2.	Monóxido de Carbono (CO) . . . . .	60
B.3.	Metano (CH <sub>4</sub> ) . . . . .	60
B.4.	Dihidrógeno (H <sub>2</sub> ) . . . . .	61
B.5.	Gases Reactivos que Contienen Nitrógeno y Oxígeno (NO <sub>x</sub> ) . . . . .	61
B.6.	Dióxido de Nitrógeno(NO <sub>2</sub> ) . . . . .	61
B.7.	Compuestos Orgánicos No Metánicos (NMOC) . . . . .	61
B.8.	Hidrocarburos sin Metano (NMHC) . . . . .	62
B.9.	Dióxido de Azufre (SO <sub>2</sub> ) . . . . .	62
B.10.	Amoníaco (NH <sub>3</sub> ) . . . . .	62
B.11.	Partículas Finas (PM2.5) . . . . .	62
B.12.	Materia Particulada Total (TPM) . . . . .	63
B.13.	Carbono Orgánico (OC) . . . . .	63
B.14.	Carbono Negro (BC) . . . . .	63
	<b>Anexo C. Sistemas de Coordenadas Geográficas</b> . . . . .	<b>64</b>
C.1.	Latitud y Longitud en Grados . . . . .	64
C.1.1.	Latitud . . . . .	64
C.1.2.	Longitud . . . . .	65
C.2.	Universal Transverse Mercator . . . . .	65

<b>Anexo D. Funciones Utilizadas</b>	<b>66</b>
D.1. Pre-procesamiento . . . . .	66
D.1.1. Añadir Margen . . . . .	66
D.1.2. Dividir Imagen . . . . .	66
D.2. Procesamiento . . . . .	67
D.3. Post-procesamiento . . . . .	67
D.3.1. Reconstrucción de la Imagen . . . . .	67
D.3.2. Sistema de Umbrales . . . . .	68

# Índice de Tablas

2.1.	Tipos de uso de suelo utilizados en [15] . . . . .	10
3.1.	Códigos extraídos de [22] y clasificación de uso de suelo . . . . .	17
4.1.	Factores de emisión (g/kg de biomasa quemada) . . . . .	22

# Índice de Ilustraciones

2.1.	Ejemplo convolución . . . . .	4
2.2.	Max-Pooling . . . . .	5
2.3.	Ilustración <i>Fully Connected Layer</i> y neurona . . . . .	6
2.4.	Diagrama Red Neuronal Convolutiva . . . . .	6
2.5.	Diagrama <i>Skip Connection</i> . . . . .	7
2.6.	Diagrama UNet . . . . .	8
2.7.	Ejemplo Segmentación de Imágenes . . . . .	8
2.8.	Zonas consideradas en [16] . . . . .	11
3.1.	Imagen descargada de Sentinel-2A con las bandas SWIR, NIR y Red . . . . .	14
3.2.	Imagen descargada de Sentinel-2A con las bandas SWIR, NIR y Red . . . . .	15
3.3.	Mapa de uso de suelo obtenido de [22] . . . . .	16
4.1.	Ejemplo <i>output</i> modelo de segmentación utilizado . . . . .	19
5.1.	Ejemplo de falso positivo en zona seca, identificando las zonas erróneamente clasificadas como quemadas en círculos en la imagen de la izquierda. . . . .	30
5.2.	Ejemplo de falso positivo en zona urbana rodeada de vegetación, identificando las zonas erróneamente clasificadas como quemadas en círculos en la imagen de la izquierda. . . . .	31
5.3.	Ejemplo de falso positivo introducido por nubosidad . . . . .	31
5.4.	Ejemplo de falso negativo de incendios activos pequeños . . . . .	32
5.5.	Ejemplo de falso negativo incendio activo pequeño cercano a montañas secas . . . . .	33
5.6.	Ejemplo de identificación exitosa de <i>burn scar</i> rodeada de vegetación . . . . .	34
5.7.	Ejemplo de identificación exitosa en zonas con nieve . . . . .	34
5.8.	Ejemplo de identificación exitosa en zonas con nieve . . . . .	35
5.9.	Ejemplo de identificación de área no quemada en zonas vegetadas . . . . .	36
5.10.	Ejemplo de identificación de área no quemada en zonas mayormente secas . . . . .	37
5.11.	Ejemplo de identificación de área no quemada en zonas de territorio mixto . . . . .	37
5.12.	Primeras 4 filas de la tabla de píxeles de área quemada. . . . .	38
5.13.	Primeras 4 filas del <i>dataset</i> tabular de tipo de uso de suelo . . . . .	38
5.14.	Primeras 4 filas del registro temporal de emisiones en Chile . . . . .	40
5.15.	Botón para ingreso de datos por el usuario . . . . .	41
5.16.	Botón de ingreso de datos . . . . .	42
5.17.	Ingreso de datos adicionales para la estimación de emisiones en quemas . . . . .	43
5.18.	Selección de región y comuna en la interfaz de usuario . . . . .	44
5.19.	Selección de la fecha de inicio y la fecha de término de interés en la interfaz de usuario . . . . .	44
5.20.	Visualización del gráfico de área quemada en la interfaz de usuario . . . . .	45
5.21.	Visualización de porcentajes de tipos de uso de suelo quemados y casilla de multi-selección de gases de interés en la interfaz de usuario . . . . .	45



5.22.	Ejemplo de visualización de emisiones y botón de descarga de datos de la interfaz de usuario . . . . .	46
6.1.	Emisiones de CO <sub>2</sub> en la región de Magallanes y la Antártica Chilena según el registro de la CONAF y el método empleado . . . . .	49
6.2.	Emisiones de CO <sub>2</sub> en la región de Arica y Parinacota según el registro de la CONAF y el método empleado . . . . .	50
6.3.	Emisiones de CO <sub>2</sub> en la Región de Coquimbo según el registro de la CONAF y el método empleado . . . . .	51
C.1.	Sistema de coordenadas Geográficas . . . . .	64
C.2.	Ilustración de las zonas UTM . . . . .	65

## GENERACIÓN DE INVENTARIO DE EMISIONES EN QUEMAS FORESTALES Y AGRÍCOLAS

Los inventarios de emisiones tienen como objetivo principal el estimar la cantidad de emisiones generadas por distintas fuentes en una zona y rango de tiempo específico. Actualmente en Chile, el registro de emisiones tiene una frecuencia anual y por comuna, la cual es baja dado los recursos satelitales que se encuentran hoy en día.

Las quemas son una gran fuente de emisión de gases de efecto invernadero. Estos gases afectan de forma tanto mundial como local. Aquellos seres vivos que son expuestos frecuentemente a estos gases pueden desarrollar enfermedades respiratorias no transmisibles. Es por esto que es importante mantener un registro de alta resolución, pudiendo identificar zonas en donde las seres vivos son expuestos de forma recurrente a las emisiones, y así tomar acciones de forma rápida.

En este trabajo se propone la generación de un inventario de emisiones para el territorio chileno con una alta resolución temporal. La base del proyecto realizado está en la utilización de imágenes satelitales y modelos de Machine Learning, con el fin de identificar zonas quemadas y, en conjunto de datos de vegetación, calcular sus emisiones. Con este fin, primero se hace un estudio del estado del arte de inventarios de emisiones en el mundo, tomando uno de ellos como método para el cálculo de emisiones. Luego, se seleccionan y descargan imágenes satelitales aptas para identificar áreas quemadas, las cuales tienen referencia temporal y espacial. En conjunto, se obtiene una base de datos de clasificación de uso de suelo para Chile de alta resolución espacial.

El método propuesto toma como entrada las imágenes satelitales, luego a través de un modelo de Machine Learning se estima el área quemada en ésta, entregando una estimación de probabilidad de zona quemada a cada píxel. A partir de esta probabilidad se estima el área quemada en la imagen y, tomando en cuenta su clasificación de uso de suelo, se calculan las emisiones de 15 gases y partículas diferentes. Este método se utiliza sobre las imágenes recolectadas del territorio chileno, creando un registro histórico de alta frecuencia temporal y espacial de emisiones. Finalmente, con el fin de facilitar el uso del modelo y la visualización del registro histórico, se genera una interfaz de usuario.

El método desarrollado es capaz de entregar, de forma casi instantánea, la estimación de área quemada, y con ella la estimación de emisiones de 15 diferentes gases y partículas. Esta metodología, al ser utilizada para la generación de un registro, gracias a los recursos satelitales actuales, puede permitir mantener un registro diario, con una resolución espacial de hasta 10 metros. El método propuesto presenta una posible solución a la baja resolución del registro actual.

# Capítulo 1

## Introducción

Según el Resumen Nacional de Ocurrencia y Daño por Incendios del año 2020 de la CONAF (Corporación Nacional Forestal [1]), en el período entre los años 2019 y 2020 hubo un total de 8.217 incendios en Chile. Es más, en este mismo resumen se puede notar un alza, cercana a 1.000, en la cantidad de quemas forestales al año desde el año 2016. Actualmente, para llevar un registro de los gases emitidos en los incendios, el Ministerio del Medio Ambiente mantiene el Registro de Emisiones y Transferencia de Contaminantes (RETC [2]), el cual entrega un total anual de emisiones en las distintas comunas de país.

Los incendios no solo producen el daño local al área quemada, también afecta a localidades viento abajo, dado que los gases y partículas emitidas generan efectos negativos, tanto en las personas que viven en el sector como al medio ambiente. Es por esto que un registro anual presenta una resolución temporal deficiente. Al solo conocer *a posteriori* las emisiones, la única opción es tomar medidas correctivas, las cuales además tienen una latencia de al menos 1 año. Se cree que con un mejor registro se podrían tomar medidas de naturaleza preventiva, identificando quemas de forma rápida y avisando a la Central de Coordinación Regional de CONAF, buscando extinguir los incendios lo antes posible, evitando exposición frecuente a gases y material particulado los cuales pueden generar consecuencias irreversibles en las comunidades viento abajo [3]. Además, los incendios en zonas abiertas no urbanas corresponden a, aproximadamente, un 20% de las fuentes de emisiones de gases de efecto invernadero [4], lo que implica que es relevante mantener un inventario de los gases y partículas liberadas producidas por ellos.

Por otra parte, actualmente existe una gran cantidad de plataformas que permiten descargar imágenes satelitales. Estas imágenes pueden tener una resolución temporal alta, incluso obteniendo muestras diarias de una zona. Es más, las imágenes satelitales pueden utilizarse con distintos fines, dado que los satélites actuales cuentan con cámaras multi-espectrales. Existen productos e imágenes satelitales que se pueden utilizar para reconocer incendios e incluso medir sus efectos, como estimar el área total que afectan.

En el Trabajo de Memoria descrito en el presente documento se propone la generación de un inventario de emisiones de incendios forestales y agrícolas, con el fin de mantener un registro de emisiones con mayor resolución tanto temporal como espacial, en distintas localidades de Chile. El inventario tiene como fin ser un registro cercano al tiempo real o *Near Real Time* (NRT), de forma de que cuando se le entregue una imagen satelital al estimador de emisiones,

éste pueda entregar la cantidad de gas y partículas liberadas en ella de forma casi instantánea.

Los inventarios de emisiones de gases y partículas corresponden a un recuento de la cantidad de contaminantes liberados a la atmósfera, identificando el período temporal y el lugar, con referencia geográfica, donde ocurren las emisiones.

El objetivo general del trabajo presentado es: generar un inventario de emisiones de Chile con alta resolución temporal y espacial, a partir de imágenes satelitales, factores de emisiones y cartografía de vegetación. Se espera que utilizando un modelo de Machine Learning, al cual se le ingresan los datos mencionados, se pueda estimar la emisión de gases y material particulado producidos por incendios forestales y agrícolas.

En cuanto a los objetivos específicos del proyecto, se tienen los siguientes:

1. Determinar una técnica utilizada actualmente para generar inventarios de emisiones, con mayor resolución temporal a la del Registro de Emisiones y Transferencia de Contaminantes del Ministerio del Medio Ambiente y la CONAF, la cual se definirá como 'estándar dorado' o *ground truth*, es decir, la verdad a la que se adhiere el método de estimación de emisiones.
2. Generar una base de datos que contenga imágenes satelitales de Chile y datos de clasificación de terreno, para así poder representar las emisiones en el territorio del país.
3. Generar un modelo de *benchmark* utilizando redes neuronales convolucionales pre-entrenadas.
4. Perfeccionar el modelo de *benchmark* ajustando sus parámetros al problema presentado.
5. Diseñar e implementar una interfaz de usuario preliminar, donde se puedan visualizar los datos históricos de emisiones y se puedan ingresar nuevas imágenes, para así obtener la estimación de de la cantidad de gases emitidos entregado por el modelo ajustado.

En el próximo capítulo se explican los términos necesarios para comprender el desarrollo del trabajo descrito en el presente documento. Además, se presentan las técnicas utilizadas actualmente para obtener emisiones a partir de productos satelitales. Una vez que se comprenden los términos base, en el capítulo 3, se presenta la adquisición de datos y lo que ellos representan. Luego, en el capítulo 4, se presenta la metodología empleada para cumplir con los objetivos mencionados anteriormente. Una vez explicado el proceso del trabajo desarrollado, se muestran los resultados obtenidos en el capítulo 5. A partir de los resultados, se genera el capítulo 6 de discusión para analizar el desempeño de la metodología. Finalmente, en el capítulo 7 de conclusiones, se discute sobre el trabajo realizado, mostrando sus principales aportes y las posibles perspectivas futuras del sistema.

# Capítulo 2

## Marco Teórico y Estado del Arte

En esta sección se describen términos necesarios para comprender la metodología a emplear. Además, se mencionan trabajos anteriores relacionados a la generación de inventarios de emisiones en el mundo.

### 2.1. Marco Teórico

#### 2.1.1. Machine Learning

Machine Learning (ML) corresponde a una rama de la Inteligencia Artificial en donde se 'entrenan' algoritmos para que aprendan el comportamiento de un sistema a través de los datos históricos de éste, sin necesidad de conocer las ecuaciones que lo modelan. Los modelos de ML son entrenados para encontrar patrones en grandes cantidades de datos con el fin de generar predicciones de la salida a partir de nuevos datos ingresados.

Actualmente, existe una gran variedad de algoritmos de Machine Learning, como árboles de decisión, redes neuronales, *gradient boosting machines*, entre otros, los cuales se usan en una gran cantidad de aplicaciones. Por ejemplo, los sistemas de recomendación y de procesamiento de lenguaje se usan diariamente de forma global. Los sistemas de recomendación entregan productos que el usuario probablemente quiere según sus compras anteriores, en donde el modelo de ML es el encargado de predecir este conjunto de productos. Por otra parte, los sistemas de procesamiento de lenguaje están en múltiples aplicaciones, pero una de las más conocidas es el de Gmail, en donde existe un modelo que sugiere las palabras siguientes de la frase según las palabras escritas por el usuario. En particular, para el procesamiento de imágenes, se utilizan redes neuronales convolucionales.

##### 2.1.1.1. Redes Neuronales Convolucionales

Como se menciona anteriormente, las redes neuronales convolucionales (CNN) [5] corresponden a algoritmos de ML utilizados principalmente para visión computacional, abarcando tareas como clasificación y segmentación de imágenes. Estos algoritmos son capaces de clasificar o estimar un valor a partir de la entrada, la cual corresponde a una imagen.

Las redes neuronales, de forma general (no solo las CNN), se componen de capas. Las capas se acoplan entre sí para formar la estructura final que realiza la estimación o predicción, es

decir, la salida de una capa es la entrada de otra (u otras), excepto en el caso de las capa de salida. En específico, para las CNN existen principalmente tres capas; éstas son: las capas convolucionales, las capas de *pooling* y las capas densas o *fully connected*.

## 1. Capa Convolutiva

Generalmente, la entrada o *input* que se entrega a una CNN es una imagen en escala de grises o a color. Esto quiere decir que el *input* tiene un 3 índices de tamaño; éstos son: el alto, el ancho y la cantidad de canales de la imagen. Cabe mencionar que las imágenes a color que usualmente se utilizan tienen 3 canales: rojo, verde y azul, mientras que las imágenes en escala de grises solo tienen 1 canal.

Las capas convolucionales son la unidad base de las CNN, dado que se llevan la mayor parte del cómputo del modelo. Esta capa realiza una convolución, como la que se muestra en la Figura 2.1, entre dos matrices: el filtro (*kernel*) y la imagen ingresada a la capa. El *kernel* se 'desliza' por el *input* con un cierto paso o *stride*, generando una representación de ésta, generalmente llamada *activation map*. Para cada uno de los canales se realiza la convolución con el o los filtros que forman la capa convolutiva, donde se apilan los *activation maps* obtenidos, generando la salida de esta capa llamada *feature map*.

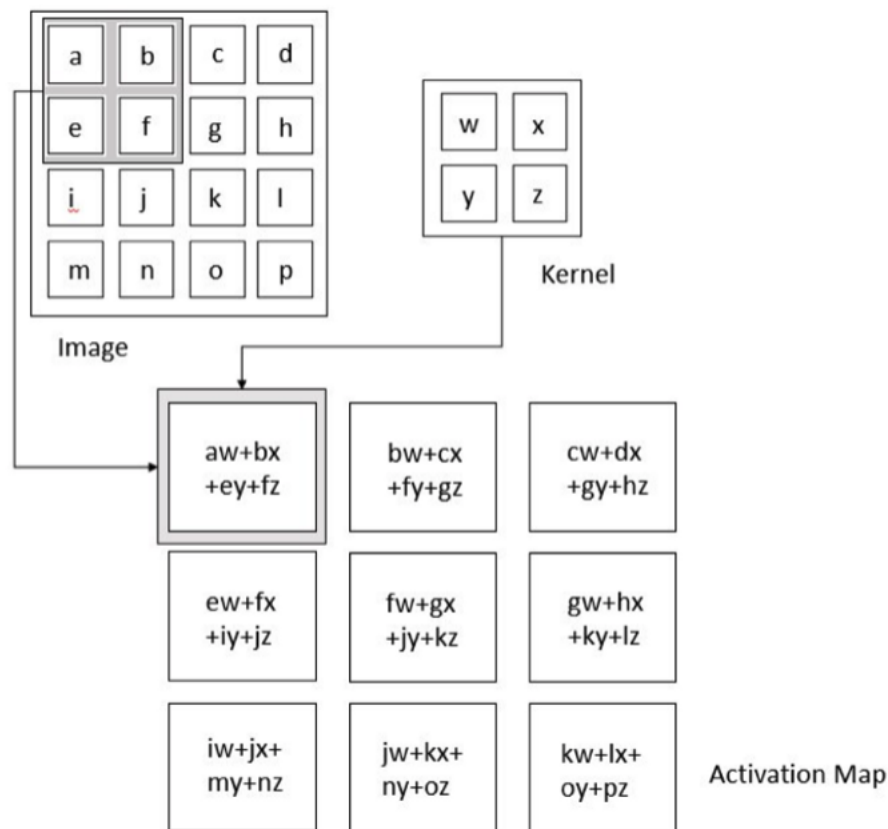


Figura 2.1: Ejemplo convolución (fuente: [6])

Las capas convolucionales logran reconocer distintas formas en las imágenes, mientras más cercanas estén a la capa de entrada las figuras identificadas tienen a ser más simples, como líneas y puntos. Al ir avanzando en la red, las formas que se perciben tienden a

ser más complejas. Es por esto que se utilizan para identificar elementos en las imágenes y para segmentar figuras en ellas.

Al final de cada capa convolucional al *feature map* generado, se le agrega no linealidad a través de una función de activación. Esta función, como se mencionó, busca añadir no linealidad al problema, evitando la suposición de linealidad en el sistema. Algunas de las funciones más utilizadas para esto son las funciones ReLU, sigmoide, GELU, entre otras.

Durante el entrenamiento de redes neuronales convolucionales, se van definiendo, a través de *backpropagation*, los valores de los distintos filtros que definen las capas convolucionales, por lo que dependiendo de la salida esperada las matrices *kernel* adquieren los valores que las componen.

## 2. Capas de *Pooling*

Estas capas se utilizan principalmente para disminuir el tamaño espacial de los *feature maps*. El fin de agregar las *pooling layers* es reducir la cantidad de cómputo requerido por la red neuronal. Existen distintos tipos de funciones de *pooling* utilizadas por estas capas, en particular, una función de interés es *Max-Pooling*. Una capa de *Max-Pooling* reduce la dimensionalidad seleccionando el valor máximo de una sección del *activation map* que recibe como entrada. Por ejemplo, en la Figura 2.2 se muestra una función de *Max-Pooling* donde se utiliza un filtro de  $2 \times 2$  y un paso de 2. Esto quiere decir que se usa un filtro de  $2 \times 2$  y se 'desliza' de 2 en dos cuadrantes, el cual selecciona el máximo de los valores que observa.

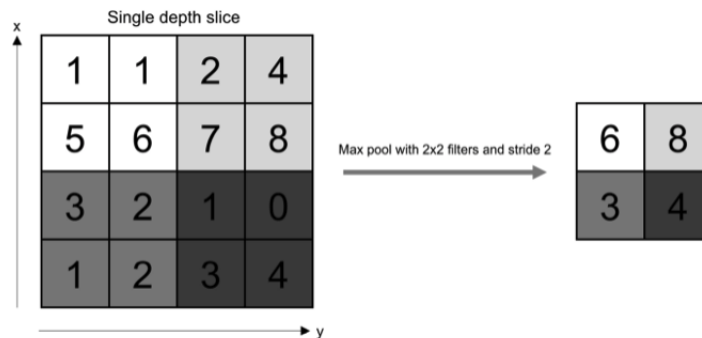


Figura 2.2: Ejemplo capa de *Max-Pooling* (fuente: [7])

Este proceso se realiza para cada uno de los canales del *activation map* obtenido como salida de la capa convolucional.

## 3. Fully Connected Layer

Las capas *fully connected* o *dense layers* corresponden a una estructura utilizada la mayoría de las redes neuronales. El objetivo de estas capas en las CNN es procesar la salida de las capas convolucionales de *pooling* y usarla para clasificar la imagen ingresada.

Las *fully connected layers* toman un vector como *input*, por lo que para pasar a estas capas en una CNN se debe 'aplanar' la salida de las capas anteriores, en el caso de que

correspondan a *feature maps*. Estas capas están compuestas por un conjunto de unidades o neuronas y su salida corresponde a la salida de cada una de ellas, como se aprecia en la Figura 2.3 (a). Cada neurona realiza una suma ponderada de las entradas, la cual luego pasa por una función de activación. Los valores por los que se pondera la entrada corresponden a los pesos o *weights*, los cuales se definen con el entrenamiento de la red. Por otra parte, esta suma ponderada tiene un valor base llamado *bias*, el cual también se define a través del entrenamiento. Finalmente, la función de activación se aplica sobre la suma ponderada, generalmente con el fin de introducir no-linealidad, por lo que algunas de las más usadas son ReLU, sigmoide, tangente hiperbólica, entre otros. La estructura base de una neurona se puede apreciar en la Figura 2.3 (b).

