



Universidad Autónoma de Yucatán

Facultad de Ingeniería

**Herramientas estadísticas y computacionales
en imágenes satelitales de Earth Engine
para la exploración de incendios forestales.**

TESIS

Presentada por:

Miguel Isaac Arroyo Velázquez

En opción al título de:

Ingeniero Físico

Mérida, Yucatán, México
2021

Aunque este trabajo hubiere servido para el Examen Profesional y hubiere sido aprobado por el sínodo, solo el autor es responsable de las doctrinas emitidas en él.

Índice general

1. Introducción	10
2. Hipótesis y objetivos	16
3. Antecedentes	17
3.1. Los incendios forestales	17
3.2. Causas de un incendio forestal	17
3.3. Comportamiento y forma en la que se propaga el fuego	18
3.4. Tipos de incendios y ecosistemas sensibles al fuego	19
3.5. Manejo de la problemática	21
3.5.1. Estudios de incendios forestales (Nivel Nacional)	21
3.5.2. Programa de Manejo del Fuego (PMF) 2020 – 2024 (Nivel Nacional) .	21
3.5.3. Instituciones en Yucatán (Nivel Estatal)	22
4. Marco Teórico	23
4.1. Imágenes satelitales	23
4.1.1. Composición de las imágenes satelitales	23
4.1.2. Tipos de imágenes satelitales	25
4.2. Google Earth Engine	27
4.2.1. Catálogo de datos	27
4.2.2. Objetos y operaciones	27
4.2.3. Aplicaciones	28
4.3. Herramientas computacionales	29
4.3.1. Lenguaje de programación: Python	29
4.3.2. Lenguaje de apoyo: R	32
4.3.3. Uso del API de Google Earth Engine	32
5. Metodología	34
5.1. Región de estudio	34
5.2. Obtención de datos	35
5.2.1. Datos nacionales: CONAFOR	35
5.2.2. Datos de la NASA: FIRMS	36
5.2.3. Imágenes satelitales: Google Earth Engine Python API	36
5.3. Filtro de datos	38
5.3.1. Rango temporal de estudio	38
5.3.2. Región de Interés	39

ÍNDICE GENERAL	3
5.3.3. Selección de variables	39
5.4. Carga, limpieza, transformación y procesamiento de datos	41
5.5. Análisis Exploratorio de Datos	41
5.5.1. Estimados de locación	41
5.5.2. Estimados de variabilidad	42
5.5.3. Prueba de hipótesis	44
5.5.4. Visualización de datos	45
6. Análisis Exploratorio de Datos	52
6.1. Datos de la CONAFOR	52
6.1.1. Superficie afectada por los incendios forestales	53
6.1.2. Municipios afectados	55
6.1.3. Tipo de vegetación y área afectada	56
6.1.4. Temporada de incendios	57
6.2. Datos de FIRMS	59
6.2.1. Nivel de confianza en los puntos de calor y ubicación de puntos de calor.	59
6.2.2. Registro de puntos de calor a través del tiempo	61
6.2.3. Temperaturas	63
6.2.4. Potencia Radiactiva del Fuego (FRP)	66
6.2.5. Comparación con datos de la CONAFOR	66
6.3. Datos de Google Earth Engine	67
6.3.1. Registro de incendios forestales por medio de MODIS (Moderate Resolution Imaging Spectroradiometer)	67
6.3.2. Variables ambientales	69
6.3.3. Índices de vegetación	75
6.3.4. Comparación: Mes de Abril del año 2017	78
7. Conclusiones	81
7.1. Trabajos a futuro en la plataforma GEE	83
7.2. Aprendizaje durante el desarrollo de la tesis	83
A. Repositorio	85
B. Visualizaciones extra	86
C. Trabajos realizados durante la tesis	89
Referencias	93

Índice de figuras

1.1. Gráfico del crecimiento poblacional y el número de incendios forestales en México. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)	11
1.2. Distribución porcentual de la superficie forestal, según el tipo de vegetación. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)	12
1.3. Comparativa de la temperatura y precipitación contra los incendios forestales y superficie forestal afectada en México. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)	12
1.4. Número de incendios forestales ocurridos y superficie afectada en el periodo 1970 - 2019. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)	13
1.5. Áreas prioritarias de protección contra incendios forestales, 2020. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)	14
3.1. Triángulo del fuego. Fuente: (CONAFOR, 2016)	17
3.2. Comportamiento del fuego - La Gran Triada. Fuente: (Comisión Nacional Forestal, 2010)	19
3.3. Ilustraciones representativas de tipos de incendios forestales	20
4.1. Representación de imágenes del canal RGB (<i>Red, Green, Blue</i>) Fuente: (Raj, 2018)	24
4.2. Representación de la captura de información para la generación de imágenes satelitales. Fuentes: (Borra, Thanki, y Dey, 2019)	24
4.3. Para mostrar las bandas más importantes del espectro electromagnético, las frecuencias y longitudes de onda tuvieron que pasar por un escalamiento logarítmico. Las fronteras entre las bandas son un tanto arbitrarias. Fuente: (Young y Freedman, 2009)	25
4.4. Gráfico de dispersión creado en Matplotlib , tomando inspiración de (Waskom, s.f.)	31
4.5. Gráfico de dispersión creado en Seaborn , tomando inspiración de (Waskom, s.f.)	31
4.6. Mapa de calor hecho en Folium representando las posiciones y cercanías entre las residencias en el estado de California, EE. UU.	32