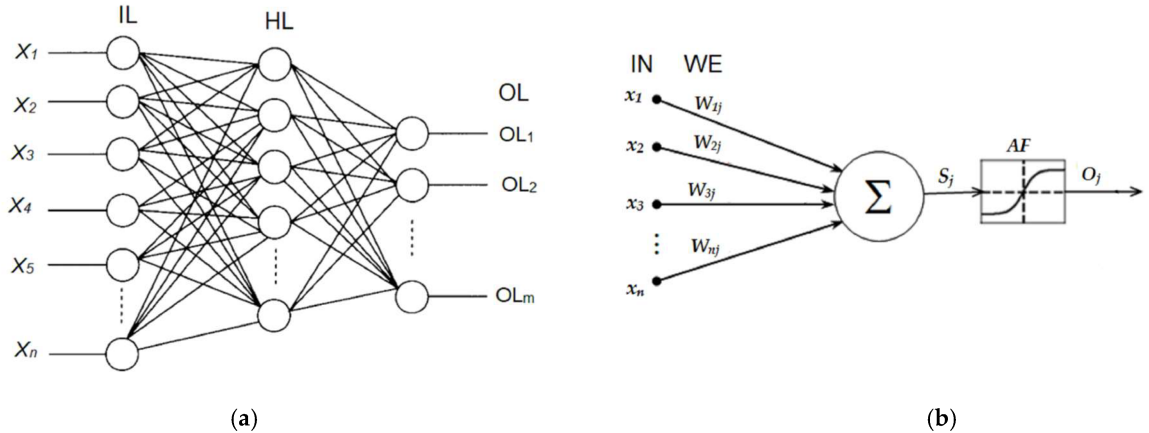


Figura 2.3: Representación unidad neuronal y *fully connected layer* (fuente: [8])

Utilizando las capas mencionadas anteriormente es que se construyen las redes neuronales convolucionales. Como se puede notar en la Figura 2.4, las capas convolucionales utilizadas pueden tener distinto tamaño de filtro, pero además pueden presentar diferente cantidad de canales. A su vez, en caso de clasificación, la *fully connected layer* de salida debe tener la misma cantidad de neuronas como clases posibles.

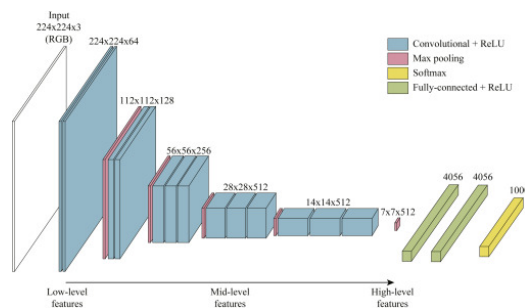


Figura 2.4: Diagrama de la red neuronal convolucional VGG16 (fuente: [9])

### 2.1.1.2. Redes Neuronales Residuales

Las redes neuronales profundas o *deep neural networks* corresponden a arquitecturas de Machine Learning que contienen una mayor cantidad de capas que los modelos tradicionales.



La intuición detrás de las *deep neural networks* se acredita a que las capas adicionales son capaces de aprender atributos más complejos de las imágenes. Sin embargo, existen estudios empíricos [10] que exponen la existencia de un umbral máximo para la profundidad con los modelos tradicionales de CNN, como los presentados anteriormente. Este máximo tiende a acreditarse a la pérdida de información al ir avanzando en las capas de la red neuronal.

Con el fin de evitar el fenómeno descrito anteriormente, se introducen las redes neuronales residuales o ResNet. El elemento básico de las ResNets es el bloque residual el cual contiene *skip connections*. Este tipo de conexiones, como se puede apreciar en la Figura 2.5, corresponde a la suma de la salida y la entrada de dos capas no consecutivas de la red neuronal. El resultado de la aplicación de *skip connections* es la habilidad de entrenar redes neuronales más profundas.

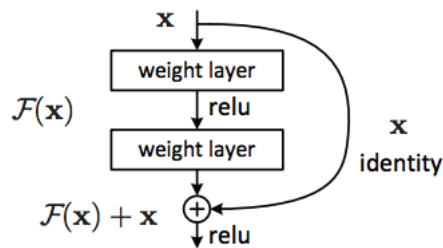


Figura 2.5: Diagrama de una *skip connection* (fuente: [11])

### 2.1.2. Red Neuronal UNet

La segmentación de imágenes o *image segmentation* en visión computacional corresponde al proceso de particionar la imagen en segmentos, es decir, en conjuntos de píxeles. La segmentación se realiza, generalmente, con el fin de localizar objetos y sus bordes en la imagen, asignándole una etiqueta a cada píxel, donde los píxeles con la misma etiqueta comparten ciertas características.

Las red neuronal artificial UNet [12] corresponde a una CNN utilizada principalmente para segmentación de imágenes, entregando una máscara en donde cada píxel obtiene una clasificación o estimación. Esta arquitectura tiene forma de 'U' como se aprecia en la Figura 2.6, de donde obtiene su nombre, compuesta por dos secciones; éstas son: compresión (*image compression*) y expansión (*image expansion*).

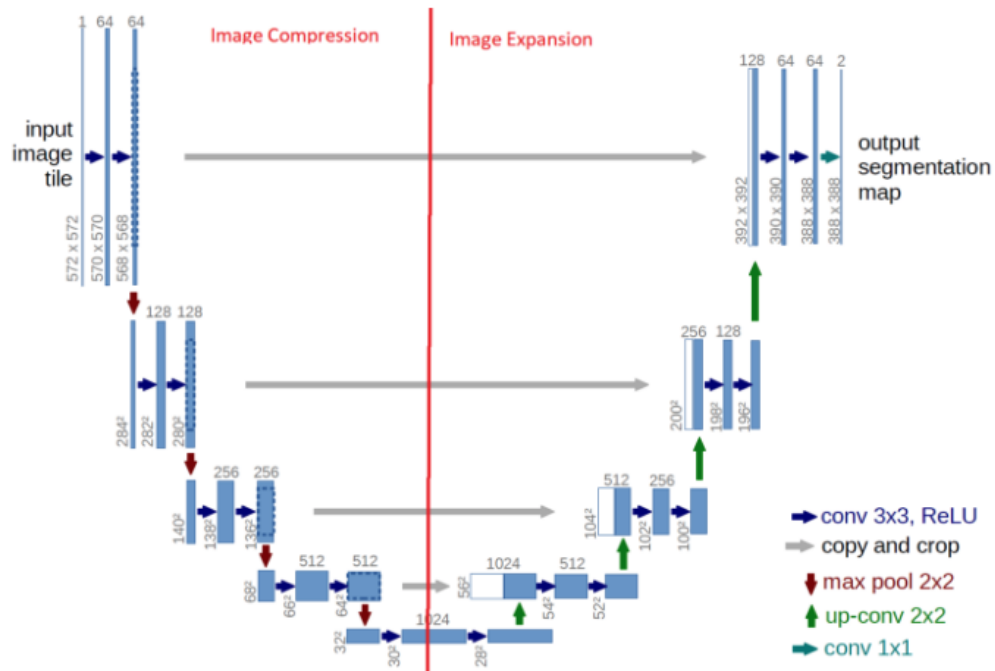


Figura 2.6: Diagrama red UNet (fuente: [12])

La sección de compresión (parte izquierda de la Figura 2.6) sigue una arquitectura típica de una CNN, con una aplicación reiterada de convoluciones con *kernels* de  $3 \times 3$ , cada una seguida de una función de activación ReLU y una capa de *maxpooling* de  $2 \times 2$  con un paso de 2, con el fin de generar submuestreo (*downsampling*). En cada paso de *downsampling* se duplica la cantidad de canales. Por otra parte, en cada etapa de la sección expansiva se realiza *upsampling*, es decir, se incrementa el tamaño en alto y ancho de la imagen. Luego, cada *feature map* pasa por una convolución de  $2 \times 2$ , que reduce a la mitad su cantidad de canales, seguido por una *skip connection* con el *feature map* recortado correspondientemente de la sección de compresión. Finalmente, se tienen dos convoluciones de  $3 \times 3$ , cada una seguida por un ReLU. La salida de este algoritmo corresponde a una imagen del mismo tamaño que la imagen de entrada, en donde se puede visualizar la segmentación, como la que se puede apreciar a la derecha de la Figura 2.7



Figura 2.7: Ejemplo resultado segmentación de imágenes (fuente: [13])

### 2.1.3. *Wildfires* y emisiones

La principal causa del cambio climático introducido por el humano es la liberación de gases de efecto invernadero, GHG por sus siglas en inglés (*Greenhouse Gases*), a la atmósfera. Una fuente importante de dióxido de carbono ( $\text{CO}_2$ ) es producido por los incendios o *wildfires*. *Wildfire* está comúnmente definido como un incendio de quema rápida y de forma incontrolable en áreas de pasto, arbustos o bosque [14].

Se definen como gases de efecto invernadero aquellos gases atmosféricos que retienen la radiación térmica [14], por lo que a ellos se atribuye el aumento en la temperatura promedio global o efecto invernadero. Las principales consecuencias del aumento de esta temperatura son:

- Deshielo de glaciares.
- Aumento en el nivel del mar, generando inundaciones en islas y ciudades costeras.
- Degradación de suelo de zonas fértiles.

Por otra parte, las personas que viven cercanas a localidades donde hay incendios de forma relativamente frecuente, tienen una peor calidad de aire, aumentando las probabilidades de desarrollar enfermedades respiratorias no transmisibles.

## 2.2. Estado del Arte

Como se menciona anteriormente, la emisión de gases y material particulado en incendios afecta tanto de forma global como local, por lo que es importante mantener un registro de ésta. Actualmente, para obtener las emisiones a partir de productos o imágenes satelitales, se presentan 2 principales acercamientos; éstos son: cálculo de emisiones a través del poder radiativo (*radiative power*) y cálculo de emisiones a través del área quemada.

### 2.2.1. *Fire Radiative Power*

En todos los ecosistemas terrestres vegetados ocurre quema de biomasa, las cuales contribuyen a la acumulación de carbono en la atmósfera. Esta metodología se basa en observaciones de la radiación termal de quemaduras de biomasa, dado que existen satélites que miden el *Fire Radiative Power* (FRP) o poder radiativo de los incendios, además de una alarma binaria de incendio. A partir del FRP se pueden derivar las tasas de combustión y de emisión de aerosoles y gases.

El estudio realizado en [15] muestra la estimación de emisiones a partir del FRP estimado por los satélites de MODIS (*Moderate Resolution Imaging Spectroradiometer*), Aqua y Terra. Es importante notar que esta metodología no logra medir a través de las nubes, por lo que en caso de que exista nubosidad este método no es capaz de extraer observaciones. Además, en este estudio se descartan observaciones sobre hielo, nieve y cuerpos de agua. Por otra parte, otra limitación de este método es que la frecuencia de muestreo se ajusta a la sincronía del sol con los satélites.

Como primer paso de este método es medir el FRP ( $F_i$ ), el área ( $A_i$ ) del píxel y el ángulo cénit ( $\Theta_i$ : ángulo entre los rayos del sol y la dirección vertical) para cada píxel  $i$  de la imagen satelital, donde las observaciones válidas de fuego son aquellas cuyo FRP es mayor a cero y las válidas de no-fuego, aquellas con FRP igual a cero. A partir de las mediciones, para cada celda  $j$  de la grilla global, se calcula su FRP ( $\langle F \rangle_j$ ) y su área ( $\langle A \rangle_j$ ). Usando los parámetros calculados anteriormente, se puede extraer la densidad de FRP por área ( $e_j$ ) y la fracción de área observada ( $\gamma_j$ ), sin nubosidad, para cada celda.

Es importante notar que existen secciones de la grilla global que están representadas en más de un producto satelital, por lo que las observaciones se deben unir o mezclar para obtener una representación más general de la grilla. Para cada celda, usando todos los productos satelitales (con índice  $k$ ) se obtienen los valores generales de la densidad de área de FRP ( $\tilde{e}_j$ ), la desviación estándar de las observaciones de la celda ( $\tilde{\zeta}_j$ ) y la fracción de área observada en cada celda ( $\tilde{\gamma}_j$ ). Por otra parte, existen celdas que contienen cuerpos de agua o que muestran un FRP mayor por presencia de emisiones infra-rojas, actividad empresarial o actividad volcánica. Para corregir esto se usan fracciones de cobertura terrestre y una máscara que identifica FPR 'falsos' (no generados por incendios con quema de biomasa). Con estos indicadores, la densidad de FRP general y la desviación estándar se pueden mejorar, generando medidas de densidad ( $\tilde{\rho}_j$ ) y desviación estándar ( $\tilde{\omega}_j$ ).

Por otra parte, dado que existen vacíos en las observaciones, en su mayoría producidos por nubosidad, con información adicional se genera una estimación discreta ( $\hat{\rho}_t$ ) de la distribución continua  $\tilde{\rho}$  utilizando el filtro de Kalman, el cual es utilizado para estimar el estado no medible de un sistema dinámico no lineal. Este estimador modela la densidad  $\rho_t$  en el período  $t$  asumiendo persistencia del momento anterior  $t - 1$ . A partir de las estimaciones se puede extraer la tasa de combustión de materia seca para cada celda de la grilla mundial. La tasa de combustión depende del tipo de terreno de cada celda, en este estudio se consideran 8 tipos de suelos (Tabla 2.1) en los que puede haber quema de biomasa. Finalmente, con la tasa de combustión de la celda se pueden obtener las tasas de emisión en ellas, para 40 constituyentes de humo, utilizando una fórmula ponderada por un factor de emisión  $\kappa_s$ , para cada gas o partícula  $s$ .

Tabla 2.1: Tipos de uso de suelo utilizados en [15]

Uso de Suelo
Sabana
Sabana con suelo orgánico
Agricultura
Agricultura con suelo orgánica
Bosque tropical
Turba
Bosque extratropical
Bosque extratropical con suelo orgánico

## 2.2.2. Área Quemada

Otra forma de estimar las emisiones en incendios es a partir de áreas quemadas, tipos de vegetación y factores de emisión de distintos gases. En el estudio [16] se genera un registro diario, de 1 kilómetro (km) de resolución para quemas abiertas de biomasa o *wildfires*.

Para poder generar el registro de emisiones con este método se requiere del registro de área quemada, es decir, un producto satelital en donde se entregue los metros cuadrados ( $m^2$ ) quemados. Por otra parte, para estimar las emisiones se requiere saber el tipo de uso de suelo y la carga de biomasa y la fracción de biomasa quemada que se encuentra en el lugar de donde se extrae la información. En [16] estos datos de área quemada y tipo de uso de suelo, se extraen de los satélites de MODIS, los cuales alcanzan una resolución de 1 km, lo que significa que cada píxel de la imagen entregada corresponde a 1  $km^2$ .

Una vez que se tienen los datos de área quemada y tipo de uso de suelo, se pueden estimar las emisiones a partir de la Ec. 2.1, donde se obtienen las emisiones  $E_i$  en gramos (g) para el gas o partícula  $i$  en un tiempo  $t$  y lugar  $x$  a partir del área quemada  $A(x, t)$  en metros cuadrados, la carga de biomasa  $B(x)$  en kilogramos por metro cuadrado, que está dada por el tipo de uso de suelo, la fracción de biomasa que se quema  $FB$  (sin unidad de medida) y el factor de emisión  $ef_i$  en gramos por kilogramo de biomasa quemada. Es importante notar que los términos de biomasa utilizados son bajo la suposición de que la biomasa está seca.

$$E_i = A(x, t) \times B(x) \times FB \times ef_i \quad (2.1)$$

Ecuación para calcular emisiones a partir del área quemada

Para la obtención de datos como la carga de biomasa y el factor de emisión de cada gas o partícula se toma en cuenta tanto la zona de donde se toman los datos como el tipo de uso de suelo en ellas. Para las zonas se toman en cuenta 14 regiones globales, las cuales se pueden apreciar en la Figura 2.8, mientras que para el tipo de uso de suelo se consideran 5 clasificaciones: bosques tropicales, bosques templados, matorrales, pastizales y cultivos.

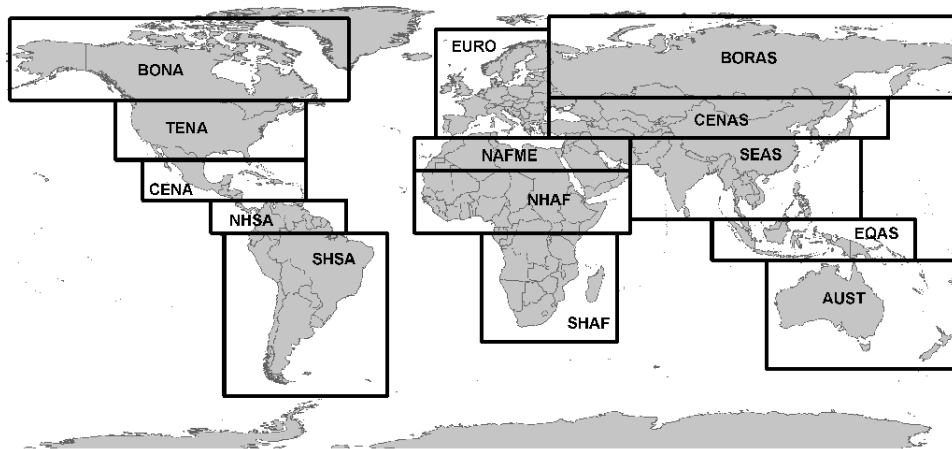


Figura 2.8: Zonas consideradas en [16]

# Capítulo 3

## Adquisición de Datos

Para el trabajo descrito en el presente documento, se requiere principalmente de 2 tipos de datos: imágenes satelitales de incendios o áreas quemadas y un mapa de uso de suelo de Chile. A continuación se presenta cómo se adquieren los datos necesarios para la generación de un inventario de emisiones en quemas forestales y agrícolas en Chile.

### 3.1. Imágenes Satelitales

Cómo se mencionó anteriormente, la estimación de las emisiones se deben extraer a partir de la combinación de imágenes satelitales con datos de vegetación, sin embargo, éstas no necesariamente corresponden a imágenes del tipo rojo, verde y azul que componen las imágenes digitales comunes. Los satélites actuales cuentan con diferentes bandas que representan señales en distintos rangos de longitud de onda, por lo que se pueden formar compuestos de las bandas para generar una variedad de imágenes. Antes de seleccionar las bandas a utilizar se debe elegir el satélite desde el cual se extraen las observaciones. Hoy en día existe una gran cantidad de satélites de los cuales se pueden extraer imágenes y productos, en particular, para el trabajo desarrollado se seleccionó el satélite Sentinel-2A. Actualmente, existen diferentes satélites de los cuales se pueden obtener imágenes, algunos de ellos son Sentinel-2A, Sentinel-2B, Aqua, Terra, Landsat-7, Landsat-8, entre otros. Se elige trabajar con Sentinel-2A dada su alta resolución espacial, alcanzando píxeles de  $100 \text{ m}^2$ , además de una mayor resolución temporal, teniendo muestras con resolución temporal entre 24 horas y 10 días de latencia.

En particular, para la visualización de incendios y cicatrices de quemas (*burn scars*) unas bandas que muestran resultados claros son las bandas *Shortwave Infrared 2* (SWIR 2), *Near Infrared* (NIR) y *Red* [17]. La señal SWIR 2 corresponde a la banda 12 (B12) de Sentinel y se encuentra entre los 2202.4 y los 2185.7 nanómetros (nm), por su parte NIR es la banda 8 (B8) y está en el rango entre los 835.1 nm y los 833 nm, finalmente, la señal *Red* representa el color rojo de las imágenes y corresponde a la banda 4 (B4) que se encuentra entre los 664.5 nm y los 665 nm.

Una vez que se seleccionan las bandas y el satélite a utilizar para obtener los datos, se deben comenzar a descargar. Para la descarga de los datos se utiliza el lenguaje de programación Python [18] y su API (*Application Programming Interface*) de Google Earth Engine [19], el cual facilita la descarga de todas las bandas de una gran variedad de productos satelitales y facilita su filtrado por fechas y geografía.

Para poder filtrar por localización las imágenes satelitales se extrae un Shapefile desde la Biblioteca del Congreso Nacional de Chile [20], que contiene las coordenadas que delimitan al país. Este tipo de archivo se lee con el paquete de Python Geopandas [21], el cual genera una geometría (multi-polígono de coordenadas) que delimitan una zona, en este caso, el territorio chileno.

Una vez que se tienen las coordenadas, se debe seleccionar un rango temporal dentro del cual se encuentran los datos. El satélite Sentinel-2A contiene muestras para Chile desde Diciembre del 2018, por lo tanto esta fecha se selecciona como límite inferior para la descarga de datos. Por otra parte, se toma como límite temporal superior el 1 de Julio del año 2021. De esta forma, al definir el rango temporal, se tienen los filtros suficientes para generar la colección de imágenes a descargar.

Es importante notar que las imágenes con una alta capa de nubosidad generar muestras ruidosas, es decir, son difíciles de interpretar para las redes neuronales. Por ende, se decide filtrar las imágenes que tengan más de un 10% de nubes. Para esto se utiliza la banda de Sentinel que contiene la probabilidad de que un píxel corresponda a una nube, generando un umbral en el 50% de probabilidad. De esta forma, si más de un 10% de los píxeles de una imagen tienen 50% o más de probabilidad de ser nubosidad, ésta se descarta y no es descargada.

Al haber filtrado por geografía, bandas, fechas y nubosidad se obtiene una base de datos con 11.084 imágenes del territorio chileno de, aproximadamente,  $10 \text{ km}^2$  (varía según la imagen). Cada píxel de las imágenes descargadas cubre  $100 \text{ m}^2$ , por lo que se logra observar incendios y *burn scars* pequeños. Esto es importante, dado que al tener una resolución espacial tan alta se visualizan con mayor claridad las quemadas, minimizando la probabilidad de pasar por alto quemadas de menos de  $250.000 \text{ m}^2$ .



Figura 3.1: Imagen satelital de las bandas SWIR, NIR y Red, extraída de Sentinel-2A con 10 m de resolución

La Figura 3.1 es un ejemplo de las imágenes descargadas para la generación del inventario de emisiones de incendios, en ella se pueden apreciar claramente dos *burn scars* de color rojizo, las cuales indican que hubo un incendio donde están marcadas en el pasado cercano. De esta forma las imágenes descargadas presentan incendios o su representación en *burn scars* con un color café rojizo. Sin embargo, en el momento de almacenamiento de las imágenes, como éstas se guardan en formato PNG (*Portable Network Graphics*), el canal rojo (SWIR 2) con el canal azul (Red) se intercambian, por lo que las imágenes almacenadas se visualizan en tonalidades azules, como se aprecia en la Figura 3.2



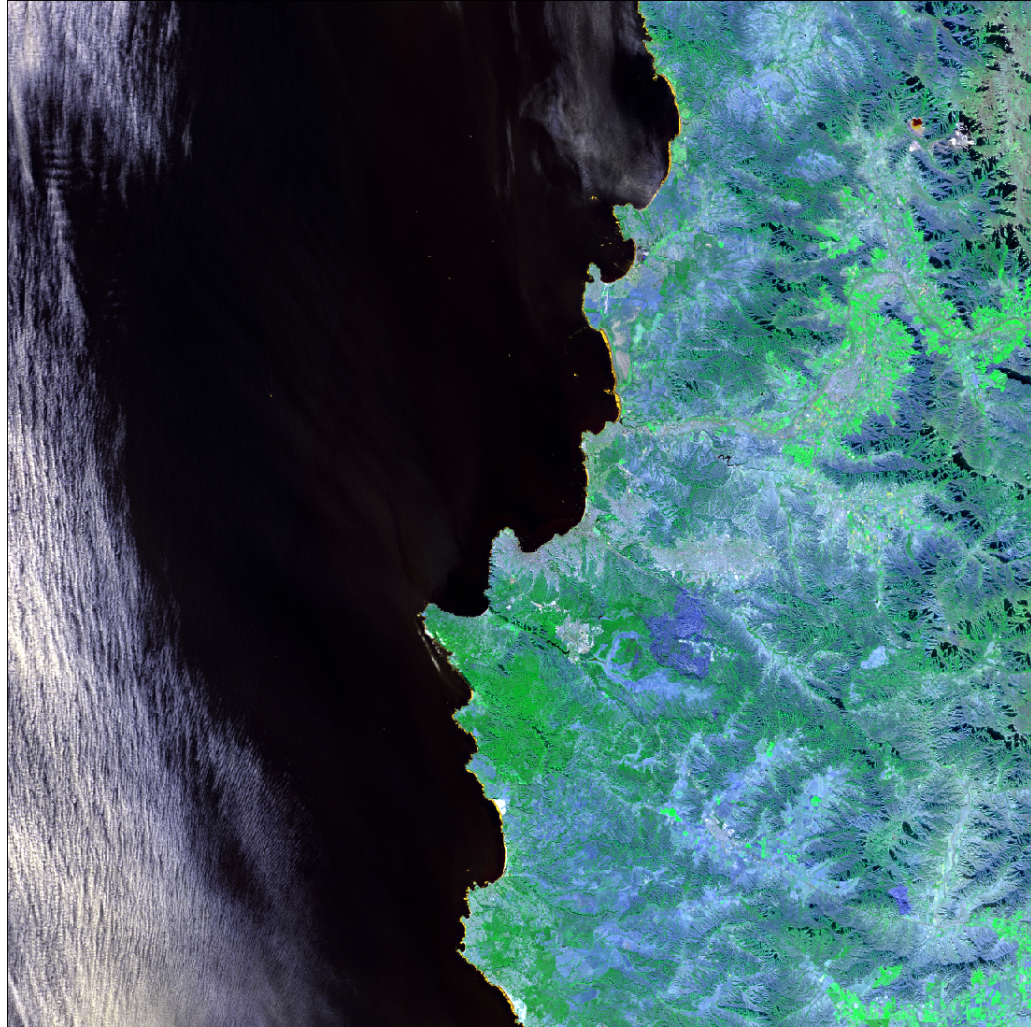


Figura 3.2: Imagen satelital de las bandas SWIR 2, NIR y Red, almacenada en formato PNG

Las imágenes descargadas tienen referencia temporal y geográfica en su nombre. Cada imagen es almacenada con un nombre en el siguiente formato:  $fecha\_longitud_{max}longitud_{min}\_latitud_{max}latitud_{min}$ , donde  $longitud_{max}$  y  $longitud_{min}$  corresponden a los valores máximos y mínimos de longitud respectivamente. Análogamente,  $latitud_{max}$  y  $latitud_{min}$  son los valores máximos y mínimos de latitud. Por su parte, el valor de la fecha corresponde a 8 dígitos, los cuales siguen el formato YYYYMMDD (año, mes y día).

## 3.2. Datos de Vegetación

Para obtener los tipos de vegetación se utiliza el mapa de uso de suelo generado en el estudio de Zhao et al. [22], el cual tiene una resolución de 30 metros y cubre el territorio chileno. El mapa de uso de suelo corresponde a un archivo GeoTIFF donde cada celda representa una referencia geográfica de  $30\text{ m}^2$  en conjunto con su tipo de uso de suelo. En la Figura 3.3 se puede apreciar la visualización del mapa de uso de suelo, en donde negro representa el valor 0 y el color más claro corresponde al valor 1210 (ver Tabla 3.1).

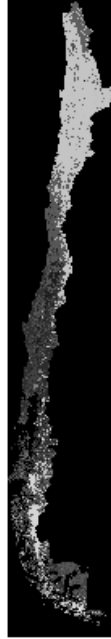


Figura 3.3: Mapa de uso de suelo obtenido de [22]

Este archivo contiene la clasificación de uso de suelo entregada en [22], pero además, como metadata tiene las coordenadas geográficas en el sistema UTM (*Universal Transverse Mercator*). Este es un sistema diferente al sistema de coordenadas en grados que se utiliza para obtener las imágenes. Las coordenadas en grados, también conocidos comúnmente como latitud y longitud, es mundial, es decir, no hay dos puntos diferentes con las mismas coordenadas. Por su parte el sistema UTM cuenta con zonas, y dentro de ellas las coordenadas son únicas.

Es importante notar que el mapa visualizado en la Figura 3.3 contiene una clasificación codificada. En la Tabla 3.1 se muestran los distintos códigos con sus respectivas clases de uso de suelo. Sin embargo, es importante mencionar que el archivo de clasificación de terreno tratado cuenta con 3 niveles, donde el primer nivel es la clasificación más general y es la utilizada en el desarrollo del método descrito. No se consideran las clasificaciones en los niveles 2 y 3, dado que estos son más específicos de lo que se requiere, y subdividen las clases mostradas en la Tabla 3.1.

Es importante notar que en el *dataset* está la clasificación 0, la cual corresponden a puntos fuera del territorio chileno, o puntos en el mapa que no tienen una clasificación definida. Por otra parte, existen clasificaciones entregadas en el archivo GeoTIFF de [22], las cuales no forman parte de terrenos forestales o agrícolas; éstas son: humedales, cuerpos de agua, superficies impermeables, tierras desnudas, hielo, nieve y nubes. Para el trabajo desarrollado en el presente informe, se consideran solo como incendios forestales y agrícolas aquellos que ocurren en terrenos clasificados como cultivos, bosque templado, bosque tropical, pastizales y matorrales.

Tabla 3.1: Códigos extraídos de [22] y clasificación de uso de suelo

Códigos en Land Cover	Clasificación para cálculo de emisiones
0	NA
110, 120, 130, 140, 150	Cultivos
211, 212, , 221, 222, 241, 251	Bosque Templado
311, 312, 320, 330	Pastizales
410, 420, 430, 440, 450	Matorrales
510, 520, 530	Humedales
610, 620, 630, 640	Cuerpos de Agua
800	Superficies Impermeables
910, 920, 931, 932	Tierras Desnudas
1010, 1020	Hielo y Nieves
1210	Nubes

# Capítulo 4

## Metodología

En este capítulo se describe los procedimientos realizados para la confección del inventario de emisiones de incendios forestales y agrícolas. A grandes rasgos lo que se hace es estimar el área quemada a partir de las imágenes satelitales descargadas utilizando un modelo de Machine Learning. Luego, utilizando la información de área quemada estimada en conjunto con los datos de tipo de uso de suelo, se calculan las emisiones de 15 gases y partículas utilizando la Ec. 2.1, presentada en el Marco Teórico.

Una vez que se obtiene el método de estimación de emisiones, éste se emplea para generar un registro de emisiones de Chile con los datos recolectados. Finalmente, utilizando tanto el método como el registro generado, se formula una interfaz de usuario preliminar, en donde se pueden obtener estimaciones para una nueva entrada y se pueden visualizar las emisiones registradas.

### 4.1. Estimación de Área Quemada

#### 4.1.1. Pre-Procesamiento

Como se menciona previamente, para estimar el área quemada de una imagen se utiliza un modelo de Machine Learning, pero para que la imagen sea admitida por la arquitectura, debe ocurrir un pre-procesamiento de la imagen. El pre-procesamiento de las imágenes descargadas es el siguiente:

1. Primero, se genera un relleno (*pad*) de la imagen. Se añaden  $h_i - 480$  píxeles de alto y  $w_i - 720$  píxeles de ancho a la imagen, donde  $h_i$  y  $w_i$  corresponden a la altura y el ancho de la imagen, respectivamente. Los píxeles agregados como relleno corresponden al color negro, es decir, tienen valor 0. Este paso se realiza para que la imagen luego pueda ser dividida en secciones de  $480 \times 720$ , dado que este es el tamaño en que la red neuronal recibe las entradas.
2. Una vez que la imagen tiene el *pad* correspondiente, esta se segmenta en secciones de  $480 \times 720$  píxeles de alto y ancho, respectivamente. Es importante notar que en este paso se guarda el tamaño original de la imagen, el cual luego se utilizará para reconstruir la salida con las mismas dimensiones de la entrada.

De esta forma, cada imagen es descompuesta en secciones de el tamaño adecuado con el cual pueden ser ingresadas y procesadas por el modelo de Machine Learning a utilizar.

### 4.1.2. Estimación del Modelo de Machine Learning

La arquitectura utilizada para generar la segmentación de área quemada corresponde a una red neuronal convolucional llamada UNet, la cual, como se describe en el el Capítulo 2 comprime la imagen a *feature maps*, los cuales en conjunto de *skip connections*, luego se expanden para obtener una máscara que representa las estimaciones de probabilidad en la segmentación.

En particular se utiliza una UNet pre-entrenada, extraída del trabajo realizado en [23], en donde se entrena la red con imágenes descargadas de la API de Google y marcadas manualmente por los autores del trabajo. La red se especializa en la segmentación de *burn scars* y *wildfires*, dado que fue entrenado con 110 imágenes que contienen esas características. Para cada una de las imágenes de tamaño  $480 \times 720$  generadas en el pre-procesamiento, el modelo genera una imagen análoga que contiene la probabilidad de cada píxel de ser clasificado como *burn scar* o *incendio*. Todas las secciones de la imagen original obtienen su máscara de probabilidad correspondiente, y se unen para conformar la máscara análoga a la imagen de entrada.

La salida final del sistema corresponde a una imagen en donde cada píxel representa la probabilidad de que éste forme parte de un *wildfire* o *burn scar*. En la Figura 4.1 se puede apreciar que el modelo marca una mayor probabilidad (colores más claros) los píxeles con colores rojizos (máscara generada para la imagen en la Figura 3.1). Es importante tomar en cuenta que esta imagen presenta condiciones poco ruidosas, es decir, hay poca nubosidad, no se visualiza nieve ni lugares sin vegetación, por lo que la segmentación logra buenos resultados.

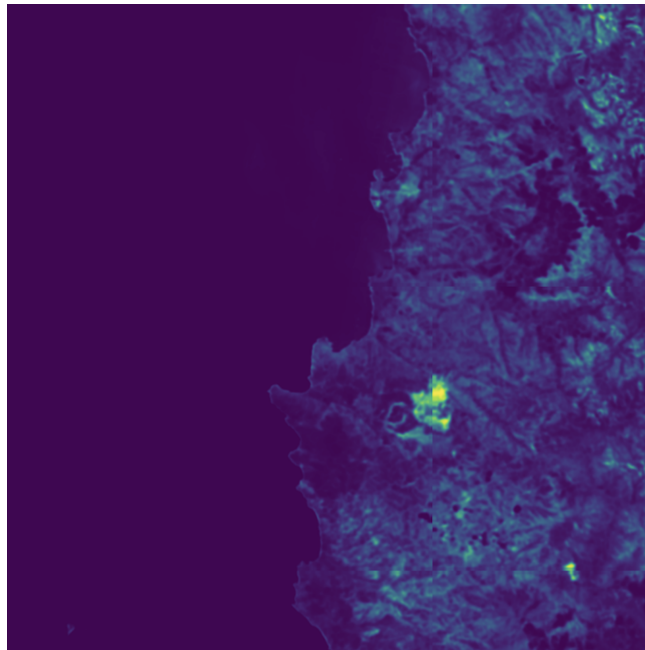


Figura 4.1: *Output* de la red UNet tomando como entrada la Figura 3.1

Sin embargo, se debe considerar que no todas las imágenes tienen condiciones tan favorables como la de la Figura 3.1, dado que se pueden presentar lugares con menos vegetación, con montañas, nieve y con una mayor presencia de nubosidad. Es por esto que para seleccio-

nar qué píxeles efectivamente corresponden a área quemada se decide utilizar umbrales. Los umbrales utilizados son ajustados de forma experimental, es decir, se cambian visualizando los resultados obtenidos para distintas condiciones de las imágenes satelitales. A partir de los experimentos se seleccionan 2 umbrales a utilizar. Con estos umbrales se filtra la salida del modelo, es decir, si los píxeles tienen un valor mayor al umbral se consideran área quemada, si no corresponden a área no quemada.

El primer umbral utilizado es el umbral de límite inferior que tiene valor 0,21 , el cual se selecciona con el fin de identificar incendios pequeños. Esto es importante dada la alta resolución de las imágenes, por lo que se pueden identificar quemadas que con otros instrumentos pueden ser ignoradas. Este umbral se obtiene, como se menciona anteriormente, de forma experimental con distintos escenarios. Para encontrarlo se prueba con distintos escenarios (zonas secas, montañas, zonas de alta vegetación, entre otras) buscando el umbral máximo con el que se identifican incendios y *burn scars* pequeños, es decir, que cubren menos del 0,1% de la imagen total.

El segundo umbral es el de límite superior de valor 0,39 , el cual se crea para eliminar altos errores de falsa clasificación de área quemada. Este umbral se obtiene con los mismo escenarios que el umbral de límite inferior, pero ahora, lo que se busca es encontrar el valor mínimo en donde se identifiquen incendios mayores (más del 0,1% de la imagen) pero en donde no se identifiquen zonas que generan falsas identificaciones de área quemada, las cuales se presentan principalmente por zonas mayoritariamente secas, territorios con una mezcla de vegetación y zonas urbanas y por nubosidad.

De esta forma, se generan dos imágenes a partir de los umbrales, una con el umbral de límite inferior y otra con el umbral de límite superior. Ambos resultados de los umbrales corresponden a imágenes con valores binarios, 1 y 0, donde el área quemada se representa con el valor 1 y a la no quemada se le asigna el valor 0. A partir de las dos imágenes generadas con los umbrales se extrae la máscara final que contiene los píxeles que se consideran área quemada. Para esto se realizan los siguientes pasos:

1. Se obtiene el valor absoluto de la diferencia entre la imagen generada con el umbral de límite inferior y la generada con el umbral de límite superior.
2. Si esta diferencia es mayor al 0,1% del número total de píxeles entonces se entrega como área quemada la salida del umbral de límite superior.
3. En el caso de que la diferencia sea menor al 0,1% del número total de píxeles de la imagen original, se utiliza el resultado extraído con el umbral de límite inferior.

La intuición detrás de esto es que los incendios pequeños por lo general corresponden a menos de un 0,1% de la imagen total, lo que representa aproximadamente 120 km<sup>2</sup> (varía según la imagen), los cuales son identificados por el umbral de límite inferior pero no por el de límite superior. Por otra parte, los terrenos que introducen error son detectados por el umbral de límite inferior y por lo general representan más del 0,1% de la imagen, pero no son detectados por el umbral de límite superior. Mientras que los incendios mayores resaltan y son detectados por ambos umbrales. De esta forma uno evita tener una sobre detección de incendios, pero asegurando que los incendios pequeños sean visualizados y tomados en cuenta como área quemada.

## 4.2. Cálculo de Emisiones

Una vez que se tiene la imagen binaria que representa incendios y *burn scars*, generada por el modelo y el sistema de umbrales descritos anteriormente, se puede calcular el área quemada de la imagen original. Lo primero que se debe hacer es sumar la cantidad de píxeles con valor 1 en la salida del sistema de umbrales. Una vez que se conoce la cantidad de píxeles que se clasifican como quema, solo se requiere conocer la resolución de la imagen para calcular el área quemada. La fórmula para obtener el área quemada de una imagen  $j$  en metros cuadrados se puede observar en la Ec. 4.1, donde  $BP(j)$  corresponde a la suma los píxeles clasificados como quemados o *burn scars* y  $R(j)$  es la resolución en metros de  $j$ . Es importante notar que en las imágenes satelitales la resolución corresponde a la medida de largo y ancho de cada píxel, por ejemplo, en el caso de Sentinel-2A la resolución es de 10 metros.

$$A(j) = BP(j) \times R(j)^2 \quad (4.1)$$

Ecuación para calcular el área quemada de una imagen

Una vez que se estima el área quemada se pueden calcular las emisiones utilizando datos de vegetación del terreno de donde se extrae la imagen. Para esto se utiliza la Ec. 2.1, la cual se propone en el estudio [16], para calcular las emisiones en gramos de los gases y partículas de interés. Esta ecuación entrega las emisiones de cierto gas o partícula  $i$  en gramos a partir de el área quemada, la carga de biomasa, la fracción de área quemada y el coeficiente de emisiones del gas correspondiente.

Para la carga de biomasa se utilizan los valores propuestos en [16] para Sudamérica, en donde se presentan 5 tipos de uso de suelo con sus correspondientes factores  $B(x)$ :

- **Bosque Tropical:** Para este tipo de terreno se asume una carga de biomasa de 25.366  $\text{g}/\text{m}^2$ .
- **Bosque Templado:** En este tipo de bosques se utiliza una carga de biomasa de 3.497  $\text{g}/\text{m}^2$ .
- **Matorrales:** En este caso la carga de biomasa es de 2.501  $\text{g}/\text{m}^2$ .
- **Pastizales:** La carga de biomasa de este tipo de terreno es de 318  $\text{g}/\text{m}^2$ .

Es importante mencionar que adicionalmente para el trabajo descrito se agrega la categoría de **Cultivo**, correspondiente a terrenos agrícolas, la cual se supone que tiene igual carga de biomasa que los pastizales, de valor 318  $\text{g}/\text{m}^2$ , dado que según [24], gran parte de los cultivos se generan a partir de pastizales.

Por su parte, en este trabajo se se considera que todo el píxel se está quemando, es decir, que la fracción de área quemada ( $FB$  según la Ec. 2.1) es siempre 1. Esto se basa en la alta resolución de las imágenes descargadas, dado que cada píxel cubre 100  $\text{m}^2$ . De esta forma, se espera que si un píxel contiene o es parte de un incendio, el considerar que su totalidad se está quemando no genere grandes márgenes de error. Por otra parte, si se utilizaran bases de datos con menor resolución, como por ejemplo de 1  $\text{km}^2$ , esta suposición podría generar

un mayor error de estimación del área quemada.

Finalmente, para poder generar el cálculo de emisiones, se extraen los factores de emisión de 15 gases y partículas diferentes a partir de [16]. Los contaminantes considerados son: dióxido de carbono ( $\text{CO}_2$ ), monóxido de carbono ( $\text{CO}$ ), metano ( $\text{CH}_4$ ), dihidrógeno ( $\text{H}_2$ ), el grupo de gases reactivos que contienen nitrógeno y oxígeno ( $\text{NO}_x$ ), dióxido de nitrógeno ( $\text{NO}_2$ ), compuestos orgánicos volátiles no metánicos (NMOC), hidrocarburos sin metano (NMHC), dióxido de azufre ( $\text{SO}_2$ ), amoníaco ( $\text{NH}_3$ ), partículas finas ( $\text{PM}_{2.5}$ ), materia particulada total (TPM), compuestos particulados totales (TPC), carbono orgánico (OC) y carbono negro (BC). Es importante notar que los factores de emisión también dependen del tipo de uso de suelo. Para cada tipo de uso de suelo existe un conjunto de factores de emisión, conteniendo los 15 gases presentados en el estudio mencionado. Estos factores son presentado en la Tabla 4.1, donde la unidad de los valores corresponden a gramos de contaminante emitido por kilogramo de biomasa quemada.

Tabla 4.1: Factores de emisión ( $\text{g}/\text{kg}$  de biomasa quemada)

GAS	Bosque Tropical	Bosque Temperado	Matorral	Pastizal	Cultivo
$\text{CO}_2$	1643	1630	1716	1692	1537
$\text{CO}$	92	102	68	59	111
$\text{CH}_4$	5,1	5	2,6	1,5	6
$\text{H}_2$	3,2	1,8	0,97	0,97	2,4
$\text{NO}_x$	2,6	1,3	3,9	2,8	3,5
$\text{NO}$	0,91	0,34	1,4	0,74	1,7
$\text{NO}_2$	3,6	2,7	1,4	3,2	3,9
NMOC	24	14	4,8	9,3	57
NMHC	1,7	5,7	3,4	3,4	7
$\text{SO}_2$	0,45	1	0,68	0,48	0,4
$\text{NH}_3$	0,76	1,5	1,2	0,49	2,3
$\text{PM}_{2,5}$	9,7	13	9,3	5,4	5,8
TPM	13	18	15,4	8,3	13
TPC	5,2	9,7	7,1	3	4
OC	4,7	9,2	6,6	2,6	3,3
BC	0,52	0,56	0,5	0,37	0,69

### 4.3. Registro Temporal de Emisiones

Con el fin de generar un inventario de emisiones de mayor resolución que el registro de emisiones que se tiene actualmente en el RETC, se debe generar un registro tanto de área quemada como de las emisiones que ésta implica. Para este registro se requiere tener referencia temporal y espacial de las quemas existentes, pero además se debe tener datos sobre el tipo de uso de suelo en estas zonas. A continuación se describe la estrategia utilizada para generar el registro de emisiones, con sus respectivos usos de suelo.



### 4.3.1. Generación de Máscaras de Área Quemada

Con el fin de obtener todas las máscaras de área quemada, las cuales representan con píxeles blancos las zonas quemadas y con píxeles negros las regiones no quemadas, se utiliza el sistema de estimación de área quemada descrito anteriormente. Primero, cada una de las imágenes satelitales, almacenadas en formato PNG, pasa por el proceso de pre-procesamiento para adaptarse al modelo. De esta forma, a cada imagen se le agrega el *pad* correspondiente y se particiona en secciones de  $480 \times 720$ .

Luego, las partes de la imagen creadas en el pre-procesamiento se introducen como entrada al modelo UNet, del cual se extrae una matriz de probabilidad de que el píxel esté quemado. Una vez que se tiene la matriz de cada imagen, ésta pasa por el sistema de umbrales descritos anteriormente, con el cual se obtiene la máscara de área quemada. Este proceso se realiza para cada imagen y su resultado es almacenado en formato PNG. Es importante mencionar que el nombre asignado a las máscaras hace referencia al día de la toma de muestra y las coordenadas dentro de las cuales se encuentra el territorio mostrado, donde su formato es: fecha(año, mes, día)\_longitud máxima\_longitud mínima\_latitud máxima\_latitud mínima.

### 4.3.2. Datos tabulares del Área Quemada

Con las máscaras de área quemada se puede generar una base de datos tabular en donde se tenga información de cada uno de los píxeles clasificados como quemados. Para esto se requiere obtener tanto la fecha de la toma de muestra como las coordenadas centrales de cada píxel. Con este fin se realiza el siguiente procedimiento para cada una de las imágenes:

1. A partir del nombre se extrae la fecha de toma de muestra, dado que los primeros 4 caracteres corresponden al año, los siguientes 2 al mes y los últimos 2 enumeran el día. Tomando estos datos se genera la fecha de toma de muestra para cada píxel  $i$ , llamada  $date^i$ .
2. Se extraen las coordenadas máximas y mínimas a partir del nombre, donde estas se separan por el símbolo '\_'. Además con estas se calcula el rango de coordenadas. Para el rango de latitudes se utiliza la Ec. 4.2, donde  $lat_{max}$  es la latitud máxima y  $lat_{min}$ , la latitud mínima. De forma análoga, se calcula el rango de longitudes con la Ec. 4.3

$$\Delta_y = lat_{max} - lat_{min} \quad (4.2)$$

$$\Delta_x = lon_{max} - lon_{min} \quad (4.3)$$

3. Se lee el archivo PNG utilizando el paquete OpenCV [25] de Python, el cual interpreta la imagen como una matriz con valores 0, que representa el color negro, y 255, correspondiente al color blanco.
4. Se recorre la matriz por fila y la columna. Las filas tienen índices entre 0 y  $N_y$ , donde  $N_y$  es el alto de la matriz. Por su parte, las columnas tienen índices entre el 0 y  $N_x$ , con  $N_x$  el ancho de la matriz. Así, para el elemento en la fila  $y$  y la columna  $x$  se calculan las coordenadas máximas y mínimas de su píxeles respectivo utilizando las Ecuaciones 4.4, 4.5, 4.6 y 4.7. De esta forma para cada píxel  $i$  clasificado como área quemada se tiene su longitud máxima ( $lon_{max}^i$ ), longitud mínima ( $lon_{min}^i$ ), latitud máxima ( $lat_{max}^i$ ) y latitud mínima ( $lat_{min}^i$ ).

$$lon_{max}^i = lon_{max} - \frac{x \times \Delta_x}{N_x} \quad (4.4)$$

$$lon_{min}^i = lon_{max} - \frac{(x + 1) \times \Delta_x}{N_x} \quad (4.5)$$

$$lat_{max}^i = lat_{max} - \frac{y \times \Delta_y}{N_y} \quad (4.6)$$

$$lat_{min}^i = lat_{max} - \frac{(y + 1) \times \Delta_y}{N_y} \quad (4.7)$$

5. Con las coordenadas máximas y mínimas se calcula, para cada píxel  $i$ , la coordenada central  $(lat^i, lon^i)$ , como se muestra en las Ecuaciones 4.8 y 4.9.

$$lat^i = \frac{lat_{max}^i + lat_{min}^i}{2} \quad (4.8)$$

$$lon^i = \frac{lon_{max}^i + lon_{min}^i}{2} \quad (4.9)$$

Los datos extraídos para cada píxel  $(date^i, lat^i, lon^i)$  son almacenados en una estructura llamada **DataFrame** del paquete Pandas [26] de Python, la cual es luego almacenada en formato CSV (*Comma Separated Values*) usando la función `to_csv` de éste mismo paquete. Es importante mencionar que un **DataFrame** puede desplegarse con un formato tabular.