5.1. Captura de pantalla del mapa de vegetaciones del estado de Yucatán. Fuentes: (INEGI, 2021a)	34
5.2. Captura de pantalla del mapa de climas del estado de Yucatán. Fuentes: (INEGI, 2021a)	35
5.3. Información sobre incendios forestales. Fuente: https://datos.gob.mx/busca/organization/conafor	35
5.4. Ejemplo de Prueba de Permutación	45
5.5. Ejemplos de distribuciones representadas en histogramas. Fuente: (Holtz, s.f.-c)	46
5.6. Representación de diferentes distribuciones a través de un gráfico de densidad. Fuente: (Holtz, s.f.-b)	46
5.7. Ejemplo de un gráfico Ridgeline con las temperaturas mensuales de la ciudad de Lincoln, Nebraska, EE.UU. (Holtz, s.f.-f)	47
5.8. Ejemplo del valor del Bitcoin entre 2013 y 2018. Fuente: (Holtz, s.f.-d)	47
5.9. Representación de un conjunto de datos en una gráfica de barras (Holtz, s.f.-a)	48
5.10. Representación de los datos previamente representados en una gráfica de barras en 5.9 a un diagrama Lollipop (Holtz, s.f.-e)	48
5.11. Elementos de un mapa de calor con datos numéricos y categóricos. Fuente: (Ribecca, s.f.-c)	49
5.12. Elementos de un diagrama de cajas y bigotes. Fuente: (Ribecca, s.f.-a)	50
5.13. Elementos de un mapa coroplético. Fuente: (Ribecca, s.f.-b)	51
 6.1. Mapa de puntos. Ubicación de los incendios forestales registrados en el estado de Yucatán en el año 2017. Información del conjunto de datos de la CONAFOR.	53
6.2. Histograma. Distribución de área afectada por incendios forestales. Se puede observar que la mayor concentración de datos se encuentran en el primer <i>bin</i> del histograma. (El ancho del bin es de 100 ha). Información del conjunto de datos de la CONAFOR.	54
6.3. Box plot. Distribución de los datos del atributo ‘Total’ separados en dos grupos: si es ANP o no es ANP.	55
6.4. Lollipop plot. Se muestra el registro de instancias (incendios forestales) durante el 2017 en los municipios de Yucatán. Información del conjunto de datos de la CONAFOR.	56
6.5. Lollipop plot. Se muestra el registro de la extensión de superficie afectada por incendios forestales durante el 2017 en los municipios de Yucatán. Información del conjunto de datos de la CONAFOR.	56
6.6. Gráfica de barras. Se muestra el registro de los tipos de vegetaciones afectadas por los incendios forestales y la extensión que estos últimos afectaron. Información del conjunto de datos de la CONAFOR.	57
6.7. Mapa de puntos. Distribución del tipo de vegetación afectada por los incendios forestales en el estado. Información del conjunto de datos de la CONAFOR.	57
6.8. Gráficas de barras. Comparación de número de incendios y superficie afectada mensual. Información del conjunto de datos de la CONAFOR.	58
6.9. Histograma y Box plot. Distribución de niveles de confianza de los puntos de calor. Información del conjunto de datos de FIRMS.	60