### 4.3.3. Pre-procesamiento de Mapa de Uso de Suelo

Como se menciona anteriormente, las coordenadas de la base de datos de píxeles quemados están en grados, mientras que las de la base de datos de tipo de uso de suelo están en UTM. Es por esto que se requiere compatibilizar ambos sistemas de coordenadas, es decir, dejar todos los datos usando el mismo formato de referencia espacial. Además, se busca transformar los datos a un formato tabular con el fin de facilitar la tarea de cruce de datos con el *dataset* tabular de área quemada generado.

#### 4.3.3.1. Proyección de coordenadas

Para compatibilizar las bases de datos obtenidas, se transforman las coordenadas del *dataset* de tipo de uso de suelo a coordenadas de latitud y longitud en grados. Para esto lo que se debe hacer es proyectar los datos de UTM a grados. Existen diversos paquetes de Python [18] para manipular archivos del tipo GeoTIFF, en particular, para este trabajo se selecciona el paquete GDAL [27].

Lo primero que se hace es leer el archivo con la función `Open` de GDAL, esta obtiene tanto los datos de clasificación de uso de suelo como la metadata de referencia espacial del archivo. De esta forma, se tiene el tipo de uso de suelo, las coordenadas en el eje  $x$  e  $y$  en UTM y la zona en la que éstas se encuentran. Para transformar las coordenadas UTM a latitud y longitud en grados, se utiliza la función `Warp` de GDAL, la cual reescribe el archivo

que se le entrega según los atributos ingresados. A la función `Warp` se le entrega el archivo y el atributo `dstSRS`, en el cual se especifica que el EPSG (*European Petroleum Survey Group*) es 4326, el cual corresponde al código de coordenadas en grados. De esta forma, se reescribe el archivo GeoTIFF con un sistema de referencia compatible al que se tiene en las imágenes satelitales.

#### 4.3.3.2. Transformación a Formato Tabular

Dado que las coordenadas de la base de datos de incendios con las del archivo GeoTIFF de tipo de uso de suelo son compatibles, es conveniente transformar el archivo GeoTIFF a datos tabulares, en donde se especifiquen las coordenadas centrales y el código de tipo de uso de suelo (ver Tabla 3.1). Lo primero que se hace es leer el archivo de uso de suelo con la función `Open` de GDAL, con la cual se obtiene una matriz con los códigos de uso de suelo, pero además se tiene la metadata con las coordenadas de latitud y longitud.

Para poder comparar y cruzar el *dataset* de área quemada con el de uso de suelo, se requiere transformar el tipo de uso de suelo a un formato tabular. Es importante notar que el archivo obtenido de [22] es extenso, por lo que el proceso de transformación no se puede realizar en todo el archivo. Para esto se divide el archivo GeoTIFF en 100 partes iguales, es decir, tienen la misma cantidad de píxeles, las cuales serán transformadas una a una a archivos CSV. Con este fin, se utiliza el paquete de Python `Rioxarray` [28], realizando el mismo proceso en cada una de las partes. Lo primero que se hace es leer cada archivo GeoTIFF con la función `open_rasterio`, para luego usar la función `to_dataframe`, la cual transforma los datos a formato tabular como se busca.

#### 4.3.4. Asignación de Tipo de Uso de Suelo

Si bien ambas bases de datos están en el mismo formato y con el mismo sistema de coordenadas, aún no pueden intersectarse, dado que solo se tienen los puntos centrales en el *dataset* de tipo de uso de suelo. Para cruzar ambos datos y poder asignar un tipo de uso de suelo a cada píxel quemado, se requiere generar cuadrados de 30 metros de largo y ancho alrededor de cada punto en el repositorio de tipo de usos de suelo, llamado *buffer*. Para realizar la intersección de *datasets* se utiliza el paquete `GeoPandas` [21] el cual se especializa en datos tabulares que incluyen información geográfica.

Lo primero que se hace es transformar la base de datos de área quemada de un `DataFrame` de `Pandas` a un `GeoDataFrame` de `GeoPandas`. Para esto se crea una nueva columna llamada `'geometry'`, la cual contiene el par de coordenadas centrales de cada píxel quemado en formato `Point` del paquete `Shapely` [29] de Python. Este proceso también se realiza con los datos tabulares de tipo de uso de suelo. Es importante mencionar que ambos `GeoDataFrames` tienen el sistema de referencia de coordenadas (CRS) de EPSG 4326, es decir, coordenadas de latitud y longitud en grados.

Al tener ambos *datasets* en formato de `GeoDataFrame` se puede crear el *buffer* e intersectar ambas bases, usando como conexión sus coordenadas geográficas. Lo primero que se debe mencionar es que como se especifica anteriormente, el archivo entregado en [22] tiene una resolución de 30 metros, esto quiere decir que para generar el *buffer* se requiere un cuadrado cuyas coordenadas estén dadas por:  $(x - 15, y - 15)$ ,  $(x - 15, y + 15)$ ,  $(x + 15, y - 15)$  y

$(x + 15, y + 15)$ , donde  $x$  corresponde a la latitud e  $y$  a la longitud, ambas en metros. Sin embargo, la formación del *buffer* no es trivial, ya que se tienen las coordenadas en grados. Es por esto que se toma la columna de 'geometry' del **GeoDataFrame** de tipo de uso de suelo y se cambia su CRS a EPSG 3395, el cual corresponde al sistema *World Mercator* que tiene sus coordenadas en metros. Al tener los puntos centrales en metros, se puede ejecutar la función **buffer** de GeoPandas, la cual como lo dice su nombre crea un *buffer* alrededor de la estructura de Shapely que se le entregue. En este caso se le entrega que la dimensión del *buffer* sea de 15 metros y de forma cuadrada, con lo que se obtiene la resolución buscada del archivo original de tipo de uso de suelo. Una vez que se obtienen la lista de *buffers* se deben volver a transformar a coordenadas en grados, para lo que nuevamente se utiliza la función **to\_crs** de GeoPandas, volviendo a el CRS de EPSG 4326. Finalmente, se reemplaza la columna 'geometry' del *dataset* de uso de suelo por los *buffers* correspondientes creados para cada punto.

Una vez que se tienen el **GeoDataFrame** de uso de suelo con los *buffer* correspondientes se pueden cruzar ambos *datasets* utilizando la función **sjoin** de GeoPandas. Esta función realiza una intersección de los datos utilizando las coordenadas de los **GeoDataFrames** que se le entregan, devolviendo como resultado una tabla que contiene las filas del **GeoDataFrame** de área quemada acompañadas de sus códigos de tipo de uso de suelo respectivo, según el *buffer* que comprende sus coordenadas centrales. Así, se obtiene finalmente, una base de datos en donde se presentan los datos de todos los píxeles que son clasificados como área quemada, con sus coordenadas centrales en grados, su fecha y el código de su tipo de uso de suelo.

Para utilizar la función creada para el cálculo de emisiones se deben traducir los códigos de tipos de uso de suelo a las categorías conocidas; éstas son: bosque templado, matorrales, pastizales y cultivos, dado que Chile no presenta bosques tropicales. Para esto, se reemplazan los códigos por las clasificaciones según la Tabla 3.1. Como se menciona anteriormente, existen más clasificaciones de tipo de uso de suelo, pero para este trabajo no se consideran como parte de incendios forestales y agrícolas, por lo que su tipo de uso de suelo se clasifica como 'Otro'. Aquellos datos con clase de uso de suelo 'Otro' no son tomados en cuenta para la generación del registro temporal de emisiones.

### 4.3.5. Cálculo de Emisiones

Finalmente, el paso restante para generar el registro temporal de área quemada y emisiones es el cálculo de éstas últimas. Para esto se utiliza la Ec. 2.1 y el *dataset* que contiene los píxeles que representan área quemada, sus coordenadas y sus tipos de uso de suelo. Dado que la base de datos a usar representa todos los píxeles que representan área quemada, se pueden calcular las emisiones para cada una de las filas tomando como área quemada  $100 \text{ m}^2$  y su tipo de uso de suelo. De esta forma, para cada píxel se calculan las emisiones en gramos ( $E_i$ ) de cada gas o partícula  $i$  según la Ec. 4.10, donde al igual que para el cálculo de emisiones se asume que la fracción de área quemada es 1, dado que con una resolución de 10 metros, esta suposición no debería generar un alto error. Como se menciona anteriormente, la carga de biomasa  $B(x)$  (extraída de [16]) y el coeficiente de emisión  $ef_i$  de cada gas o partícula  $i$  dependen del tipo de uso de suelo.

$$E_i = 10^2 \times B(x) \times 1 \times ef_i \quad (4.10)$$

## 4.4. Generación de Interfaz de Usuario

Para la generación de la interfaz de usuario se busca generar dos secciones de la plataforma: la primera tiene como fin el estimar las emisiones en una imagen proporcionada por el usuario, mientras que la segunda sección busca exponer el registro temporal obtenido a partir de los datos adquiridos.

A continuación, se muestran los requerimientos que se eligen para la interfaz preliminar. En la primera parte, donde el usuario puede ingresar una imagen, se necesita una forma de entregar la entrada al modelo, además, se busca que la interfaz muestre de forma clara los resultados tanto de las emisiones como del área quemada. Mientras que, en la sección del registro temporal, se busca que el usuario pueda obtener información filtrada según sus preferencias, desplegando de forma gráfica los resultados que se buscan.

### 4.4.1. Estimación de Área Quemada a Partir de Imagen Proporcionada por el Usuario

Para esta sección de la interfaz de la aplicación web se espera que a partir de una imagen que entregue el usuario se puedan estimar las emisiones y el área quemada en ésta misma. Es importante notar que no se tienen restricciones sobre la resolución de la imagen, pero se espera que la imagen no tenga un alto nivel de nubosidad (menos al 10%), las bandas presentadas sean las bandas SWIR, NIR y *Red* (en este orden) y el archivo entregado esté en formato PNG (*Portable Network Graphics*).

De esta forma los mínimos entregables para la interfaz de usuario donde se estiman las emisiones de una foto ingresada son:

1. Un botón en donde se puedan cargar las imágenes en formato PNG.
2. Una visualización tanto de la imagen ingresada como la máscara que representa los píxeles clasificados como áreas quemadas.
3. Una casilla en donde se pueda entregar la información de la resolución de la imagen en metros.
4. Una casilla de selección en donde se pueda elegir el tipo de uso de suelo de la zona de la cual se extrae la imagen.
5. Una casilla en donde se seleccione el gas o partícula de interés o se puedan seleccionar todos ellos.
6. Entrega del área quemada estimada en la imagen entregada con su respectiva unidad de medida.
7. Entrega de las emisiones de los gases y partículas seleccionados con su respectiva unidad de medida.

Para la generación de la interfaz de usuario se utiliza el paquete Streamlit [30] de Python, el cual corresponde a un paquete de código libre, que simplifica la tarea de formar aplicaciones web con conexión a modelos de Machine Learning.

#### 4.4.2. Registro Temporal de Emisiones y Área Quemada

En esta sección de la aplicación web se espera que se pueda mostrar al usuario el registro de emisiones generado de forma fácil para interpretar, pudiendo visualizar las emisiones con gráficos con referencia temporal. De esta forma, se espera que el usuario pueda ingresar un período de tiempo, dentro del rango temporal estipulado donde se encuentran datos satelitales, y se muestren las emisiones y el área quemada en él. Es por esto que se espera que los entregables para esta porción de la interfaz sean:

1. Una casilla en donde se pueda ingresar una fecha de inicio de los datos a analizar
2. Una casilla en la que se pueda ingresar la fecha de término de los datos a visualizar.
3. Un gráfico que muestre el total de área quemada en las fechas determinadas anteriormente.
4. Una casilla para seleccionar el o los contaminantes a visualizar en el rango de fechas definido anteriormente.
5. Un conjunto de casillas para seleccionar la localidad o comuna a estudiar.

Es importante notar que no se tienen imágenes satelitales de todo el terreno chileno, por lo que no se tendrán disponibles todos los sectores del territorio nacional. De esta misma forma, puede no tenerse registros diarios de cada localidad, alcanzando una latencia de a lo más 10 días por lugar. Sin embargo, se espera que la pérdida de detección por la latencia sea disminuida, dado que el modelo no solo detecta incendios activos, sino que también detecta *burn scars* que quedan luego de incendios.

# Capítulo 5

## Resultados Obtenidos

### 5.1. Estimación de Área Quemada

A continuación se exponen visualizaciones de resultados de la estimación de área quemada. Es importante notar que, si bien el fin del proyecto es la generación de un inventario de emisiones, dado a las diferencias de resolución y la cantidad de datos, la comparación de emisiones se hace compleja con otros productos que calculan emisiones. Es por esto que el método toma como 'verdad adherida' el cálculo de emisiones utilizado por FINN [16], el cual es uno de los estándares mundiales para la estimación de éstas. De esta forma, la opción seleccionada para evaluar el método descrito en el presente informe es analizando el desempeño en estimación de área quemada.

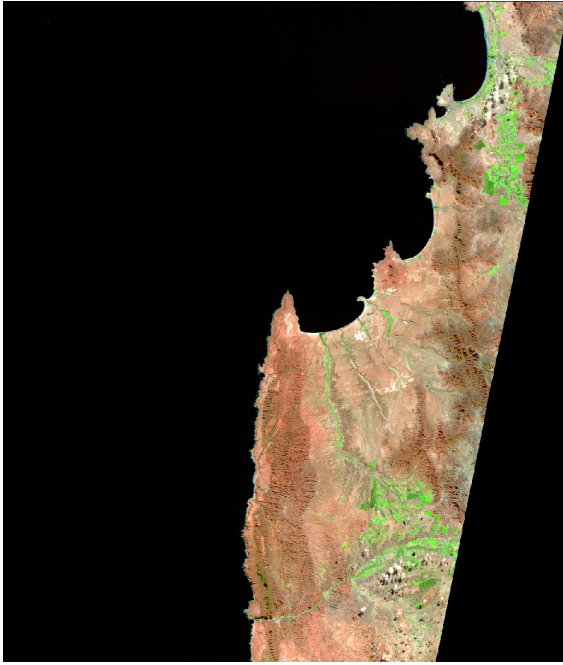
Uno de los mayores referentes de área quemada es el producto satelital es *MODIS Burned Area* [31], el cual tiene resolución temporal diaria y resolución espacial de 500 metros. Esto significa que cada uno de sus píxeles corresponde a una superficie de 250.000 m<sup>2</sup>, lo que corresponde a 2.500 veces lo que cubre uno de los píxeles de las imágenes de Sentinel-2, de donde fueron descargadas las utilizadas para generar el registro. Es por lo anterior que la comparación entre ambos productos no es directa, y para mostrar el desempeño del modelo de segmentación a continuación se muestran ejemplos de resultados obtenidos con él.

#### 5.1.1. Ejemplos de Desempeño

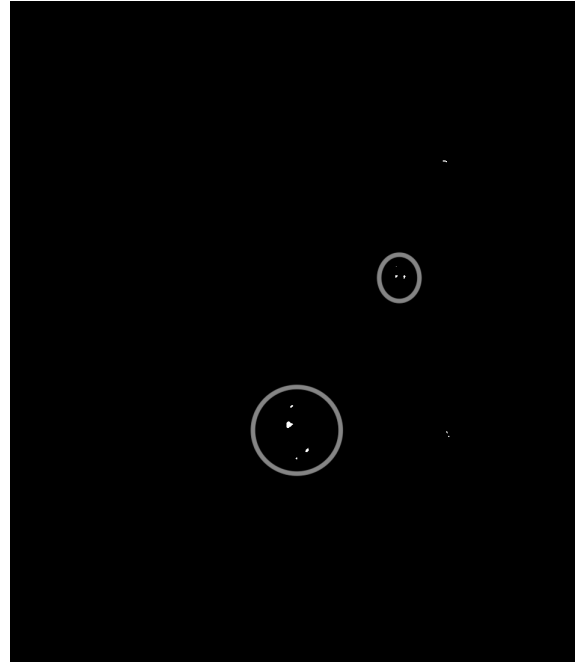
Para exponer el desempeño de la estimación de área quemada, en esta sección se realiza un proceso cualitativo, dado que no se tiene un estándar dorado. De esta forma, se mostrarán ejemplos en donde existen errores de detección y donde hay detección exitosa.

##### 5.1.1.1. Falsos Positivos

Un falso positivo corresponde a una identificación errónea, es decir, se identifica un incendio o *burn scar* donde no lo hay. Uno de los ejemplos más claros de falsos positivos es cuando las imágenes presentan zonas secas cercanas a zonas vegetadas, dado que éstas son detectadas como incendios dado el contraste entre ambos tipos de suelo. Como se puede visualizar en la Figura 5.1 las zonas secas se muestran en un color rojo anaranjado, el cual si bien no representa área quemada, se asemeja. Es por esto que se nota una pequeña detección en su máscara de área quemada, la cual está mal clasificada.



a Imagen satelital



b Máscara de área quemada

Figura 5.1: Ejemplo de falso positivo en zona seca, identificando las zonas erróneamente clasificadas como quemadas en círculos en la imagen de la izquierda.

Otro caso en donde se presentan falsas clasificaciones de incendios es en presencia de zonas urbanas rodeadas por vegetación. Los sectores en donde no hay vegetación se visualizan de un color similar a las *burn scars*, pero como se aprecia en la Figura 5.2 tienen una forma regular y por lo general geométrica, mientras que las quemadas tienden a tener bordes irregulares.





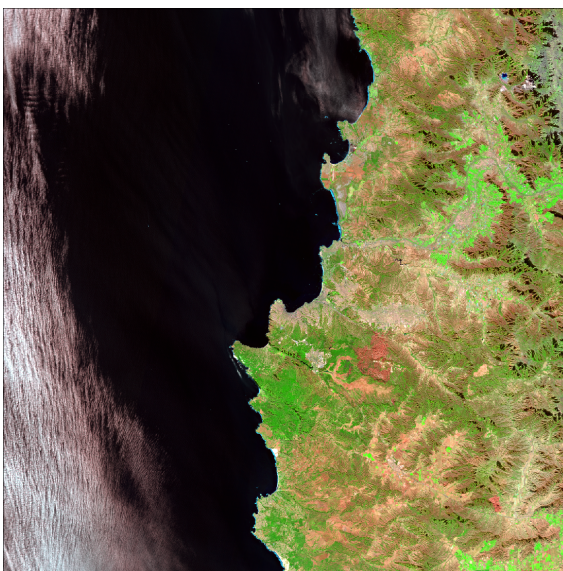
a Imagen satelital



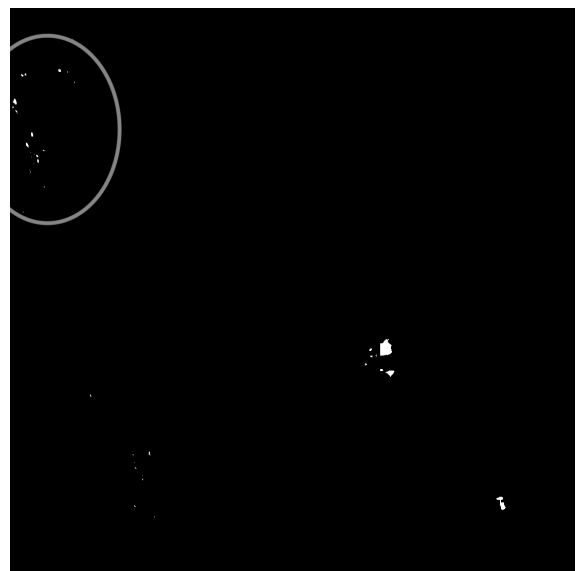
b Máscara de área quemada

Figura 5.2: Ejemplo de falso positivo en zona urbana rodeada de vegetación, identificando las zonas erróneamente clasificadas como quemadas en círculos en la imagen de la izquierda.

Finalmente, es importante notar que la nubosidad introduce ruido a las imágenes satelitales. Es por esto que las nubes pueden introducir falsos positivos en la estimación de área quemada, como por ejemplo la estimación de área quemada en la Figura 5.3 (b) donde en la esquina superior, se notan clasificaciones de incendio, cuando en la Figura 5.3 (a) se nota que en realidad se trata de nubes sobre un cuerpo de agua, por lo que la probabilidad de que se trate realmente de área quemada es casi nula.



a Imagen satelital



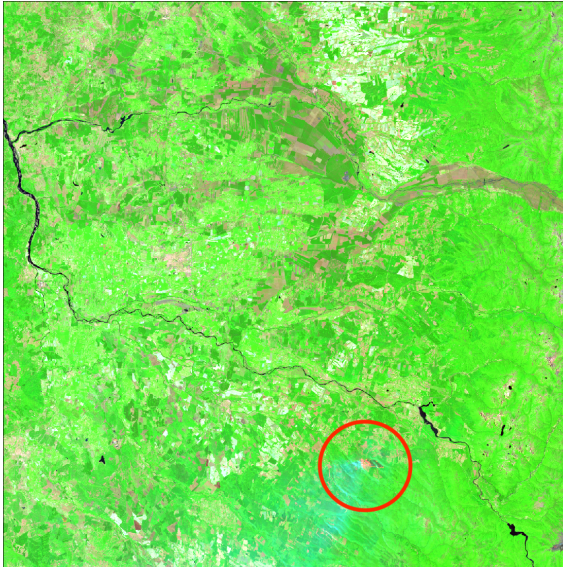
b Máscara de área quemada

Figura 5.3: Ejemplo de falso positivo introducido por nubosidad

### 5.1.1.2. Falsos Negativos

Falsos negativos hace alusión a clasificaciones erróneas porque existe un incendio pero éste no es identificado. Esto ocurre principalmente en la presencia de incendios activos muy pequeños como el que se puede visualizar en la Figura 5.4 (a) con un color naranja fuerte y una estela de humo color blanco. La detección en este caso solo toma en cuenta la zona de color café rojizo que está cercana al incendio activo. Esto mismo ocurre en presencia de incendios activos en conjunto con montañas secas, las cuales se visualizan con color café como se observa en la Figura 5.5.