ÍNDICE DE FIGURAS

6

6.10. Mapa de y gráfico de densidades en dos dimensiones. Información del conjunto de datos de FIRMS.	61
6.11. Serie de tiempo. Distribución del registro del número de puntos de calor con un nivel de confianza arriba del 84 % a través de año en el estado de Yucatán. Información del conjunto de datos de FIRMS.	62
6.12. Gráfica de barras. Número de puntos de calor registrados con un nivel de confianza por arriba del 84% por mes durante el periodo del 2001 al 2020, en el estado de Yucatán. Información del conjunto de datos de FIRMS.	62
6.13. Serie de tiempo. Distribución del registro del número de puntos de calor con un nivel de confianza arriba del 84 % a través de los años del 2001 al 2020, en el estado de Yucatán. Información del conjunto de datos de FIRMS.	62
6.14. Matriz de correlación. Correlación entre el número de puntos de calor registrados por año y el año, en el estado de Yucatán. Información del conjunto de datos de FIRMS.	63
6.15. Gráfico de dispersión y ajuste lineal. Número de registros de puntos de calor a través de los años del 2001 al 2020. Información del conjunto de datos de FIRMS.	63
6.16. Histogramas. Comparación entre los atributos <code>brightness</code> y <code>bright_t31</code> . Información del conjunto de datos de FIRMS.	64
6.17. Histograma. Distribución de la diferencia de las medias muestrales tras la realización de una prueba de hipótesis con la Prueba de Permutación.	65
6.18. Matriz de correlación. Correlación entre la temperatura media anual de los puntos de calor y el año. Información del conjunto de datos de FIRMS.	65
6.19. Gráfico de dispersión y ajuste lineal. Temperatura media anual de los puntos de calor a través de los años 2001 al 2020. Información del conjunto de datos de FIRMS.	65
6.20. Gráfico de densidades en 2 dimensiones. Concentración y comparación de los datos en los atributos de FRP y Temperatura del punto de calor. Información del conjunto de datos de FIRMS.	66
6.21. Mapa de puntos. Comparación incendios forestales registrados por la CONAFOR en el año 2017 y los puntos de calor registrados por FIRMS durante el periodo Marzo-Mayo del 2017.	67
6.22. Mapa. Superficie afectada por incendios forestales en el estado de Yucatán durante el periodo de tiempo de 2001 a 2019 con un nivel de confianza mayor a 84 %. Fuente: (Lizundia-Loiola, Otón, Ramo, y Chuvieco, 2020) vía Google Earth Engine	68
6.23. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y temperatura mediana de la superficie en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuentes: (Lizundia-Loiola et al., 2020) y (Wan, Hook, y Hulley, 2015) vía Google Earth Engine	70
6.24. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y humedad del suelo mediana en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuentes: (Lizundia-Loiola et al., 2020) y (Abatzoglou, Dobrowski, Parks, y Hegewisch, 2018) vía Google Earth Engine.	70

6.25. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuente: (Lizundia-Loiola et al., 2020) y (Abatzoglou et al., 2018) vía Google Earth Engine.	71
6.26. Histograma. Distribuciones y estimados de locación y variabilidad de las variables ambientales: Temperatura de la superficie, Humedad del suelo y PSDI. Información obtenida por la plataforma Google Earth Engine	72
6.27. Mapa de calor de serie de tiempo. Cambios a través de los años 2001 a 2019 de las variables ambientales: Temperatura de la superficie, Humedad del suelo y PSDI. Información obtenida por la plataforma Google Earth Engine	72
6.28. Ridgeline plot. Distribución de temperatura de la superficie a través de los meses en los años 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.	73
6.29. Gráfico de barras. Media mensual de las variables ambientales, Humedad del suelo y PDSI, en los años 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.	73
6.30. Mapa de calor de matriz de correlaciones. Matriz de correlaciones entre las variables ambientales: Temperatura de la superficie, Humedad del suelo y PSDI. Información obtenida por la plataforma Google Earth Engine.	73
6.31. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y temperatura mediana de la superficie en el estado de Yucatán en los meses Marzo, Abril y Mayo en el los años 2001 a 2019. Fuentes: (Lizundia-Loiola et al., 2020) y (Wan et al., 2015) vía Google Earth Engine	74
6.32. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y humedad del suelo mediana en el estado de Yucatán en los meses Marzo, Abril y Mayo en el los años 2001 a 2019. Fuentes: (Lizundia-Loiola et al., 2020) y (Abatzoglou et al., 2018) vía Google Earth Engine.	74
6.33. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI en el estado de Yucatán en los meses Marzo, Abril y Mayo en el los años 2001 a 2019. Fuentes: (Lizundia-Loiola et al., 2020) y (Abatzoglou et al., 2018) vía Google Earth Engine.	75
6.34. Mapa. Mediana de NDVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2001 a Diciembre 2003 (periodo de tiempo base). Fuentes (Didan, 2015) vía Google Earth Engine	76
6.35. Mapa. Mediana de NDVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2004 a Diciembre 2019 (periodo actual). Fuentes (Didan, 2015) vía Google Earth Engine	76
6.36. Mapa. Cambios en el NDVI entre el periodo base y el periodo actual. Fuentes (Didan, 2015)	76
6.37. Mapa. Mediana de EVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2001 a Diciembre 2003 (periodo de tiempo base). Fuentes (Didan, 2015) vía Google Earth Engine	77
6.38. Mapa. Mediana de EVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2004 a Diciembre 2019 (periodo actual). Fuentes (Didan, 2015) vía Google Earth Engine	77

6.39. Mapa. Cambios en el EVI entre el periodo base y el periodo actual. Fuentes (Didan, 2015)	77
6.40. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y temperatura mediana de la superficie en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: (Lizundia-Loiola et al., 2020) y (Wan et al., 2015) vía Google Earth Engine	78
6.41. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y humedad del suelo mediana en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: (Lizundia-Loiola et al., 2020) y s (Abatzoglou et al., 2018) vía Google Earth Engine	79
6.42. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI mediana en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: (Lizundia-Loiola et al., 2020) y (Abatzoglou et al., 2018) vía Google Earth Engine	79
6.43. Mapa. Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y mediana de cambios en el EVI (con respecto al mes anterior) en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: (Lizundia-Loiola et al., 2020) y (Didan, 2015) vía Google Earth Engine	80
6.44. Mapa. Acercamiento a las observaciones de incendios forestales de la CONAFOR, FIRMS y Google Earth Engine.	80
A.1. Repositorio de GitHub con el código de programación utilizado en la tesis	85
B.1. Mapa coroplético. Distribución de área afectada en los municipios de Yucatán en el 2017. Información del conjunto de datos de la CONAFOR.	86
B.2. Series de tiempo y gráfica de densidad Comportamiento de las variables ambientales: Temperatura de la superficie, humedad del suelo y PDSI. Información obtenida por la plataforma Google Earth Engine.	87
B.3. Diagrama de Sankey. Relación entre las causas que ocasionaron el incendio forestal y el tipo de incendio en el que resultó en el estado de Yucatán en el 2017. Información del conjunto de datos de la CONAFOR	88
C.1. Publicación en towardsdatascience. Titulo: Visualizing Mexican wildfires during 2017 with folium	89
C.2. Publicación en tacosdedatos. Titulo: Tutorial introductorio a las imágenes satelitales	90
C.3. Poster presentado en el "9th International Fire Ecology and Management Congress"	91

Índice de cuadros

3.1. Descripción y relación del fuego con las diferentes categorías de ecosistemas de acuerdo con el régimen de fuego (Comisión Nacional Forestal, 2010)	21
5.1. Listado de variables y conjuntos de datos extraídos del catálogo de datos de Google Earth Engine	37
5.2. Lista de Índices de Vegetación de Landsat 8	38
5.3. Listado de atributos a usar y su descripción. Datos FIRMS. Fuente: (Earth Science Data Systems Program, 2021)	40
6.1. Estimados de locación y variabilidad del atributo ‘Total’ del conjunto de datos de la CONAFOR	53
6.2. Estimados de locación y variabilidad del atributo ‘Total’ para ANP del conjunto de datos de la CONAFOR	54
6.3. Atributos del área máxima afectada por incendios forestales del conjunto de datos de la CONAFOR	55
6.4. Atributos del área máxima afectada por incendios forestales del conjunto de datos de la CONAFOR	58
6.5. Estimados de locación y variabilidad del atributo ‘confidence’ del conjunto de datos de FIRMS	59
6.6. Estimados de locación de las imágenes satelitales de las variables ambientales: Temperatura de la superficie, Humedad del suelo y PDSI durante el periodo de tiempo del 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.	69

Capítulo 1

Introducción

Perspectiva general

El fuego y los ecosistemas forestales forman de manera conjunta una dinámica que puede ser de gran beneficio o puede contribuir al deterioro ambiental, según sean las condiciones en las que se presente: Los incendios forestales ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

Desde una perspectiva social, el uso del fuego es benéfico, siempre y cuando se tenga un control en la región de afectación, tal como en campos agrícolas o como tratamiento complementario en actividades de manejo forestal. Dentro de los beneficios que el fuego trae ante las comunidades humanas está el uso en la cocina, fuente de calor, luz, caza, cultivos y producción de energía; sin embargo, puede ser dañino para los ecosistemas que no están adaptados al fuego ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

A escala mundial y nacional, el Convenio Marco de las Naciones Unidas sobre el Cambio Climático (CMNUCC), la Convención de las Naciones Unidas de Lucha contra la Desertificación (CNULCD), el Convenio de las Naciones Unidas sobre la Diversidad Biológica (CDB) y el Foro de las Naciones Unidas sobre Bosques (FNUB) reconocen el papel crítico del fuego en el mantenimiento de los ecosistemas forestales dependientes del fuego. Pero también reconocen que el fuego inadecuado es un factor que contribuye a la deforestación, la degradación del bosque y la destrucción de medios de subsistencia, de la biodiversidad y la infraestructura ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

Actualmente, la humanidad se ha convertido en la principal responsable de incendios forestales, sobrepasando otros sucesos naturales, tales como los rayos. Ligado a lo que se comenta, existen otras afectaciones que han influido en el comportamiento del fuego, como la crisis climática y el cambio en patrones de lluvia y sequías ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

Situación nacional y prioridades

Contexto social

La situación de los bosques de México se identifica por su carácter social de la tenencia y por la pobreza en que vive la mayoría de sus habitantes (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021).