De los ejemplos vistos se infiere que existen falsos negativos cuando se trata de incendios activos acompañados de zonas que generan error; éstas son: zonas secas, nieve, nubes o zonas quemadas de mayor área. Este error de clasificación se atribuye al sistema de umbrales en donde se prima el evitar una gran cantidad de falsos positivos (como los que podría generar la montaña seca de la Figura 5.5) por sobre detectar un incendio pequeño, dado que estos tienden a ser menores al 0,1 %, mientras que las zonas que generan error son mayores a este porcentaje.

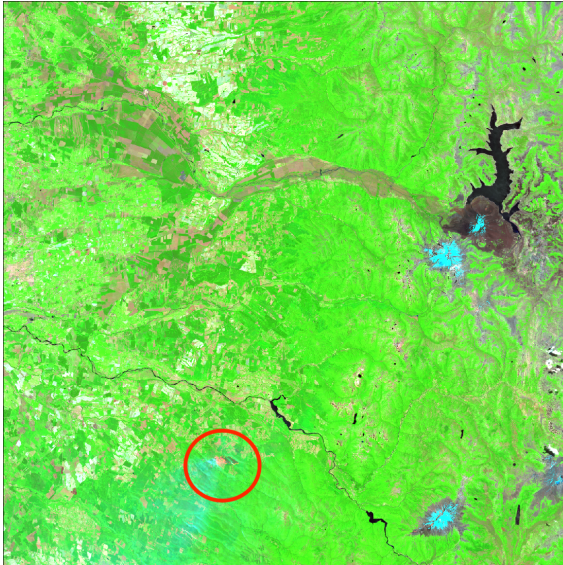


a Imagen satelital

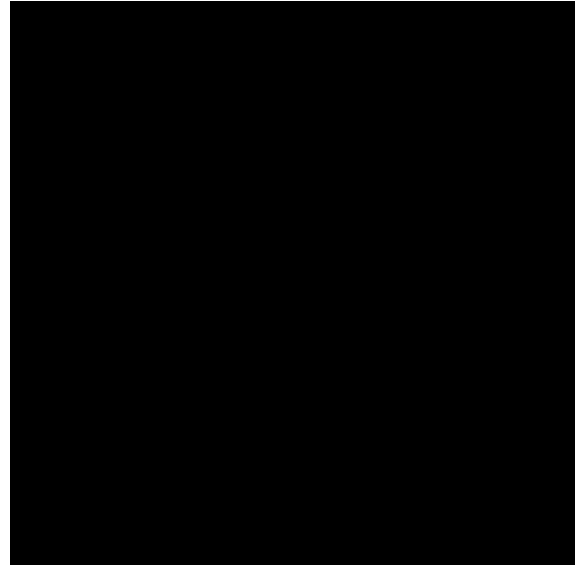


b Máscara de área quemada

Figura 5.4: Ejemplo de falso negativo de incendios activos pequeños



a Imagen satelital



b Máscara de área quemada

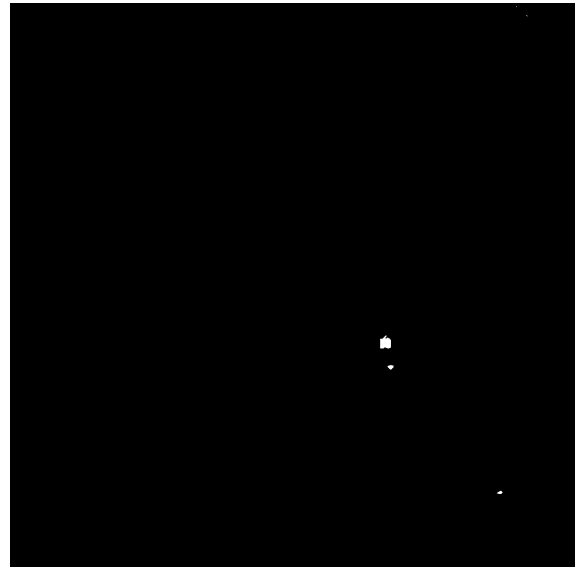
Figura 5.5: Ejemplo de falso negativo incendio activo pequeño cercano a montañas secas

### 5.1.1.3. Verdaderos Positivos

Los verdaderos positivos se refieren a las clasificaciones de área quemada cuando realmente se encuentra un incendio o *burn scar* en la imagen satelital. El caso con mayor claridad de clasificación correcta de zonas quemadas es un zonas de alto nivel de vegetación donde se visualizan *burn scars*. Un ejemplo claro de identificación exitosa se muestra en la Figura 5.6, donde la zona quemada visualizada en la imagen satelital con color café rojizo se identifica claramente en la máscara de área quemada, es más, incluso se identifica la *burn scar* pequeña cercana a la esquina inferior derecha. Sin embargo, es importante notar que, si bien el área quemada es detectada, la mascara no tiene la misma forma que se muestra en la imagen satelital dado que el modelo puede presentar un error de predicción.



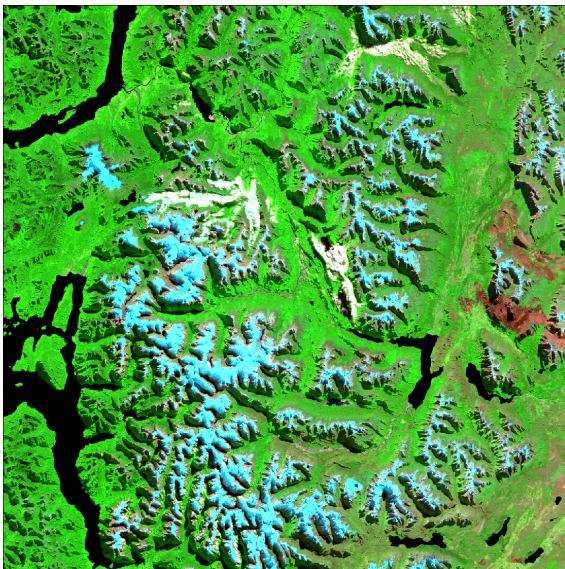
a Imagen satelital



b Máscara de área quemada

Figura 5.6: Ejemplo de identificación exitosa de *burn scar* rodeada de vegetación

Por otra parte, se puede notar que en la presencia de *burn scars* la detección suele ser exitosa sin importar el ambiente. Esto se puede apreciar en el caso mostrado en la Figura 5.7, donde la detección es correcta a pesar de la presencia de nieve, la cual se observa de color celeste en la imagen satelital.



a Imagen satelital



b Máscara de área quemada

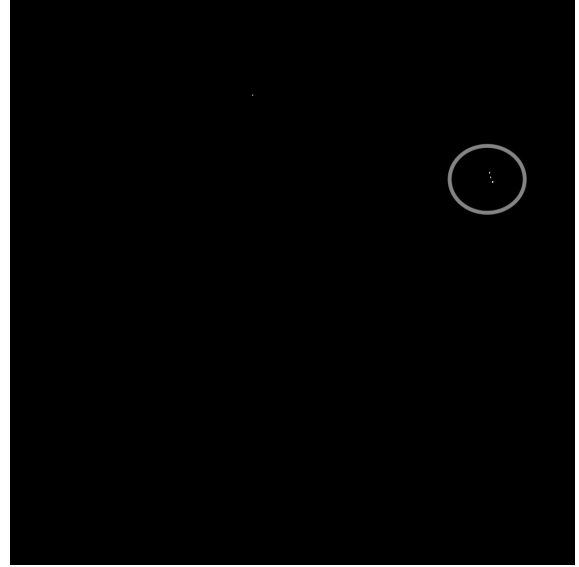
Figura 5.7: Ejemplo de identificación exitosa en zonas con nieve

También es importante mencionar que la detección no solo sirve para *burn scars* grandes como las observadas en las Figuras 5.6 y 5.7, sino que el método propuesto también es capaz de identificar marcas pequeñas como la que se aprecia en la Figura 5.8 (a) en la esquina

superior derecha.



a Imagen satelital



b Máscara de área quemada

Figura 5.8: Ejemplo de identificación exitosa en zonas con nieve

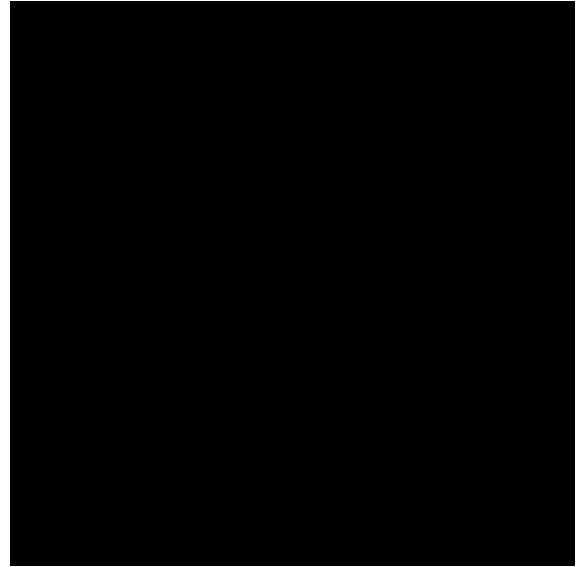
#### 5.1.1.4. Verdaderos negativos

Las muestras clasificadas como verdaderos negativos corresponden, en este caso, a imágenes en donde no se identifican incendios y realmente no los había. Si bien en el dataset se muestra una gran cantidad de verdades negativos a continuación se exponen solo aquellos que representan un agregado de imágenes similares.

Un conjunto importante de confirmar que contiene una alta tasa de verdaderos negativos son zonas vegetadas, como bosques y cultivos, en donde no se visualizan incendios activos ni *burn scars*. Un ejemplo claro de identificación de zonas vegetadas no quemadas es el que se ve en la Figura 5.9, en donde la máscara (b) no muestra identificación de incendios.



a Imagen satelital

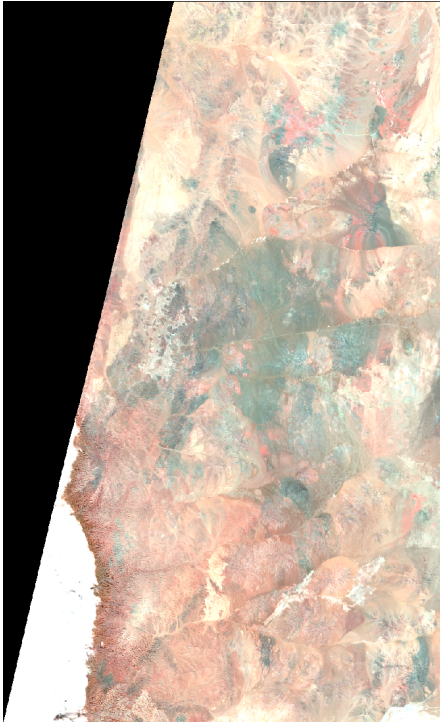


b Máscara de área quemada

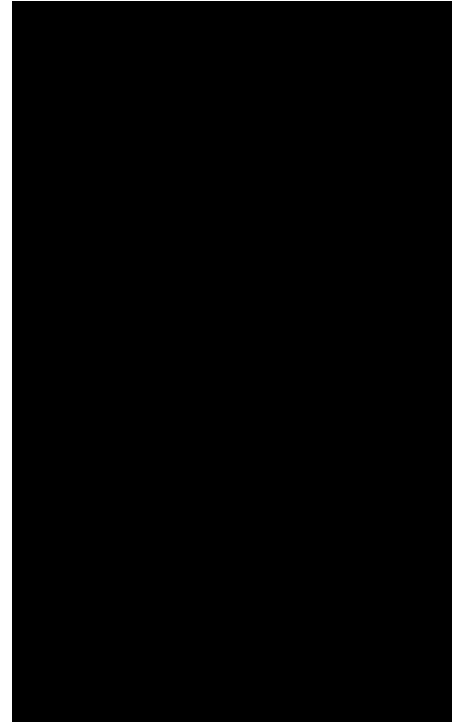
Figura 5.9: Ejemplo de identificación de área no quemada en zonas vegetadas

Otro tipo de terrenos en donde es importante que no se presenten falsos positivos, es decir, se tengan verdaderos negativos, es en zonas mayormente secas. Como se menciona anteriormente, las zonas secas tienden a presentar colores rojizos por lo que potencialmente podría existir una alta tasa de identificación errónea de área quemada. A pesar de los colores rojizos en este tipo de suelo, el sistema de umbrales es capaz de mitigar este error, obteniendo identificación exitosa de zonas no quemadas como se aprecia en la Figura 5.10. Se destaca este agregado de imágenes dado que, por las tonalidades que tienen en las bandas *Short Wave Infrared 2*, *Near Infrared* y *Red*, su representación se asemeja a zonas quemadas, por lo que potencialmente podrían agregar falsos positivos.

Finalmente, otro grupo importante a considerar son las imágenes en donde se encuentra territorio mixto. Un ejemplo de este conjunto es el mostrado en la Figura 5.11, en donde se encuentra nieve, vegetación y zonas secas. Este tipo de terrenos, al igual que los de zonas secas, también tienen un gran potencial de error dado que existe un alto contraste dentro de la imagen, además de presentarse zonas de color café, similar a la representación de área quemada.

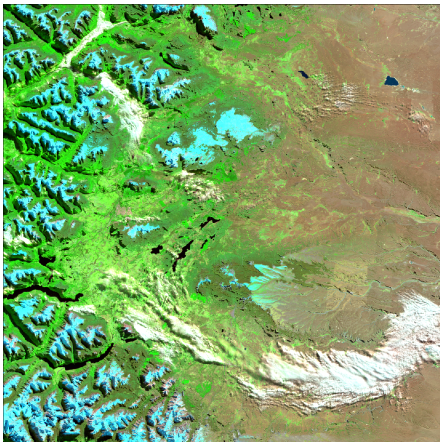


a Imagen satelital

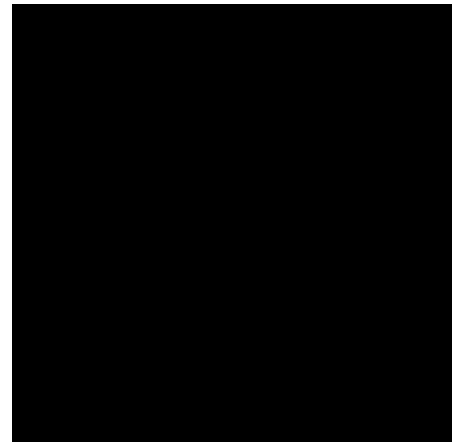


b Máscara de área quemada

Figura 5.10: Ejemplo de identificación de área no quemada en zonas mayormente secas



a Imagen satelital



b Máscara de área quemada

Figura 5.11: Ejemplo de identificación de área no quemada en zonas de territorio mixto

## 5.2. Registro Temporal

Como se menciona anteriormente, se tiene un registro de 11.084 imágenes satelitales con menos de 10% de nubosidad del territorio chileno. Además, se tiene un archivo que cubre todo Chile, etiquetándolo según su tipo de uso de suelo. A partir de estas dos bases de datos

se crea un registro temporal en donde se especifique la zona, la fecha y las emisiones durante un incendio. Para esto se realizan los procesos especificados en la sección de Metodología; éstos son: transformar las máscaras de área quemada a datos tabulares, el pre-procesamiento de mapa de uso de suelo, asignar el tipo de uso de suelo a los datos tabulares de área quemada y finalmente el cálculo de emisiones.

### 5.2.1. Datos Tabulares de Área Quemada

Al realizar el proceso descrito, se obtiene una base de datos con 3 columnas: **date**, **lon** y **lon**, como se puede apreciar en la Figura 5.12. De esta forma, cada fila representa 100 m<sup>2</sup> de área quemada cuyo punto central es (**lon**, **lat**) y la fecha del incendio está por el atributo **date**. Al transformar los datos a formato tabular se obtiene un total de 2.035.677 filas, lo que significa que se identifican 203.567.700 m<sup>2</sup> quemados en todo el *dataset*.

	<b>date</b>	<b>lon</b>	<b>lat</b>
<b>0</b>	2019-02-06	-70.930778	-29.232902
<b>1</b>	2019-02-06	-70.931879	-29.232902
<b>2</b>	2019-02-06	-70.932980	-29.232902
<b>3</b>	2019-02-06	-70.928576	-29.233802
<b>4</b>	2019-02-06	-70.929677	-29.233802

Figura 5.12: Primeras 4 filas de la tabla de píxeles de área quemada.

### 5.2.2. Datos Tabulares de Tipo de Uso de Suelo

Al haber transformado el sistema de referencia a coordenadas en grados, se genera un **DataFrame** del archivo GeoTIFF. Como se puede observar en la Figura 5.13, este *dataset* tiene 5 columnas, de las cuales se utilizan 3: **y**, **x** y **data**. La variable **y** representa la latitud del punto central de cada píxel, mientras que **x**, la longitud de éste mismo punto. Por su parte, la columna **data** representa la clasificación de uso de suelo entregada en [22]. Al tener todos los datos cargados se eliminan las columnas con valor 0 en la columna **data**, de lo que se obtiene un total de 997.589.107 filas.

	<b>band</b>	<b>y</b>	<b>x</b>	<b>spatial_ref</b>	<b>data</b>
<b>0</b>	1	-55.618845	-69.043662	0	150
<b>1</b>	1	-55.623666	-68.063248	0	241
<b>2</b>	1	-55.625809	-69.052231	0	150
<b>3</b>	1	-55.616703	-69.052767	0	150
<b>4</b>	1	-55.731328	-67.420262	0	1010

Figura 5.13: Primeras 4 filas del *dataset* tabular de tipo de uso de suelo



### 5.2.3. Registro temporal de emisiones

Al tener ambas bases de datos, éstas se relacionan con el proceso descrito en la asignación de tipo de uso de suelo, donde es importante notar que no necesariamente todos los píxeles que representan área quemada son parte del territorio chileno. Lo anterior se debe a que las imágenes descargadas contienen partes del territorio chileno, pero pueden haber imágenes que además contengan terrenos de otros países o incluso cuerpos de agua. Al intersectar el *dataset* de uso de suelo con el de área quemada, asignar las comunas y regiones según las coordenadas, filtrar por uso de suelo que se considera como quema forestal y agrícola y calcular las emisiones, se obtiene una base de datos de 620.151 filas.

Como se aprecia en la Figura 5.14, la base de datos del registro de emisiones tiene 24 columnas, con las cuales se pueden referenciar los píxeles de área quemada tanto en el tiempo como geográficamente, además de obtener las emisiones de 15 gases y partículas generados por la quema en él. A continuación se describe, para cada columna, el significado de su valor.

- **date:** Este valor corresponde a la fecha en la que se toma la imagen satelital.
- **lat:** Esta columna contiene el valor de la latitud en grados de la coordenada central del píxel que representa el área quemada.
- **lon:** De forma complementaria a la columna **lat**, **lon** corresponde a la longitud en grados del punto central del píxel.
- **landcover:** Esta columna comprende el tipo de uso de suelo. Como ya fue mencionado, para este estudio se utilizan 4 clasificaciones para calcular las emisiones: pastizales (**Grasslands**), matorrales (**Shrublands**), bosques templados (**Temperate Forest**) y cultivos (**Croplands**).
- **region:** Esta columna entrega el nombre de la región de Chile en la que se encuentra el píxel de área quemada.
- **comuna:** Este valor corresponde al nombre de la comuna en la que se encuentra el píxel de área quemada.
- **provincia:** Corresponde a la provincia en la que se detecta el píxel que representa área quemada.
- **burned\_area\_m2:** Esta columna entrega el valor es el área quemada en metros cuadrados.
- **co2:** Este valor corresponde a la emisiones de dióxido de carbono ( $\text{CO}_2$ ) en gramos.
- **co:** Esta columna comprende las emisiones en gramos de monóxido de carbono ( $\text{CO}$ ).
- **ch4:** Este valor entrega las emisiones en gramos de metano ( $\text{CH}_4$ ).
- **h2:** Emisiones de dihidrógeno ( $\text{H}_2$ ) en gramos.
- **nox:** Esta columna entrega las emisiones de gramos de la combinación gases reactivos que contienen nitrógeno y oxígeno ( $\text{NO}_x$ ).

- **no2**: Emisiones de dióxido de nitrógeno (NO<sub>2</sub>) en gramos.
- **nmoc**: Este valor representa las emisiones de compuestos orgánicos no metánicos (NMOC) en gramos.
- **nmhc**: Esta columna entrega las emisiones en gramos de hidrocarburos sin metano (NMHC).
- **so2**: Emisiones en gramos de dióxido de azufre (SO<sub>2</sub>).
- **nh3**: Este valor corresponde a la emisión de amoníaco (NH<sub>3</sub>) en gramos.
- **pm25**: Esta columna entrega las emisiones en gramos de partículas finas (PM2.5).
- **tpm**: Este valor representa las emisiones de materia particulada total (TPM) en gramos.
- **tpc**: Emisión en gramos de compuestos particulados totales (TCP).
- **oc**: Esta columna entrega las emisiones de carbono orgánico (OC) en gramos.
- **bc**: La última columna, representa las emisiones en gramos de carbono negro (BC).

	date	lat	lon	landcover	region	comuna	provincia	burned_area_m2								
0	2019-04-11	-55.675554	-67.387563	Grasslands	Región de Magallanes y Antártica Chilena	Cabo de Hornos	Antártica Chilena	10								
1	2019-04-11	-55.676480	-67.387563	Grasslands	Región de Magallanes y Antártica Chilena	Cabo de Hornos	Antártica Chilena	10								
2	2019-04-11	-55.677405	-67.387563	Grasslands	Región de Magallanes y Antártica Chilena	Cabo de Hornos	Antártica Chilena	10								
3	2019-02-04	-55.276240	-70.191360	Temperate Forest	Región de Magallanes y Antártica Chilena	Cabo de Hornos	Antártica Chilena	10								
4	2019-02-04	-55.276240	-70.192971	Temperate Forest	Región de Magallanes y Antártica Chilena	Cabo de Hornos	Antártica Chilena	10								

	co2	co	ch4	h2	nox	no	no2	nmoc	nmhc	so2	nh3	pm25	tpm	tpc	oc	bc
	53805.6	1876.2	47.70	30.846	89.04	23.532	101.76	295.74	108.12	15.264	15.582	171.72	263.94	95.40	82.68	11.766
	53805.6	1876.2	47.70	30.846	89.04	23.532	101.76	295.74	108.12	15.264	15.582	171.72	263.94	95.40	82.68	11.766
	53805.6	1876.2	47.70	30.846	89.04	23.532	101.76	295.74	108.12	15.264	15.582	171.72	263.94	95.40	82.68	11.766
	570011.0	35669.4	1748.50	629.460	454.61	118.898	944.19	4895.80	1993.29	349.700	524.550	4546.10	6294.60	3392.09	3217.24	195.832
	570011.0	35669.4	1748.50	629.460	454.61	118.898	944.19	4895.80	1993.29	349.700	524.550	4546.10	6294.60	3392.09	3217.24	195.832

Figura 5.14: Primeras 4 filas del registro temporal de emisiones en Chile

### 5.3. Interfaz de Usuario

A continuación se ilustran los atributos de la aplicación web generada para la visualización de resultados. Como se menciona anteriormente, esta interfaz de usuario se divide en dos grandes partes; éstas son: la estimación de emisiones en imágenes ingresadas por el usuario y la visualización del registro temporal de emisiones en Chile. A modo general, se busca que el resultado de la aplicación web sea una interfaz de uso intuitivo en donde los resultados desplegados sean de fácil análisis para quienes visualicen.