En el caso de los ejidos y comunidades que existen, 8928 poseen bosques y selvas, dentro de los que se estiman 3056 ejidos con su principal fuente de ingresos en las actividades relacionadas con los recursos forestales. (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021).

Bajo el contexto del carácter social, otro de los factores que alteran los regímenes del fuego son el crecimiento de las comunidades rurales y urbanas, el abandono rural, tala ilegal y conversión de ecosistemas. La tendencia de estos efectos se pueden observar en la Figura 1.1.

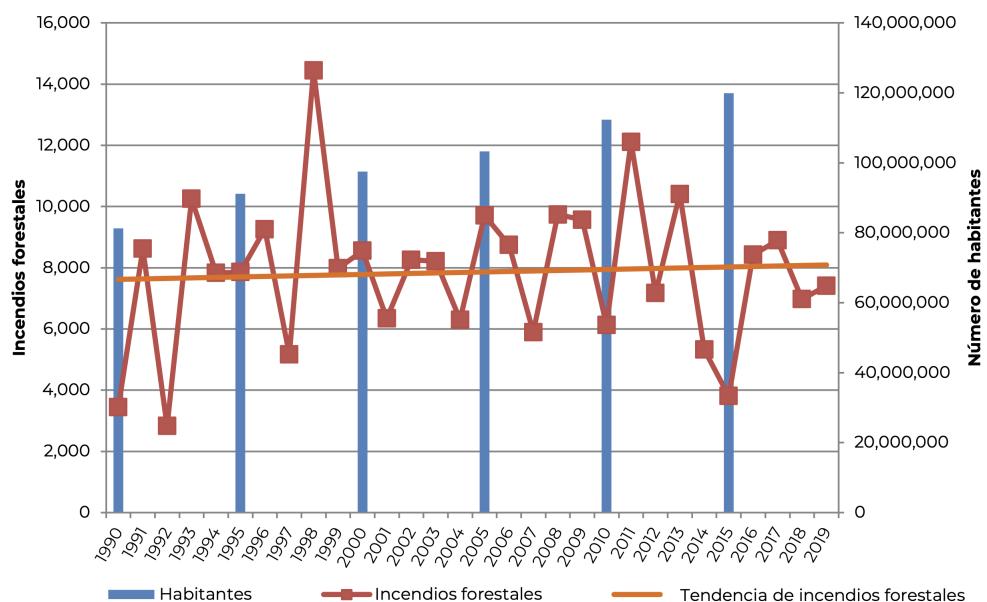


Figura 1.1: Gráfico del crecimiento poblacional y el número de incendios forestales en México. Fuente: (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021)

Contexto ambiental y natural

La superficie forestal de México, es de 138 millones de hectáreas (ha), equivalente al 70 % del territorio nacional y los principales ecosistemas forestales son los matorrales xerófilos (41.2 %), los bosques templados (24.2 %), las selvas (22.8 %) y otros tipos de asociaciones de vegetación forestal (11.8 %) (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021). Más información en la Figura 1.2.

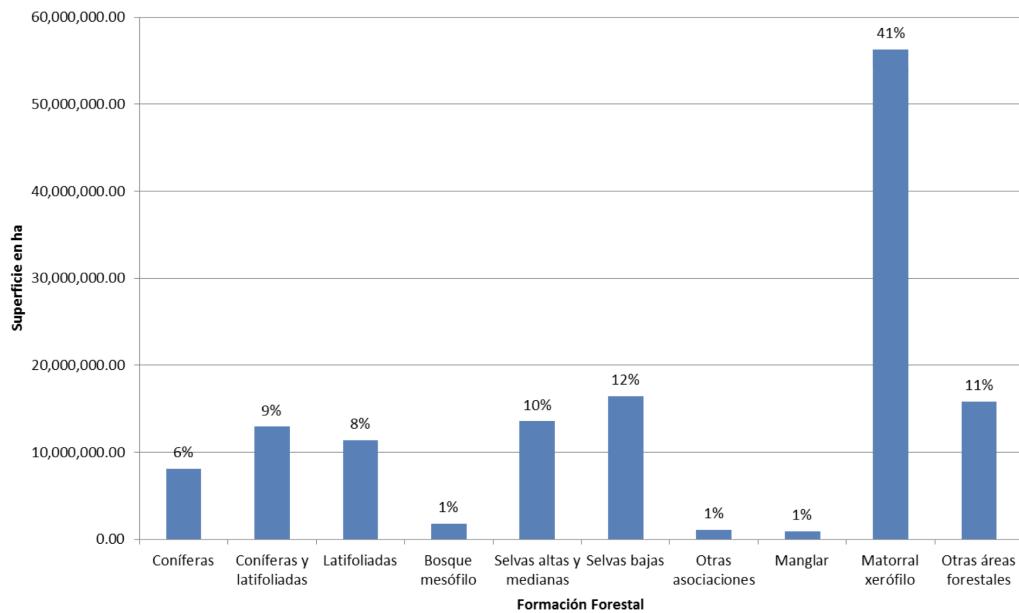


Figura 1.2: Distribución porcentual de la superficie forestal, según el tipo de vegetación. Fuente: ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#))

En los últimos años, los efectos de la crisis climática han influenciado en la frecuencia y la propagación de los incendios forestales mediante la alteración de la temperatura, la precipitación, la humedad en el ambiente, el viento, la dinámica de los combustibles forestales, la composición de las especies, la estructura de la vegetación y la humedad del suelo ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

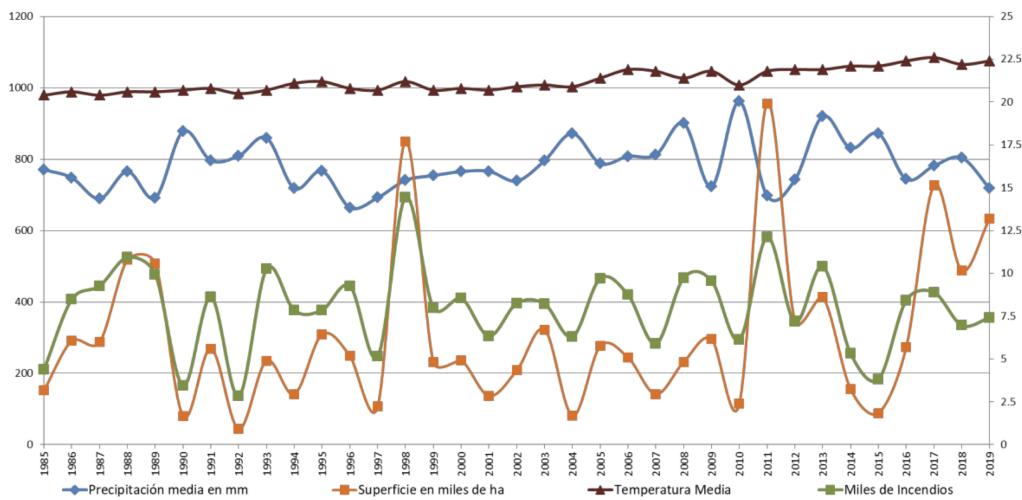


Figura 1.3: Comparativa de la temperatura y precipitación contra los incendios forestales y superficie forestal afectada en México. Fuente: ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#))

Como se mencionó con anterioridad, una de las afectaciones son por cambios en temperatura y efectos ambientales. Esto se refleja en la Figura 1.3, la temperatura media ha estado en

constante incremento, mientras que la precipitación sigue una tendencia dentro del promedio, lo que ha ocasionado mayor sequedad y disponibilidad de combustibles forestales en los ecosistemas, dando como consecuencia incendios con mayor dificultad de control ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

Durante el periodo de 1970 al 2019 se presentó como un promedio anual 7095 incendios y 260998 ha, también como media anual. Los años a destacar son 1998, 2011 y 2017 con presencia de fuego de 849632, 956405 y 726361 ha respectivamente ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)). Ver Figura 1.4

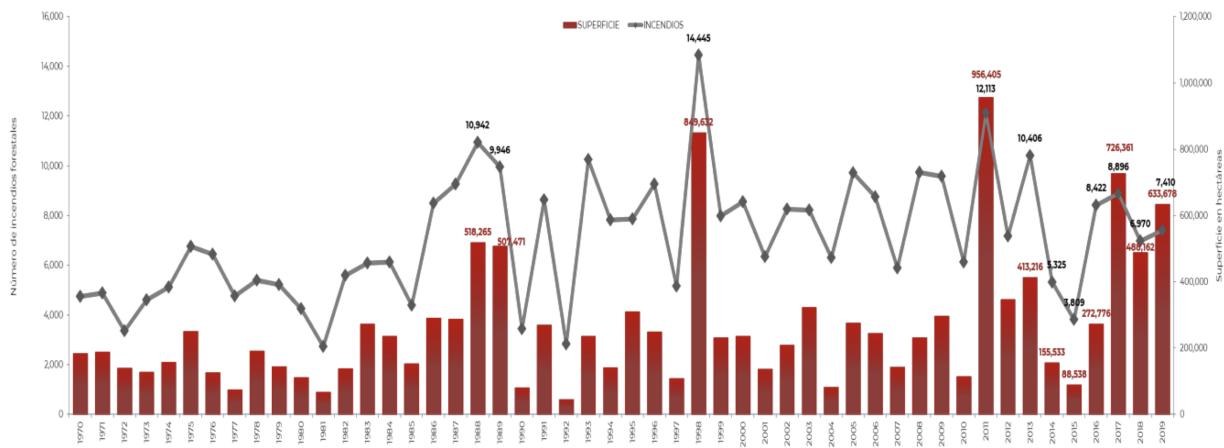


Figura 1.4: Número de incendios forestales ocurridos y superficie afectada en el periodo 1970 - 2019. Fuente: ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#))