La primera función clave desarrollada es una caja de selección en donde el usuario pueda elegir si quiere ingresar una imagen satelital o si quiere visualizar el registro de emisiones

generado. Se puede ver que en la barra lateral de la Figura 5.15, encerrado en rojo, el botón de selección.

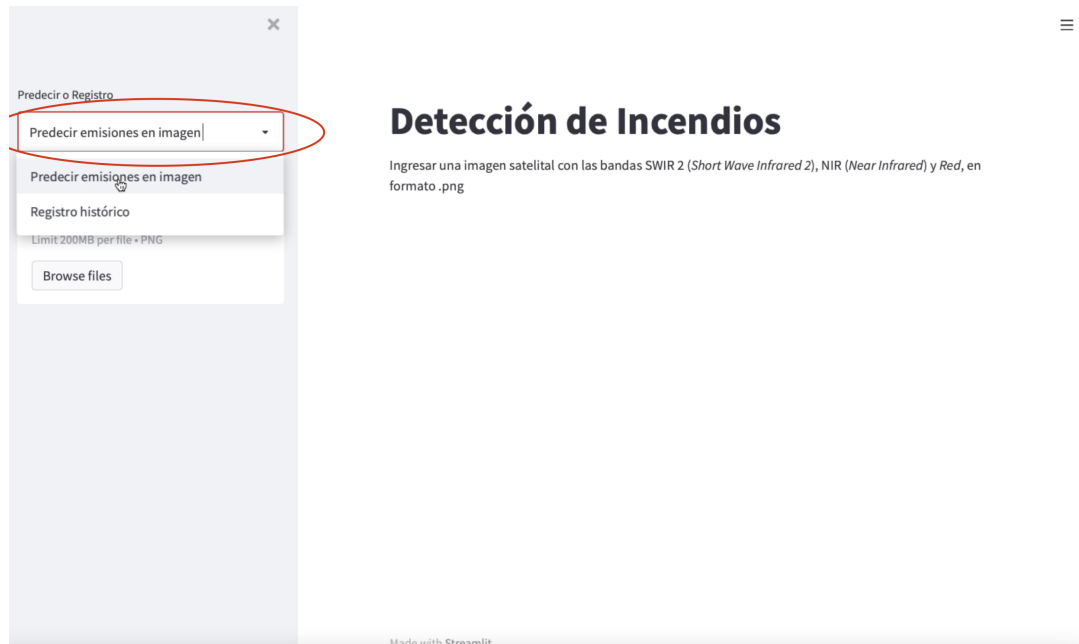


Figura 5.15: Botón para ingreso de datos por el usuario

### 5.3.1. Estimación de Emisiones por Imagen Ingresada

Como se menciona en el Capítulo 4 se busca tener características mínimas en la interfaz de ingreso de datos. El primer rasgo importante es un botón para ingreso de datos, el cual al igual que la caja de selección, se encuentra en la barra lateral de la aplicación web. Este botón se visualiza en la Figura 5.16, para ingresar una imagen se tienen 2 opciones:

1. Hacer click sobre el el botón que dice '**Browse files**', luego se abre un buscador desde donde ingresar a los archivos locales del computador y seleccionar la imagen satelital.
2. Arrastrar el archivo PNG a la caja encerrada en rojo de la Figura 5.16.

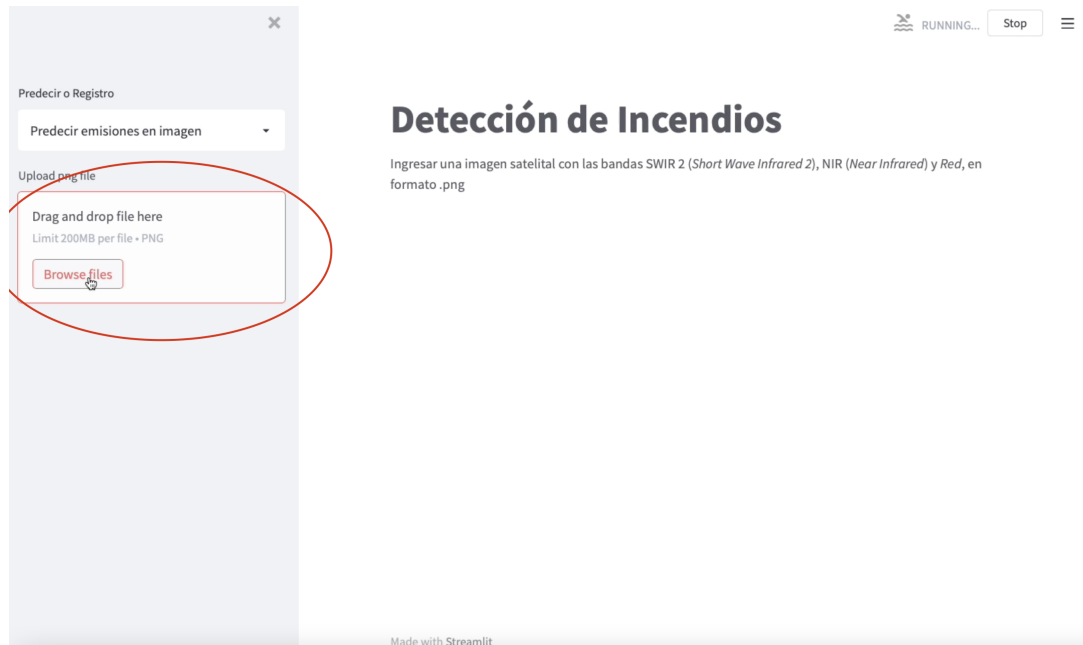


Figura 5.16: Botón de ingreso de datos

Una vez que el archivo PNG de la imagen satelital de la cual se espera estimar sus emisiones fue ingresada, esta se despliega en la interfaz. Para esto se intercambian nuevamente los canales 1 y 3, dado que como se menciona anteriormente, al guardar el archivo PNG éstos se intercambian. Luego, se muestra la máscara de área quemada entregada por el modelo y el sistema de umbrales. De ésta máscara se calcula la cantidad de píxeles quemado.

Una vez que se tienen los píxeles quemados el usuario debe ingresar los datos restantes para la estimación de emisiones. Es por esto que se agrega una caja de selección para ingresar el tipo de uso de suelo, la cual se puede ver encerrada en rojo en la Figura 5.17. Además, se requiere la resolución en metros de la imagen, dado que como se muestra en la Ec. 4.1 el área quemada depende de la medida de los píxeles. Para que el usuario entregue la resolución de la imagen satelital se genera la caja de ingreso de datos numéricos que se ve encerrada en azul en la Figura 5.17.

Finalmente, el usuario puede seleccionar en el botón encerrado en verde de la Figura 5.17 el contaminante del cual quiere obtener las emisiones. En esta lista hay 15 opciones de emisiones: dióxido de carbono ( $\text{CO}_2$ ), monóxido de carbono ( $\text{CO}$ ), metano ( $\text{CH}_4$ ), dihidrógeno ( $\text{H}_2$ ), el grupo de gases reactivos que contienen nitrógeno y oxígeno ( $\text{NO}_x$ ), dióxido de nitrógeno ( $\text{NO}_2$ ), compuestos orgánicos no metánicos (NMOC), hidrocarburos sin metano (NMHC), dióxido de azufre ( $\text{SO}_2$ ), amoníaco ( $\text{NH}_3$ ), partículas finas ( $\text{PM}_{2.5}$ ), materia particulada total (TPM), compuestos particulados totales (TPC), carbono orgánico (OC) y carbono negro (BC). Además, está la opción de seleccionar todos los gases y partículas, donde se muestra una tabla con todas las emisiones en toneladas, subrayando en amarillo el compuesto con mayor emisión.

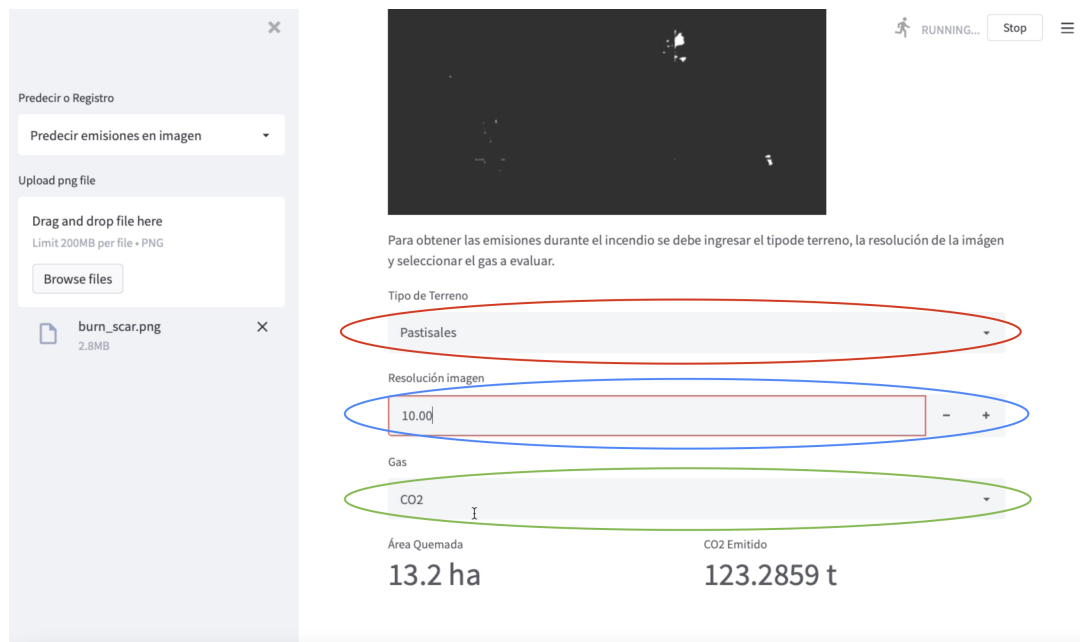


Figura 5.17: Ingreso de datos adicionales para la estimación de emisiones en quemas

Para visualizar de mejor manera el funcionamiento de esta sección de la interfaz de usuario, se puede observar el [video](#) publicado en Youtube. En él se muestra todo el proceso desde el ingreso de la imagen hasta la obtención de las emisiones en ella.

### 5.3.2. Registro temporal

La otra sección de la interfaz de usuario corresponde a la visualización del registro temporal de Chile. Esta parte de la aplicación web tiene como fin que el usuario pueda seleccionar la zona del país y el rango de tiempo en el cual quiere visualizar las emisiones por quemas forestales y agrícolas.

El primer rasgo de esta sección de la interfaz es una caja de selección de la región a analizar, la cual se se ve encerrada en verde en la Figura 5.18. La opciones posibles son las 16 regiones de Chile y 'Todo el País', dependiendo de la alternativa seleccionada se muestran datos para esa región. Una vez que se selecciona la región, se puede optar a visualizar los datos de una comuna en específico o ver los datos de toda la región en la caja de selección encerrada en rojo en la Figura 5.18.

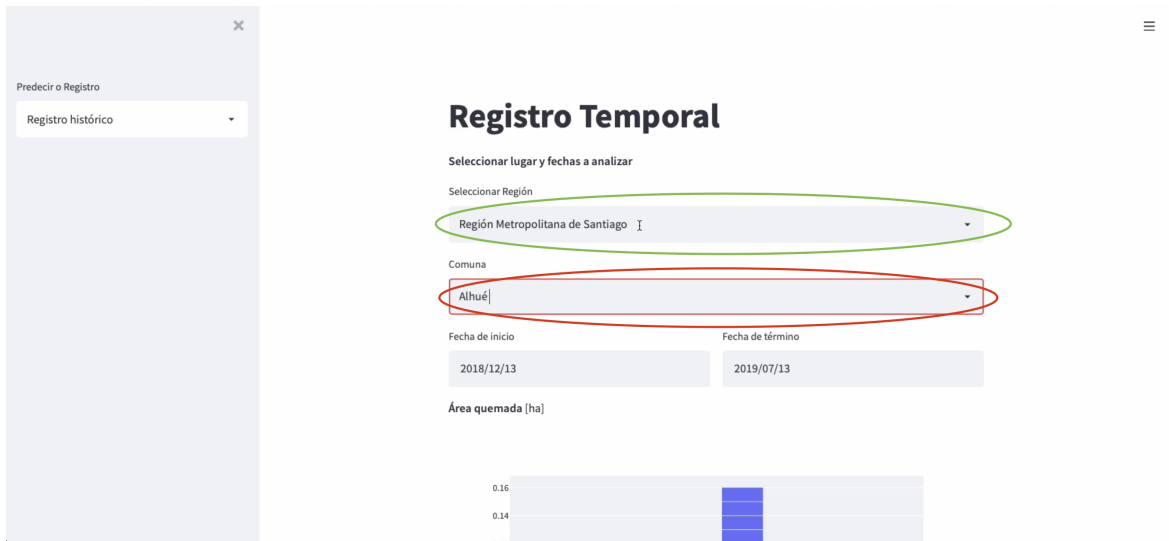


Figura 5.18: Selección de región y comuna en la interfaz de usuario

Una vez que se selecciona la zona de interés, se puede escoger el rango de fechas para la visualización de los datos. Primero, se tiene la caja de selección de la fecha de inicio, la cual está encerrada en verde en la Figura 5.19. Luego, se puede seleccionar la fecha de término en la casilla encerrada en rojo de la Figura 5.19. Es importante notar que las fechas están restringidas al registro, por lo que el mayor rango de fechas posible está entre el 13 de Diciembre de 2018 y el 26 de Junio de 2021.

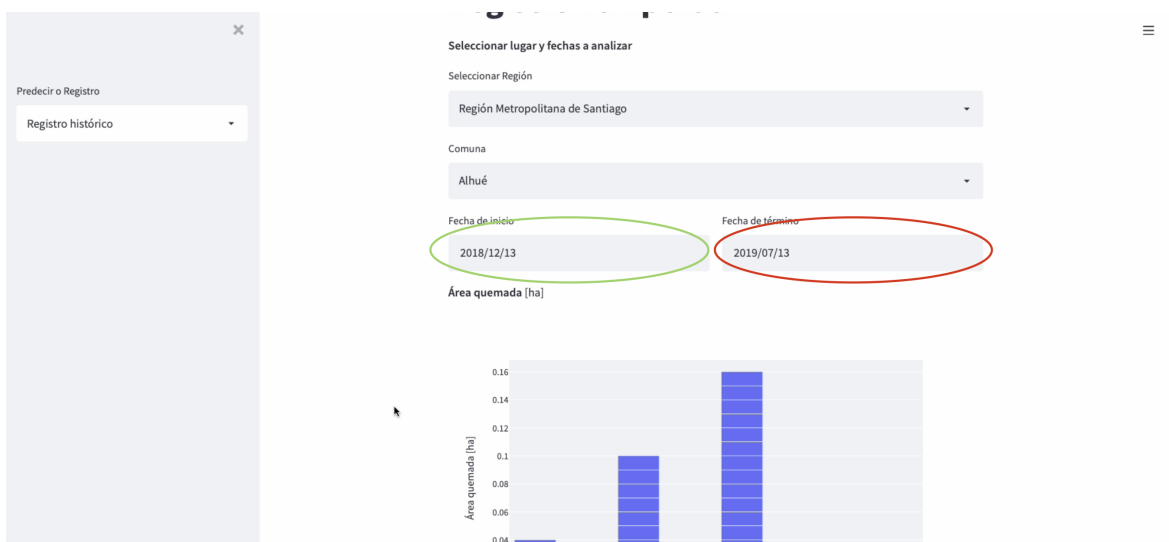


Figura 5.19: Selección de la fecha de inicio y la fecha de término de interés en la interfaz de usuario

Una vez que el usuario selecciona la zona y las fechas de interés se pueden visualizar los resultados de área quemada como se muestra en la Figura 5.20. Esta visualización corresponde a un gráfico de barras en donde se suman todas las áreas quemadas en cada comuna, según la fecha en que se tomó la imagen satelital utilizada para estimar ésta.



Figura 5.20: Visualización del gráfico de área quemada en la interfaz de usuario

Además, es importante conocer la procedencia de las quemas y las emisiones, por lo que la interfaz entrega los porcentajes de quemas de bosque templado, matorrales, pastizales y cultivos que se identifican. La visualización de los porcentajes de tipos de usos de suelos identificados como área quemada se puede apreciar en la Figura 5.21.



Figura 5.21: Visualización de porcentajes de tipos de uso de suelo quemados y casilla de multi-selección de gases de interés en la interfaz de usuario

Luego, para poder visualizar el registro de emisiones, el usuario debe seleccionar el o los contaminantes de interés. Es por esto que la interfaz tiene una casilla de multi-selección en donde aparecen los 15 gases y partículas mencionados anteriormente, la cual se aprecia encerrada en rojo en la Figura 5.21. El usuario puede escoger uno o más de los gases y partículas disponibles y se desplegarán gráficos de barra de las emisiones para cada uno de los gases seleccionados, como el observado en la Figura 5.22.



Figura 5.22: Ejemplo de visualización de emisiones y botón de descarga de datos de la interfaz de usuario

Finalmente, de forma adicional, el usuario puede descargar los datos de las emisiones seleccionadas utilizando el botón encerrado en rojo de la Figura 5.22. Una vez que este botón es presionado se descarga un archivo CSV que contiene los datos de fecha, latitud central, longitud central, área quemada y las emisiones en gramos para la localidad, las fechas y los gases seleccionados. De esta forma el usuario puede acceder al registro con una mayor resolución espacial y analizar los datos con mayor profundidad si lo desea.

Para observar de mejor manera el funcionamiento de la visualización del registro temporal de emisiones proporcionado en la interfaz de usuario se puede ver el [video](#) publicado en Youtube, en el cual se muestran las opciones de las casillas de selección y un ejemplo del resultado de los datos a descargar.



# Capítulo 6

## Discusión

Como se describe en la sección anterior, este trabajo de memoria presenta 3 principales derivados: estos son: el modelo de estimación de emisiones, el registro temporal y la interfaz de usuario de visualización de datos. A continuación se discuten los resultados obtenidos para cada una de las partes respectivas.

### 6.1. Modelo de Estimación de Emisiones

Se puede apreciar en los resultados que el modelo de estimación de área quemada no es perfecto, es decir, presenta errores de clasificación. Estos errores, como ya se mencionó, corresponden a falsos positivos que clasifican como área quemada una zona sin incendio, y falsos negativos en donde se falla en detectar área quemada.

En cuanto a los falsos positivos, como se observa en las Figuras 5.1, 5.2 y 5.3 estos errores son introducidos principalmente por zonas secas, zonas urbanas cercanas a vegetación y por nubes. Esto tiene razón en que el modelo utilizado, extraído de [23], fue entrenado con datos de incendios en zona con alto nivel de vegetación y sin nubosidad. De esta forma, el modelo tiende a tener un mejor rendimiento en un terreno en que los incendios generan un alto contraste con el entorno, como lo son bosques y zonas que se visualizan con un verde brillante. Es por esto, que tanto la nubosidad como las zonas con menor nivel de vegetación introducen falsos positivos. Sin embargo, estas fallas de clasificación no son de alto nivel e introducen un error de menos del 0.1 % de la imagen, gracias al sistema de umbrales introducido.

Por otra parte, se tienen falsos negativos como se muestra en las Figuras 5.4 y 5.5, en donde el modelo falla en clasificar como área quemada incendios pequeños. Si bien esto hace que se subestimen las emisiones de contaminantes, los incendios no detectados tienden a ser menores al 0.1 %, por lo que se asume que el error introducido no afecta con gran impacto la estimación.

El objetivo general de este Trabajo de Título es lograr generar inventarios de emisiones en incendios forestales y agrícolas a través de la utilización de modelos de Machine Learning. Como se puede ver en los resultados, este objetivo es logrado, dado que al entregar una imagen satelital se puede obtener una estimación de las emisiones en ella, a partir del área quemada identificada, casi en tiempo real.

Para usar este modelo se debe tener en consideración que fue entrenado con imágenes satelitales de 3 canales, utilizando las bandas *Short Wave Infrared*, *Near Infrared* y *Red*, por lo que para su correcto funcionamiento la entrada debe cumplir con estas condiciones. Por otro lado, es importante comprender que la arquitectura UNet utilizada tiene un mejor desempeño en zonas de alta vegetación, lo que se debe a su *dataset* de entrenamiento. Finalmente, es importante mencionar que zonas de baja vegetación o zonas con alto nivel de nubosidad introducen ruido, por lo que la clasificación entregada por el modelo puede no ser la óptima.

## 6.2. Registro Temporal de Emisiones

En cuanto al registro temporal de emisiones, una buena forma de discutir sobre él es la comparación con el registro actual de emisiones de Chile de la CONAF [2]. Actualmente, el registro de emisiones de Chile contiene los años hasta el 2019, por lo que se utilizarán solo los datos del 2019 del registro generado para compararlos.

Al analizar ambos registros se puede ver que en regiones en donde existe mayor nubosidad a lo largo del año, es decir, especialmente las regiones del sur, el registro generado por el método de cálculo de emisiones a partir del área quemada tiende a mostrar emisiones mucho menores a las del registro de la CONAF. Esto se debe principalmente, a un déficit de imágenes satelitales dado que fueron filtradas según la nubosidad que presentan. Como se aprecia en la Figura 6.1, el registro de oficial chileno (`co2_ret`) de emisiones de dióxido de carbono son mayores a los estimados por el método propuesto (`co2`) en la comuna de Punta Arenas. En esta comuna llueve una gran cantidad de días, es más, según el reporte climático del año 2019 [32] hubo alrededor de 335 días de precipitaciones, lo que implica un alto nivel de nubosidad por la mayor parte del año. Por su parte, las otras comunas tienen una mayor semejanza entre a las emisiones estimadas con las registradas oficialmente.

Esta sub-estimación también se ve representada en la comuna de Traiguén de la Región de la Araucanía, donde la CONAF reporta la mayor cantidad de emisiones de CO<sub>2</sub>. En esta comuna se tiene una gran diferencia entre los registros, el registro de la CONAF entrega  $3.77496 \times 10^5$  toneladas de CO<sub>2</sub>, mientras que el registro generado entrega  $1.20272 \times 10^2$  toneladas. La diferencia es de 3 órdenes de magnitud, lo que es bastante considerable. Esta diferencia se atribuye a la nubosidad presente en esta comuna. Según el registro de [33], en Traiguén hay como mínimo un 18% de nubosidad constantemente. Es por lo anterior que en esta comuna solo se tienen registrados dos días donde se identifican incendios.