Áreas y zonas prioritarias

De acuerdo al Informe de resultados del Inventario Nacional Forestal y de Suelos 2009 – 2014, de las 138 millones de hectáreas de superficie forestal, 38.7 millones de extáreas (un 28 % de la extensión forestal) corresponden a una prioridad alta y muy alta en materia de incendios forestales. ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)). La Figura 1.5 muestra un mapa las prioridades de cada zona forestal.

Causas y consecuencias

Algunos de los factores que afectan esta problemática, sea desde su origen, propagación o dinámica de los incendios forestales, son:

- Fragmentación del paisaje e introducción de plantas exóticas: Cuando se introducen al ecosistema especies que prosperan con el fuego, haciendo que ecosistemas que rara vez experimentan fuego, ahora lo experimenten frecuentemente. ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
- Acumulación y disponibilidad de combustibles: Para este aspecto pueden ser los siguientes:

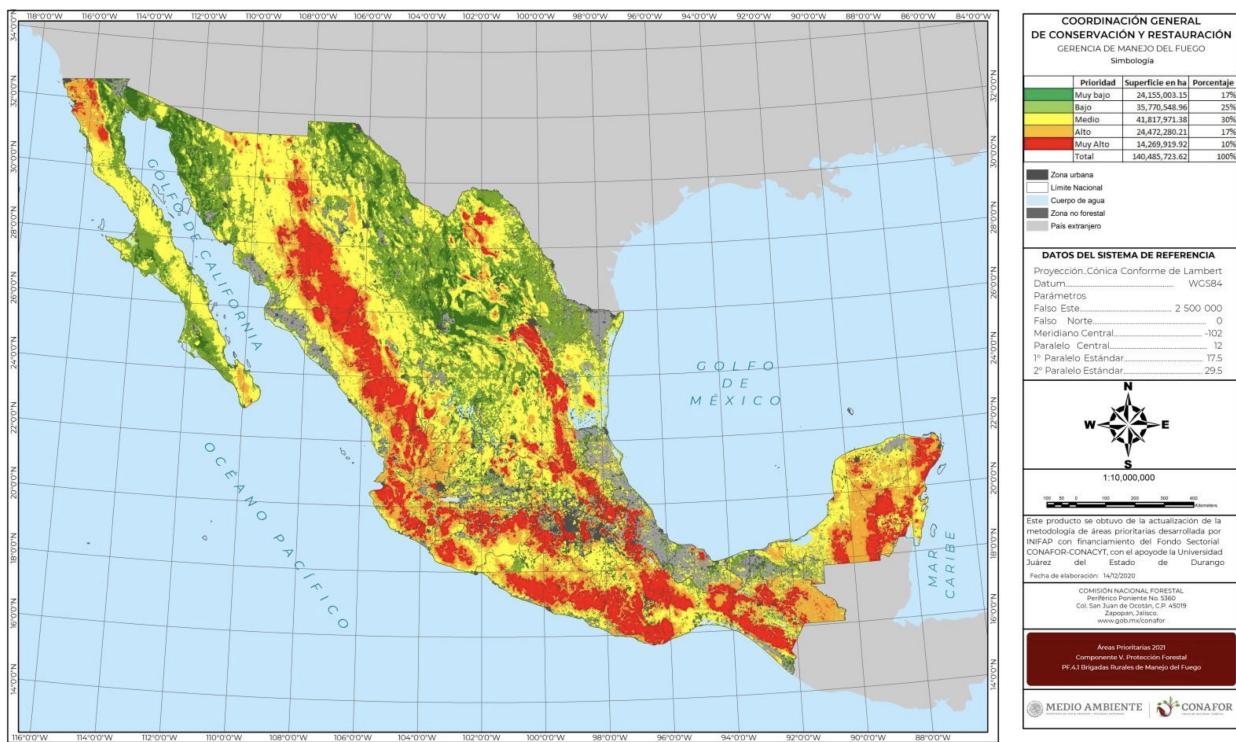


Figura 1.5: Áreas prioritarias de protección contra incendios forestales, 2020. Fuente: ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#))