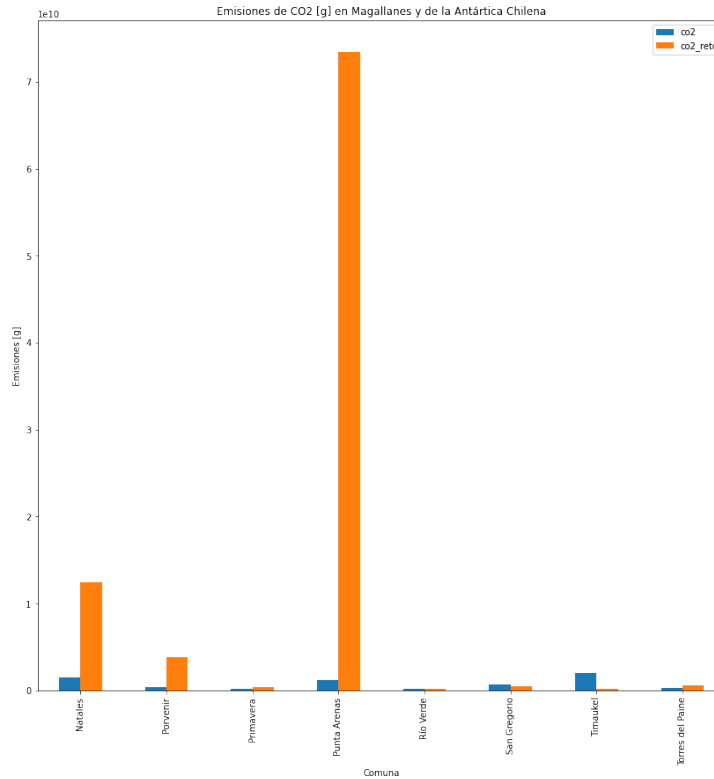


Figura 6.1: Emisiones de CO<sub>2</sub> en la región de Magallanes y la Antártica Chilena según el registro de la CONAF y el método empleado

La sub-estimación de emisiones en sectores del sur del país puede deberse a la baja cantidad de muestras en el registro. Es importante notar que para la descarga de imágenes satelitales se filtra por la cantidad de nubes, usando como cota superior un 10% de nubosidad en la imagen. Es por esto que se cree que en las regiones del sur del país hay una falta de muestras, por ejemplo en la región de Magallanes y la Antártica Chilena solo se registran incendios en 57 días del año 2019. Además, es importante notar que el registro está adherido a la periodicidad de toma de muestras del satélite Sentinel-2, si bien la frecuencia con la que este satélite toma muestras es alta, esta no es diaria y por lo tanto pueden perderse imágenes de incendios en los períodos en que no hay imágenes en el territorio chileno.

Lo anterior muestra el fuerte impacto que tiene la decisión de filtrar las imágenes satelitales según su nubosidad. Al hacer esto lo que se busca es evitar obtener errores de estimación, tanto en falsos positivos como en falsos negativos. Pero de esta misma forma, se eliminan muestras de distintas partes del país, obteniendo así un registro incompleto para zonas con alto nivel de nubes de forma constante. Una posible solución a esto es buscar un porcentaje de nubosidad mayor para filtrar, siempre teniendo en cuenta que la presencia de nubes introduce error, evitando generar un registro demasiado ruidoso. En otras palabras, se debe buscar un nuevo porcentaje de nubes para filtrar las imágenes, evitando tanto introducir un gran nivel de error de estimación de área quemada como descartar una cantidad relevante de muestras para ciertos sectores.

Por otra parte, en regiones más al norte de Chile en donde hay una menor nubosidad promedio, se puede ver que las estimaciones del método propuesto se asemejan más a las

emisiones del registro de la CONAF, al menos en orden de magnitud. Esto se puede observar en la Figura 6.2, en donde el método formulado en este Trabajo de Título entrega estimaciones de emisiones ( $\text{CO}_2$ ) de dióxido de carbono que incluso superan en algunas comunas a las emisiones registradas. Sin embargo, se puede apreciar que los órdenes de magnitud de ambos registros son iguales para las 4 comunas, lo que implica que el modelo de Machine Learning no sobre o subestima en grandes magnitudes el área quemada, y por lo tanto las estimaciones de emisiones de gases no presentan grandes errores en zonas más despejadas.

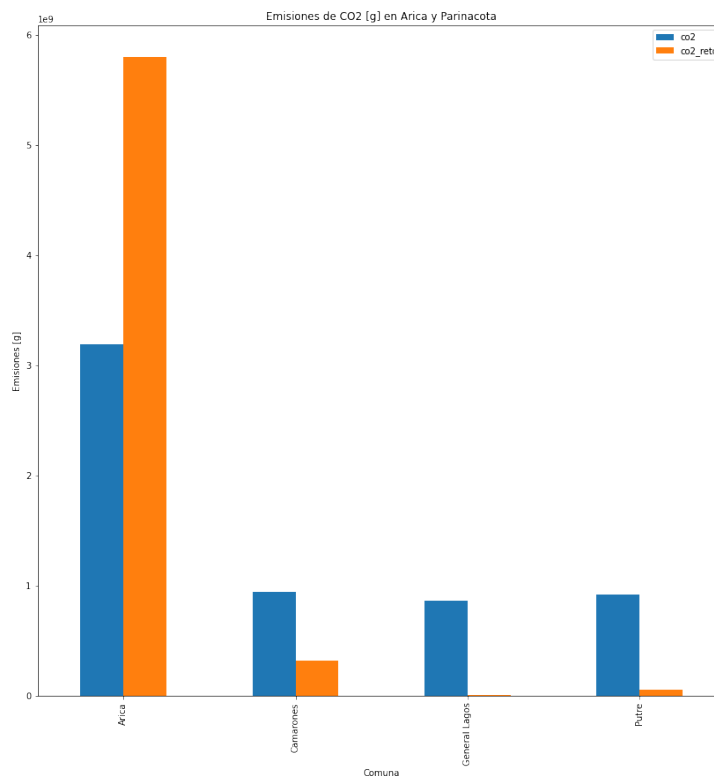


Figura 6.2: Emisiones de  $\text{CO}_2$  en la región de Arica y Parinacota según el registro de la CONAF y el método empleado

Mientras que en regiones más centrales, como la Región de Coquimbo, donde la nubosidad oscila entre 8% y 42% existen comunas donde los registros se asemejan y otras comunas donde se presenta una mayor diferencia. En la Figura 6.3 se puede notar que en la mayoría de las comunas ambos registros se mantienen en el mismo orden de magnitud.

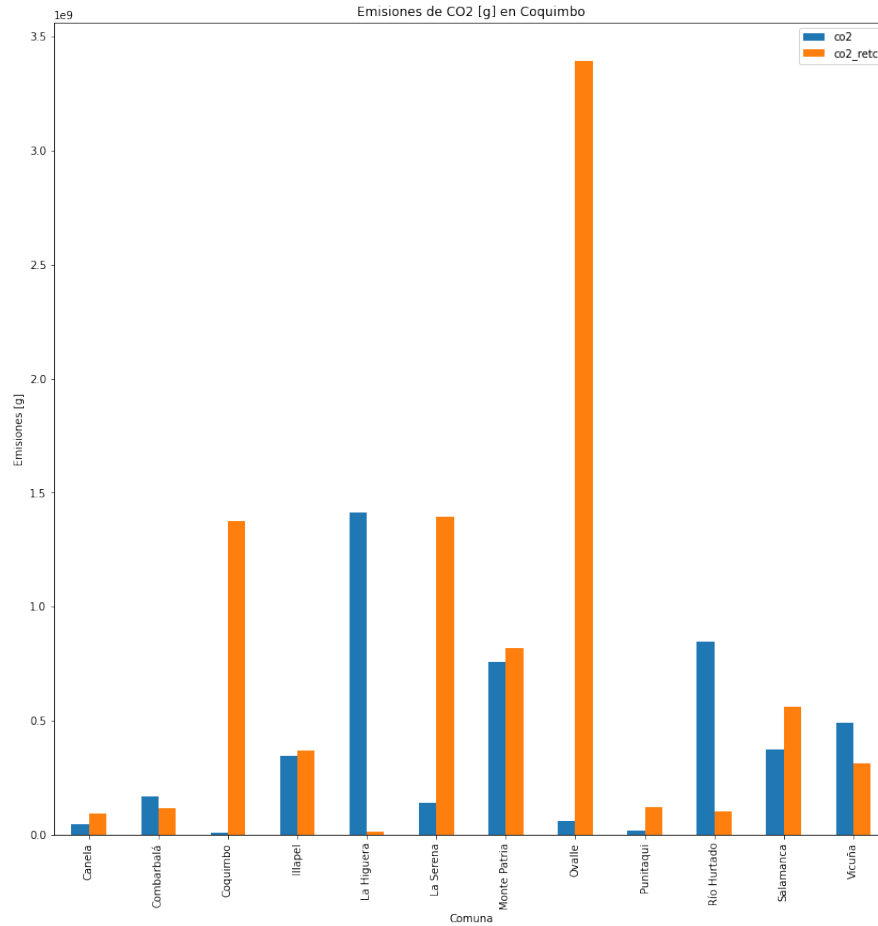


Figura 6.3: Emisiones de CO<sub>2</sub> en la Región de Coquimbo según el registro de la CONAF y el método empleado

Se menciona en los resultados del modelo de Machine Learning para estimar el área quemada que existen falsos positivos en cuerpos de agua, como se visualiza en la Figura 5.3. En este registro se eliminan este tipo de errores con la clasificación de tipo de uso de suelo. Como se muestra en la Tabla 3.1 existe una clasificación de uso de suelo que corresponden a cuerpos de agua, pero al momento de generar el registro se asume que este tipo de uso de suelo no genera emisiones. Esto mismo ocurre con zonas de tierras desnudas, hielos, nieves y en algunos casos nubes. De esta forma, el registro minimiza el error introducido por la identificación de área quemada en zonas donde no pueden haber incendios forestales ni agrícolas.

Es importante observar que, dado que solo se han usado datos del satélite Sentinel-2, el registro está limitado a la toma de muestras de éste. Como se menciona anteriormente en la sección de Adquisición de Datos, Sentinel-2 obtiene imágenes de Chile con alta frecuencia, pero en una misma zona puede existir una latencia de hasta 10 días de toma de muestras, lo que implica que no existe el registro de incendios en estos días a menos de que quede una *burn scar* que cubra un área considerable. Es por esto que para obtener un registro más completo, una posible solución a la falta de muestras es utilizar las muestras de otros satélites adicionalmente, asegurándose de no cubrir las mismas zonas durante el mismo período de tiempo. A pesar de que el agregar muestras de otros satélites podría generar una mejora, se debe considerar que estos tienen, por lo general, una resolución menor a la de los satélites

de Sentinel y por lo tanto no se mantendría la resolución actual, teniendo mayor riesgo de sobre-estimación de área quemada y emisión de gases y partículas.

Por otra parte, al filtrar las imágenes según su porcentaje de nubosidad se obtiene la mayor limitación del registro, como se visualiza en la Figura 6.1. Sin embargo, la nubosidad también introduce error en la estimación de área, por lo que el utilizar imágenes con un mayor nivel de nubes puede ser perjudicial para el registro temporal, identificando área quemada donde no la hay. Es por esto que se cree que, si con el fin de obtener un mejor registro en las zonas de alta nubosidad de Chile, se utilizan muestras con un mayor porcentaje de nubes, se debe utilizar un porcentaje en donde la cantidad de falsos positivos introducidos no sea tan importante que afecte las órdenes de magnitud de la estimación de las emisiones.

Los aportes que entrega este registro recae en su resolución tanto temporal como espacial. El registro actual se entrega de forma anual y por comuna, mientras que el registro generado tiene una resolución de  $100 \text{ m}^2$  y le asigna un día a la emisión de gases y partículas por quemas. Con este registro se puede localizar de forma mucha más certera las emisiones, teniendo un mayor conocimiento de éstas en diferentes puntos y por lo tanto de las emisiones a las que se ven expuestas localidades viento abajo. Además, se puede entregar la fecha en son expuestos a las emisiones, por lo que en el caso de que el registro se mantenga actualizado se pueden tomar medidas de forma más rápida en las localidades cercanas a quemas y expuestas a gases de efecto invernadero. Incluso, si se mantiene una descarga constante de imágenes satelitales, se podría identificar incendios en las distintas zonas y tomar acciones rápidas para extinguirlos.

### 6.3. Interfaz de Usuario

La aplicación web generada en este Trabajo de Título busca ser una interfaz de usuario preliminar, en donde se visualicen los resultados de la inferencia del modelo de Machine Learning y las estimaciones de emisiones de gases de efecto invernadero. Es por esto que se divide en dos partes, donde el usuario pueda ingresar una imagen en particular y obtener las emisiones en ella, pero también pueda ver resultados históricos del modelo en Chile.

En cuanto a la primera sección de la aplicación, en la cual el usuario puede ingresar una imagen satelital propia extraída de Sentinel u otro satélite ([34]), se cumple el objetivo de que se visualice la inferencia realizada por el modelo, además de que se entreguen las emisiones calculadas. Sin embargo, se debe tener en cuenta que en esta sección el usuario solo puede ingresar un único tipo de uso de suelo, por lo que si hay más de un incendio o área quemada en la imagen se asume que ambos tienen igual tipo de vegetación. Es importante notar que esta interfaz no cumple su propósito si el usuario no entrega una imagen apta, es decir, utilizando las bandas en el orden específico, dado que la inferencia no será certera. A pesar de esto, la imagen de entrada no está limitada en tamaño por el pre-procesamiento realizado. El fin de esta parte de la interfaz es que el usuario, sin conocer el modelo y sin necesidad de programar pueda obtener un registro de las emisiones en la imagen que ingrese.

Por otra parte, la interfaz de registro temporal de emisiones busca que el usuario pueda tener conocimiento de las emisiones en la región y espacio temporal que sea de su interés. Sin embargo, es importante mencionar que se está utilizando el registro generado con las

imágenes de Sentinel-2 por lo que tanto las zonas como las fechas están limitadas a la toma de muestra de éste satélite y al filtro de nubosidad aplicado en la recolección de datos.

Como se menciona anteriormente, esta aplicación web busca ser una interfaz preliminar y por lo tanto se le pueden agregar distintas funcionalidades con el fin de mejorarla. Una posible mejora podría ser que el usuario ingrese las latitudes y longitudes mínimas de la imagen en conjunto con la fecha de toma de muestra, para así poder geo-referenciar los píxeles de área quemada y entregar un archivo CSV con las emisiones por píxel. En forma complementaria, podría agregarse una casilla para ingresar archivos CSV en la interfaz de registro, de forma de que el usuario ingrese datos al registro. Sin embargo, se decide no integrar estas funcionalidades, ya que esto podría generar inconsistencias en el registro si el usuario ingresa algún dato erróneo.

Una funcionalidad que mejoraría de forma importante la aplicación web de visualización del registro sería el que esté conectada a un registro dinámico, el cual se actualice cada vez que se el satélite entregue una nueva muestra en el país. Sin embargo, esta funcionalidad no está dentro del alcance de la generación de inventarios de emisión, ya que requiere una conexión constante a algún satélite.

La aplicación web resultante, como ya se mencionó es creada con Streamlit. Este paquete de Python genera interfaces simples, tal como la mostrada en los resultados. Si bien, la interfaz no es susceptible a distintas mejoras estéticas, para el alcance del proyecto, la utilización de este paquete muestra ser suficiente para cumplir con las características deseadas. Lo anterior se basa en que la interfaz muestra funcionalidades de uso simple para el usuario. El principal aporte de esta aplicación es su simplicidad de uso, permitiendo a usuarios no expertos obtener las emisiones en sus imágenes satelitales y visualización de un registro histórico de alta resolución temporal de emisiones.

# Capítulo 7

## Conclusiones

En el Trabajo de Título presentado se trata el problema de la generación de inventarios de emisiones, presentado por la baja frecuencia temporal que se tiene en el registro chileno actual. Luego, se describe el proceso para la recolección de datos y la metodología empleada para la estimación de área quemada y emisiones de gases y partículas, la generación de un registro temporal de emisiones en Chile utilizando el método propuesto y la creación de una interfaz para que usuarios puedan visualizar los resultados del método propuesto. Una vez descrita la metodología de desarrollo del proyecto, se presentan los resultados obtenidos con el modelo de estimación en imágenes, el resultado de aplicar este modelo y estimar las emisiones sobre los datos recolectados para la generación del registro temporal. Por otra parte, se muestran las funcionalidades de la interfaz creada y su funcionamiento. Finalmente, una vez presentados los resultados se discute sobre ellos, sus posibles mejoras y los principales aportes que éstos entregan.

El objetivo principal de este proyecto es la generación de inventarios de emisión de quemas forestales y agrícolas a partir de imágenes satelitales y la utilización de un modelo de Machine Learning. Este objetivo se logra gracias a la utilización y adaptación de un modelo generado para clasificar los píxeles de una imagen satelital como área quemada, entregando las estimaciones de forma casi instantánea una vez ingresada la entrada. A partir de la estimación de área quemada del modelo de Machine Learning, se utiliza el método presentado en [16] para calcular las emisiones en las imágenes satelitales que se le entregan como entrada al modelo. Además, se recolectan datos de vegetación e imágenes satelitales durante un período extenso del territorio chileno, los cuales junto con el uso del método propuesto, son utilizados para crear un registro de alta resolución temporal y espacial de las emisiones por incendios en Chile. Finalmente, se genera una aplicación web que funciona como interfaz entre el método y el usuario, facilitando a éste último la visualización tanto de la estimación de emisiones del modelo ante una nueva entrada, como los resultados obtenidos para la base de datos recolectada.

El principal aporte del Trabajo de Título descrito es el desarrollo de un método que entrega de forma casi instantánea la estimación de emisiones a partir de imágenes satelitales, aumentando la resolución tanto temporal como espacial del registro actual. Como se menciona anteriormente, el registro actual de emisiones mantiene una frecuencia anual, la cual es baja dadas las condiciones actuales. De esta forma el método descrito entrega una herramienta para mantener un inventario actualizado y de alta frecuencia, dado que solo se requiera una imagen satelital para obtener la estimación de las emisiones. Actualmente, existe un gran



cantidad de satélites que obtienen imágenes de Chile de forma diaria, por lo que al utilizar este método se puede obtener un registro de emisiones de alta frecuencia temporal y espacial. Con este registro se podrían tomar acciones más rápidas sobre incendios y localidades que han sido expuestas a la emisión de sus contaminantes. Por otra parte, la generación de la interfaz facilita la utilización del método propuesto y visualización de resultados, con el fin de que éstos sean claros y el usuario pueda interpretarlos.

Los inventarios de emisiones pueden utilizarse con un objetivo informativo, es decir, para mantener los registros de las emisiones y poder entregar esta información al público o a una institución. Pero, esta información también puede ser utilizada para desarrollar estrategias o políticas públicas en base a ella. Por otra parte, este tipo de inventarios se utilizan para modelar y monitorear la calidad de aire. Es por esto que es importante mantener un registro actualizado y de alta resolución, con el fin de poder generar políticas públicas y medidas correspondientes, no basándose en información anticuada que puede no aplicar a la actualidad. Por otra parte, para que el monitoreo de calidad de aire esté al día, se necesita tener la mayor cantidad de información sobre las emisiones posible hasta la fecha.

Si bien el trabajo desarrollado cumple con los objetivos presentados, existen oportunidades de mejora con las cuales se podría mantener un registro completo de las emisiones en el territorio chileno. Un posible perfeccionamiento del proyecto sería realizar *fine-tuning* al modelo de estimación de área quemada utilizando imágenes de Chile, clasificadas por un experto previamente. Por otra parte, en cuanto al registro, para que este sea más completo se podría complementar la base de datos con imágenes de otros satélites, sin limitarse a la frecuencia de toma de muestras de Sentinel-2. Para obtener un registro lo más actualizado posible una propuesta de mejora es el generar una conexión, de tal forma que cada vez que un satélite entregue una imagen nueva del territorio del país, esta sea ingresada, se le asigne un tipo de uso de suelo y el modelo estime su área quemada y se calculen sus emisiones. Finalmente, se nota que la clasificación de uso de suelo es más específica que las utilizadas en el método de cálculo de emisiones, por lo que se podría encontrar un método que explote de mejor manera esta clasificación.

# Bibliografía

- [1] CONAF, “Estadísticas históricas,” 2020.
- [2] M. del Medio Ambiente, “Retc,” 2020.
- [3] CONAF, “Detección de incendios forestales,” 2020.
- [4] N. S. Khadka, “Climate change: Worries over co2 emissions from intensifying wildfires,” 2018.
- [5] M. Black and Y. Yacoob, “Tracking and recognizing rigid and non-rigid facial motions using local parametric models of image motion,” in *Proceedings of IEEE International Conference on Computer Vision*, pp. 374–381, 1995.
- [6] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [7] M. Mishra, “Convolutional neural networks, explained,” 2020.
- [8] S. Bekesiene, R. Smaliukiene, and R. Vaicaitiene, “Using artificial neural networks in predicting the level of stress among military conscripts,” *Mathematics*, vol. 9, p. 626, Mar. 2021.
- [9] W. H. L. Pinaya, S. Vieira, R. Garcia-Dias, and A. Mechelli, “Convolutional neural networks,” in *Machine Learning*, pp. 173–191, Elsevier, 2020.
- [10] K. H., X. Z., S. R., and J. S., “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.
- [11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [12] R. O., F. P., and B. T., “U-net: Convolutional networks for biomedical image segmentation,” in *Lecture Notes in Computer Science*, vol. 9351, 2015.
- [13] H. Sankesara, “Unet, introducing symmetry in segmentation,” 2019.
- [14] Iberdrola, “Efecto Invernadero las consecuencias del efecto invernadero: desde la desertificación a las inundaciones,” 2019.
- [15] K. J. W., H. A., A. M. O., B. A., C. N., J. L., M. J.J., R. M., S. M. G., S. M., and van der Werf G., “Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power,” in *Biogeoscience*, vol. 9, pp. 527 – 554, 2012.
- [16] W. C., A. S.K, Y. R. J., E. L.K., A.-S. J. A., O. J.J., and S. A.J., “The fire inventory

- from near (finn): a high resolution global model to estimate the emissions from open burning,” in *Geoscientific Model Development*, vol. 4, pp. 625 – 641, 2011.
- [17] G. S., “Wildfires in california spotted by sentinel-2a,” 2016.
- [18] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [19] Gorelick, Noel, Hancher, Matt, Dixon, Mike, Ilyushchenko, Simon, Thau, David, Moore, and Rebecca, “Google earth engine: Planetary-scale geospatial analysis for everyone,” *Remote Sensing of Environment*, 2017.
- [20] B. del Congreso Nacional, “Mapas vectoriales,” 2018.
- [21] K. Jordahl, J. V. den Bossche, M. Fleischmann, J. Wasserman, J. McBride, J. Gerard, J. Tratner, M. Perry, A. G. Badaracco, C. Farmer, G. A. Hjelle, A. D. Snow, M. Cochran, S. Gillies, L. Culbertson, M. Bartos, N. Eubank, maxalbert, A. Bilogur, S. Rey, C. Ren, D. Arribas-Bel, L. Wasser, L. J. Wolf, M. Journois, J. Wilson, A. Greenhall, C. Holdgraf, Filipe, and F. Leblanc, “geopandas/geopandas: v0.8.1,” July 2020.
- [22] Z. Y., F. D., Y. L., W. X., C. Y., B. Y., H. H., G. M., E. C., B. G., R. J., and G. P., “Detailed dynamic land cover mapping of chile: Accuracy improvement by integrating multi-temporal data,” *Remote Sensing of Environment*, 2016.
- [23] D. An, B. Han, and Y. Y. Zhijuan Zhang, “Wildfire detection proyect,” 2020.
- [24] T. Lark, S. Spawn, and M. Bougie, “Cropland expansion in the united states produces marginal yields at high costs to wildlife,” in *Nat Commun*, vol. 11, 2020.
- [25] G. Bradski, “The OpenCV Library,” *Dr. Dobb’s Journal of Software Tools*, 2000.
- [26] W. McKinney *et al.*, “Data structures for statistical computing in python,” in *Proceedings of the 9th Python in Science Conference*, vol. 445, pp. 51–56, Austin, TX, 2010.
- [27] GDAL/OGR contributors, *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation, 2021.
- [28] A. D. Snow, “rioxarray.” <https://pypi.org/project/rioxarray/>.
- [29] S. Gillies *et al.*, “Shapely: manipulation and analysis of geometric objects,” 2007.
- [30] J. Roes, K. McGrady, A. Treuille, and T. Teixeira, “streamlit.” <https://pypi.org/project/streamlit/>.
- [31] L. Giglio, L. Boschetti, D. Roy, M. Humber, and C. Justice, “The collection 6 modis burned area mapping algorithm and product,” Nov 2018.
- [32] S. de Climatología y Meteorología Aplicada, “Reporte climático,” 2019.
- [33] W. Spark, “El clima y el tiempo promedio en todo el año en traiguén,” 2019.
- [34] E. Engine, “Earth engine data catalog,” 2021.

# Anexos

## Anexo A

### Satélites

Existen distintos satélites de los cuales se pueden extraer imágenes y productos. En particular aquellos que pasan de forma regular en Chile son Landsat, Sentinel y MODIS. En particular, para el Trabajo de Título desarrollado se considera usar dos satélites y finalmente se selecciona Sentinel, por lo que se describirán ambos en conjunto con sus bandas.

#### A.1. Sentinel

Los satélites de la flota Sentinel son propiedad del programa Copernicus. En particular el satélite Sentinel-2 tiene una cámara multi-espectral que entrega 13 bandas con hasta 10 metros de resolución. Este satélite se utiliza para monitorizar vegetación, tierra y agua, observar áreas costales y vías navegables interiores. A continuación se describen algunas de las bandas y sus resoluciones.

- **B1:** Esta banda representa los aerosoles en el ambiente con una resolución de 60 metros. Esta banda está dentro del rango de los 443.9 y los 442.3 nanómetros.
- **B2:** La banda B2 representa el color azul con una resolución de 10 metros, conteniendo las longitudes de onda entre 496.6 y 492.1 nanómetros.
- **B3:** Esta banda representa con una resolución de 10 metros el color verde, que se encuentra entre las longitudes de onda 560 y 559 nanómetros.
- **B4:** Al igual que las bandas anteriores, esta tiene una resolución de 10 metros, pero ahora representando el color rojo. Las longitudes de onda de esta banda se encuentran entre 664.5nm y 665nm.
- **NIR:** Esta banda representa las longitudes de onda entre los 835.1nm y los 833nm con una resolución de 10 metros. Las longitudes de onda en este rango representan los colores cercanos al infrarrojo.
- **SWIR 1:** Esta banda representa, con una resolución de 20 metros, las señales de onda corta cercanas al infrarrojo (*Short Wave Infrared*) con longitudes de onda entre 613.7nm y 1610.4nm.
- **SWIR 2:** Esta banda es muy similar a la anterior, pero ahora se representan este tipo de señales pero en el rango entre 2202.4nm y 2185.7nm.

## A.2. MODIS

Este proyecto obtiene sus productos a partir de la combinación de los satélites Aqua y Terra. En particular un producto de interés para el proyecto fue *MODIS Burned Area Monthly Global 500m*, el cual como dice su nombre, entrega de forma mensual un producto de área quemada con una resolución de 500.

Si bien este producto podría simplificar el proceso de estimación de emisiones, evitando estimar a través de un modelo de Machine Learning el área quemada, su resolución temporal es baja, por lo que no cumpliría con el objetivo principal. Por otra parte, dada su resolución espacial, puede ocurrir una sobre-estimación de área quemada. Lo anterior se basa en que tiene una resolución baja, por lo que dentro de uno de sus píxeles puede existir un incendio, pero no necesariamente toda el área en este está quemada, sin embargo, todo el píxel se clasifica como área quemada.

A diferencia del producto de Sentinel-2, éste solo contiene 2 bandas relevantes.

- **BurnDate:** Esta banda representa la fecha del año en que se identifica el incendio o quema en el píxel. Su rango está entre 0 y 366, donde 0 representa en que no se encuentra incendio ningún día del año.
- **Uncertainty:** Esta banda está relacionada a la anterior, representando la incertidumbre que se tiene con respecto a la fecha de que se identifica una quema.

# Anexo B

## Gases y Partículas

Durante el proyecto se habla, principalmente, de 15 gases de efecto invernadero. Para comprender más sobre las emisiones estimadas a continuación se describen los gases tratados.

### B.1. Dióxido de Carbono ( $\text{CO}_2$ )

Este compuesto está formado por carbono y oxígeno, que existe como gas incoloro en condiciones de temperatura y presión estándar. El  $\text{CO}_2$  atmosférico es la principal fuente de carbono para la vida en la Tierra. Su concentración, antes de la industrialización era regulado por organismos que realizan fotosíntesis.

Este gas aporta al efecto invernadero dado que absorbe y emite radiación infrarroja. Si bien no solo este gas incrementa el efecto invernadero, el aumento de su emisión luego de la industrialización aporta de forma importante a este fenómeno.

### B.2. Monóxido de Carbono ( $\text{CO}$ )

Este gas está, al igual que el  $\text{CO}_2$ , esta compuesto por carbono y oxígeno, siendo incoloro y altamente tóxico. Este gas puede causar la muerte cuando se está expuesto a niveles elevados.

Este gas se produce por la combustión de sustancias como gas, gasolina, carbón, petróleo, tabaco y madera. Existen distintas fuentes de  $\text{CO}$ , como por ejemplo vehículos, chimeneas, calderas, entre otros.

### B.3. Metano ( $\text{CH}_4$ )

El metano es un gas natural, incoloro y sin olor que se produce debido a la descomposición de materia orgánica, como las plantas. Este gas está compuesto por carbono e hidrógeno. Una de las propiedades más importantes del metano es que produce calor y luz en su emisión.

Si bien este gas no es tóxico, su principal peligro para la salud son las quemaduras que puede provocar si este está expuesto a combustión. Es altamente inflamable y puede formar mezclas explosivas con el aire.

## B.4. Dihidrógeno ( $H_2$ )

Este gas está compuesto únicamente por hidrógeno. Este compuesto de dos moléculas es incoloro, inodoro e inflamable a temperatura ambiente. El dihidrógeno se emplea en la producción de amoníaco, como combustible alternativo y o como suministro de energía en las pilas de combustible.

El dihidrógeno no es tóxico y está clasificado como un simple asfixiante. Sin embargo, es inflamable, por tanto, se prohíbe su entrada en áreas que contengan mezclas inflamables debido al peligro inmediato de incendio o explosión. El dihidrógeno se puede absorber por las vías respiratorias y a través de la piel.

## B.5. Gases Reactivos que Contienen Nitrógeno y Oxígeno ( $NO_x$ )

$NO_x$  es un término genérico para gases muy reactivos compuestos por nitrógeno y oxígeno. Muchos de los óxidos de nitrógeno son incoloros e inodoros en condiciones estándar. Este tipo de gases son emitidos por combustión, por lo que sus fuentes principales son vehículos, las centrales eléctricas y los incendios.

En la atmósfera, los óxidos de nitrógeno pueden contribuir a la formación de smog y tener consecuencias para la salud, en particular para aquellos seres vivos expuestos frecuentemente. También contribuye al aumento de la temperatura global promedio y puede provocar lluvia ácida.

## B.6. Dióxido de Nitrógeno( $NO_2$ )

El dióxido de nitrógeno es de color café-amarillo. Es emitido en los procesos de combustión a altas temperaturas, como en los vehículos motorizados y las plantas eléctricas. Por ello es un contaminante frecuente en zonas urbanas y rurales.

Este gas es tóxico e irritante, dado que genera partículas de nitrato. De esta forma, su emisión lleva a la producción de ácido y altos niveles de partículas finas en el ambiente. Este gas afecta principalmente al sistema respiratorio irritando las vías aéreas.

## B.7. Compuestos Orgánicos No Metánicos (NMOC)

Los compuestos parte de NMOC por lo general contiene contaminantes de aire peligrosos, tanto así que algunos de ellos están asociados a la reducción de la capa de ozono.

La mayor cantidad de emisiones de NMOC se atribuye a la volatilización de compuestos orgánicos contenidos en residuos de vertedero.

## B.8. Hidrocarburos sin Metano (NMHC)

Los hidrocarburos son sustancias químicas producidas en la naturaleza que están formados carbono e hidrógeno. Los compuestos orgánicos no metánicos corresponden a hidrocarburos que no contiene metano.

Este gas es emitido por combustión, por lo que sus fuentes son incendios, industrias, vehículos, entre otros. Este gas influye en el calentamiento global dado que retiene la radiación.

## B.9. Dióxido de Azufre (SO<sub>2</sub>)

Como lo dice su nombre, este compuesto está formado por oxígeno y azufre, el cual es incoloro con un característico olor irritante. El dióxido de azufre es el principal causante de la lluvia ácida ya que en la atmósfera es transformado en ácido sulfúrico.

Este gas es liberado principalmente a través de la combustión de petróleo, diésel o de gas natural contienen ciertas cantidades de compuestos azufrados. Este gas es tóxico y afecta sobre todo las mucosidades y los pulmones, provocando ataques de tos. La exposición de altas concentraciones por cortos períodos de tiempo puede irritar el tracto respiratorio, causar bronquitis y congestionar los conductos bronquiales de los asmáticos.

## B.10. Amoníaco (NH<sub>3</sub>)

El amoníaco está formado por nitrógeno e hidrógeno, es un gas incoloro con un característico olor fuerte. Este gas puede producirse naturalmente por descomposición de la materia orgánica o se puede fabrica industrialmente.

El amoníaco en el ambiente no afecta a personas ni mamíferos, ya que existe un mecanismo para que sea excretado del cuerpo. Sin embargo, los peces y anfibios no tiene este mecanismo, por lo que es tóxico para ellos. Incluso en concentraciones diluidas este compuesto afecta a los animales acuáticos, por lo que clasifica como peligroso para el ambiente.

## B.11. Partículas Finas (PM<sub>2.5</sub>)

La materia particulada 2.5 corresponde a partículas finas suspendidas en el aire que tienen un diámetro de menos de 2.5 micras. PM<sub>2.5</sub> incluye partículas de sustancias químicas orgánicas, polvo, hollín y metales. Esta materia puede provenir de toda clase de combustión, como quemas de madera o quemas agrícolas.

Estas partículas son consideradas en contaminante por lo nocivas que son para la salud. El tamaño de estas partículas les permite desplazarse profundamente dentro de los pulmones respirando. Es más, su inhalación causa efectos serios en la salud incluyendo enfermedades cardíacas y pulmonares.



## **B.12. Materia Particulada Total (TPM)**

Se denomina material particulado a una mezcla de partículas líquidas y sólidas, de sustancias orgánicas e inorgánicas, que se encuentran en suspensión en el aire.

Estas partículas son nocivas para salud de los seres vivos que han sido expuestos frecuentemente. Por ejemplo, las personas que han inhalado con frecuencia material particulado tienen mayor riesgo de desarrollar enfermedades cardiovasculares, respiratorias y cáncer de pulmón

## **B.13. Carbono Orgánico (OC)**

El carbono orgánico se genera principalmente de la descomposición de materia orgánica o como resultado del metabolismo de los seres vivos. La cantidad de carbono orgánico es un indicador de la calidad de agua.

Si bien no se conocen efectos adversos del OC en los seres humanos, elevadas concentraciones de éste en aguas superficiales, pueden generar una disminución importante en el oxígeno disuelto. De esta forma, la emisión excesiva de este gas tiene como consecuencia la pérdida de biodiversidad marina.

## **B.14. Carbono Negro (BC)**

El carbono negro es parte de las partículas finas PM<sub>2.5</sub>. Éste consta de carbono puro en varias formas enlazadas. Este gas es emitido a través de la combustión incompleta de combustibles fósiles y biomasa.

El carbono negro es un agente que contribuye al calentamiento global, de forma directa e indirecta. El BC absorbe la luz solar, permaneciendo en la atmósfera durante varios días o semanas, reteniendo radiación y aumentando la temperatura global promedio.

# Anexo C

## Sistemas de Coordenadas Geográficas

Como se aprecian en el Trabajo de Título realizado existen diferentes sistemas de coordenadas geográficas. En particular las 2 presentadas en la metodología y adquisición de datos son las coordenadas en grados y el sistema *Universal Transverse Mercator* (UTM).

### C.1. Latitud y Longitud en Grados

Estas coordenadas geográficas se utiliza para referenciar cualquier ubicación en la Tierra a, partir de los valores de su posición horizontal. Para obtener la posición horizontal se requieren las coordenadas de latitud y longitud. Este sistema tiene coordenadas angulares esféricas cuyo centro es el centro de la Tierra y se expresan en grados sexagesimales, las cuales se pueden apreciar en la Figura C.1.

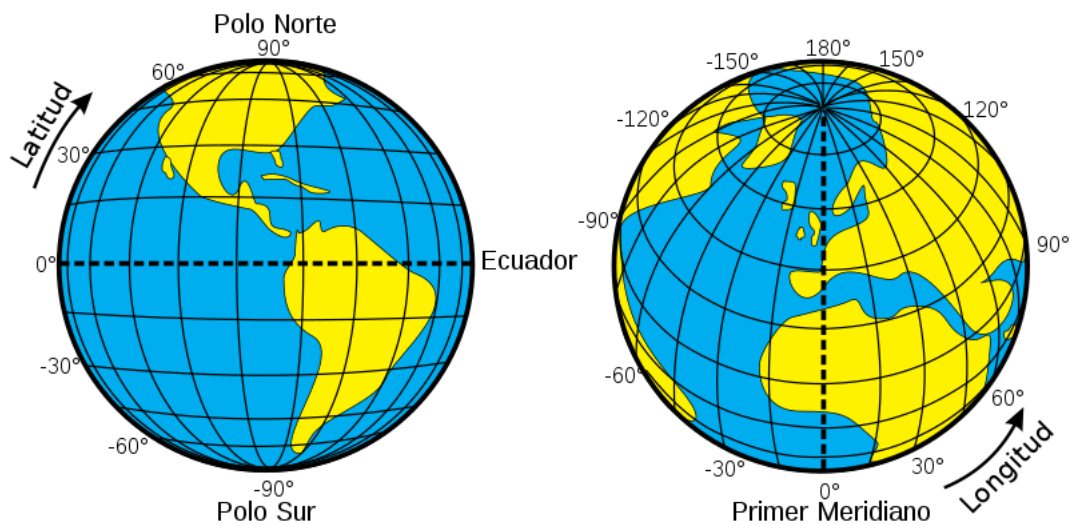


Figura C.1: Sistema de coordenadas Geográficas

#### C.1.1. Latitud

La latitud es el ángulo entre el plano ecuatorial y la línea que pasa por este punto y el centro de la Tierra. Los puntos que tienen el mismo valor de latitud forman un plano paralelo a la línea del ecuador.

La línea del ecuador corresponde al paralelo  $0^\circ$  y divide el planeta en los hemisferios norte y sur. El punto más lejano al ecuador tiene latitud  $90^\circ$ , mientras que el punto más lejano en el hemisferio sur tiene latitud  $-90^\circ$ .

### C.1.2. Longitud

La longitud es la medida del ángulo entre el meridiano de Greenwich y el meridiano que pasa por el punto de interés. Las líneas de longitud forman semi-círculos que pasan por los polos.

El meridiano reflejado del meridiano de Greenwich tiene un valor de  $180^\circ$ . Al observar la tierra desde la latitud  $90^\circ$ , desde el meridiano de Greenwich hasta su contrario, siguiendo el sentido de las manillas del reloj, las longitudes crecen hasta los  $180^\circ$ . Por otra parte, al repetir este proceso en el sentido contrario, las longitudes decrecen desde los  $0^\circ$  hasta los  $-180^\circ$ .

## C.2. Universal Transverse Mercator

El sistema de coordenadas UTM divide el mundo en 60 zonas de norte a sur, las que se observan en la Figura C.2, cada uno con 6 grados de ancho en longitud. Cada zona tiene un número consecutivo empezando en el número 1. Dentro de cada zona las coordenadas se miden según norte y este en metros. Cada zona tiene un meridiano central en donde el valor de este es de 500.000 metros.

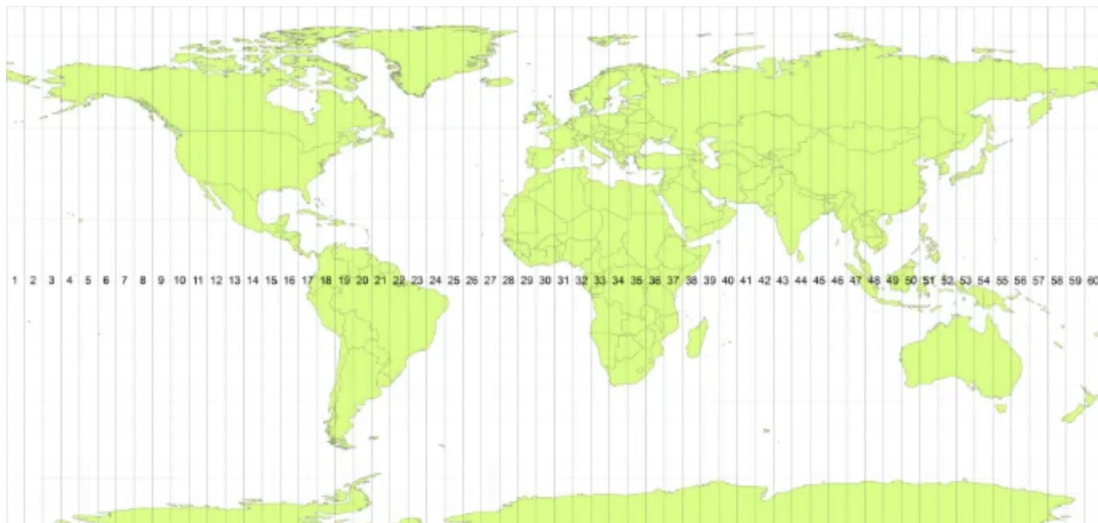


Figura C.2: Ilustración de las zonas UTM

# Anexo D

## Funciones Utilizadas

A continuación se muestran los códigos de algunas de las funciones utilizadas según la sección de la metodología en que se requieren.

### D.1. Pre-procesamiento

Para ajustar las imágenes y transformarlas a un formato apto para ser ingresadas a la red neuronal UNet utilizada, se utilizan funciones de pre-procesamiento.

#### D.1.1. Añadir Margen

La primera función utilizada agrega un margen de valores nulos a la imagen con el fin de que su tamaño permita dividirla en partes de  $480 \times 720$  píxeles.

```
1 import numpy as np
2
3 def add_image_margin(input_image, im_width=720, im_height=480, color=0):
4     # padd
5     height, width, vec = input_image.shape
6     new_image = np.pad(input_image, ((0, im_height-height), (0, im_width-width),
7     (0, 0)), 'constant', constant_values=color)
8     return new_image
```

#### D.1.2. Dividir Imagen

La segunda función se utiliza para dividir la imagen de entrada, de forma que sus partes tengan el tamaño necesario para ser ingresados a la red neuronal. Esta entrega una lista con las secciones de  $480 \times 720$  de la imagen original y las medidas de ésta.

```
1 def preprocess_input_image(input_image, im_height=480, im_width=720):
2     # normalize the pixel value
3     input_image = input_image / np.max(input_image.astype('float'))
4     height, width, vec = input_image.shape
5     row_num = height//im_height if height%im_height==0 else height//im_height+1
6     col_num = width//im_width if width%im_width==0 else width//im_width+1
7
```

```

8     new_image_array = np.zeros((row_num*col_num, im_height, im_width, vec))
9
10    for i in range(row_num):
11        for j in range(col_num):
12            temp_image = input_image[im_height*i:im_height*(i+1),
13                                    im_width*j:im_width*(j+1), :]
14            new_image_array[i*col_num+j] = add_image_magin(temp_image,
15                                                            im_width=im_width,
16                                                            im_height=im_height,
17                                                            color=0)
18    return new_image_array, row_num, col_num

```

## D.2. Procesamiento

Una vez que las imágenes son aptas para ingresar al modelo de Machine Learning, se pasa a procesarlas para estimar su área quemada. La función utilizada durante el procesamiento toma como entrada una lista de imágenes y el modelo entrenado, para luego retornar una lista con las predicciones para cada una de las secciones de la imagen entregadas por el modelo UNet.

```

1 def batch_predict(input_image_array, model):
2     num, height, width, vec = input_image_array.shape
3     preds_array = np.zeros((num, height, width, 1))
4     # to avoid OOM
5     for ii in range(input_image_array.shape[0]):
6         preds_array[ii] = model.predict(np.expand_dims(input_image_array[ii, :, :, :],
7                                                         axis=0),
8                                         verbose=1)
9     return preds_array

```

## D.3. Post-procesamiento

Al tener la lista de salidas del modelo a partir de las secciones de la imagen ingresada, estas se deben unir para reconstruirla. Una vez que esta es reconstruida, se debe utilizar el sistema de umbrales para filtrar el área quemada.

### D.3.1. Reconstrucción de la Imagen

La primera función utilizada toma la lista de imágenes de salida del modelo y el tamaño de la imagen original ingresada, con el fin de generar una salida donde los píxeles hacen referencia a los de la imagen de entrada.

```

1 def combine_image(input_image, row_num, col_num,
2                  original_height, original_width,
3                  im_height=480, im_width=720,
4                  remove_ghost=True):
5     num, height, width, vec = input_image.shape

```

```

6 new_image = np.zeros((height*row_num, width*col_num, vec))
7 for i in range(row_num):
8     for j in range(col_num):
9         # Remove the ghost caused by CNN model prediction boundary artifacts.
10        # by padding same 4 pixels on the boundary
11        if remove_ghost:
12            # padd all four edges
13            input_image[i*col_num+j, :, :] = np.pad(input_image[i*col_num+j,
14                4:height-4, 4:width-4, :],
15                ((4, 4), (4, 4), (0, 0)),
16                'edge')
17
18        new_image[im_height*i:im_height*(i+1),
19            im_width*j:im_width*(j+1), :] = input_image[i*col_num+j,
20            :, :]
21    return new_image[:original_height, :original_width, :]

```

La siguiente función transforma la imagen de salida en una que se pueda visualizar en Python, transformando los valores de cada píxel a un número entero. Además, los píxeles son transformados para que estén dentro del rango entre 0 y 255.

```

1 def conv_float_int(image):
2     # convert array back to [0, 255] for display
3     if not np.any(image):
4         return image.astype(int)
5     return ((image-np.min(image))/(np.max(image)-np.min(image))*255).astype(int)

```

### D.3.2. Sistema de Umbrales

Al tener la imagen de salida del modelo reconstruida solo falta utilizar el sistema de umbrales para obtener la salida final del método de estimación de área quemada. A continuación se muestra el código utilizado para obtener esta salida, incluyendo el pre-procesamiento, el procesamiento y el post-procesamiento de la imagen ingresada (`uploaded_image`).

```

1 input_image_array = np.array(uploaded_image)
2 original_width, original_height, pix_num = input_image_array.shape
3 new_image_array, row_num, col_num = preprocess_input_image(input_image_array)
4
5 preds = batch_predict(new_image_array, model)
6
7 preds_t_lower = (preds > 0.19).astype(np.uint8)
8 preds_t_upper = (preds > 0.39).astype(np.uint8)
9
10 dif_preds = np.sum(np.abs(preds_t_lower - preds_t_upper))
11 area1 = round(0.001 * output_pred.shape[0] * output_pred.shape[1])
12
13 if dif_preds >= area1:
14     preds_t = preds_t_upper
15 else:
16     preds_t = preds_t_lower
17

```

```
18 output_mask = conv_float_int(combine_image(preds_t, row_num, col_num,  
19                                     original_width, original_height,  
20                                     remove_ghost=False)[:,:0])
```

Así, finalmente la imagen de salida mostrada en la interfaz corresponde la matriz `output_mask`.