1. Supresión de incendios forestales ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
 2. Falta de manejo o cultura de silvicultura (disciplina encargada del manejo de los bosques) ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
 3. Fenómenos naturales como huracanes, plagas, enfermedades, especies invasoras o sequías ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
 4. Tala clandestina ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
 5. Condiciones meteorológicas críticas ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
- Uso inadecuado del fuego en actividades como limpieza de terrenos urbanos, quema bajo objetivos productivos como control de plagas o preparación de terrenos agropecuarios, cultivos ilícitos y la cacería; no se puede descartar el uso de fogatas en recreación y quema de residuos en derechos de vía. Cuando se pierde el control del fuego en estas acciones, provocan incendios forestales con altos impactos negativos desde una perspectiva social, económica y ecológica ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
 - Limitada cultura de prevención de incendios forestales: Esto involucra el bajo conocimiento

miento de la dinámica entre el fuego, los recursos naturales y las comunidades rurales. Dentro de este apartado se pone atención a la falta de conciencia ambiental en la ciudadanía y la planificación urbana, ya que uno de los elementos que favorecen el origen y propagación de los incendios es la proliferación de fraccionamientos ecológicos en zonas arboladas de interfaz urbana-forestal ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

- Limitada información científica para la toma de decisiones. La literatura gris es un gran obstáculo ya que genera un déficit de conocimiento frente a las amenazas que sufren los ecosistemas y la sociedad por la degradación de los ecosistemas forestales, o caso contrario, la alta frecuencia de incendios forestales ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

Las consecuencias del deterioro de los ecosistemas forestales debido a la alteración de los regímenes de fuego, desde una perspectiva social, económica y ecológica son:

- Pérdida de superficie arbolada a causa de incendios forestales severos ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
- Afectación al hábitat de las especies de flora y fauna de los ecosistemas ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
- Incremento de emisiones de gases de efecto invernadero ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).
- Erosión, compactación y pérdida de permeabilidad del suelo ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

El presente trabajo aporta conocimiento sobre la incidencia de los incendios forestales en la región del estado de Yucatán. Se presenta información sobre los conjuntos de datos de imágenes satelitales que serán usados para la exploración de los incendios forestales así como el proceso y las herramientas para la extracción de la información para este tipo de datos

Capítulo 2

Hipótesis y objetivos

Hipótesis

La adquisición de imágenes satelitales a través de la plataforma **Google Earth Engine (GEE)** nos proporciona variables relevantes para realizar un análisis histórico de incendios forestales en la Península de Yucatán.

Objetivo general

Examinar diferentes fuentes de datos que brindan información de incendios forestales en la región de la Península de Yucatán para su exploración y análisis.

Objetivos específicos

- Determinar las variables temporales, climatológicas y de vegetación que puede proporcionar las imágenes satelitales de la plataforma de Google Earth Engine.
- Implementar una metodología basada en procesos estadísticos y computacionales que permita una manipulación, exploración y visualización de los datos provenientes de las imágenes satelitales en la región de la Península de Yucatán.

Capítulo 3

Antecedentes

3.1. Los incendios forestales

No se ha encontrado algún registro sobre si los incendios forestales son clasificados como desastre natural en México, sin embargo, en EE. UU. es clasificado como uno por la Agencia de Protección del Medio Ambiente (US EPA, por su nombre en inglés). Interesante saber que aproximadamente el 90 % de los incendios forestales ocurridos son causados por seres humanos y el otro 10 % por la propia naturaleza (Wolters, 2020)

3.2. Causas de un incendio forestal

Para que un incendio forestal inicie se necesitan de 3 elementos, llamados en conjunto “Triángulo del fuego” (Comisión Nacional Forestal, 2010) (Figura 3.1):



Figura 3.1: Triángulo del fuego. Fuente: (CONAFOR, 2016)

- Combustible: Es cualquier material inflamable que se encuentre en el área: árboles y ramas secas, pastizales, arbustos e incluso hogares y vehículos .

- Calor: La ayuda de una chispa o fuente de calor es primordial para iniciar el fuego, las fuentes solo son 2:
 - Naturales: Erupciones volcánicas, rayos de tormentas eléctricas o el simple calor que proporciona el sol ([Wolters, 2020](#)).
 - Humanas: Aquí las fuentes son múltiples, estas pueden ser accidentales, como roturas de líneas eléctricas o colisiones entre medios de transporte. Existen también las negligencias humanas como: quema de basura, limpieza de vías de carretera, quemadas agropecuarias no controladas, quemadas por conflictos entre comunidades, fogatas, colillas de cigarros y uso del fuego en otras actividades dentro de áreas forestales ([Comisión Nacional Forestal, 2010](#))
- Oxígeno: Para hacer combustión no solo se necesita de una fuente de calor y combustible, el oxígeno es un gran aportador para la continua quema del combustible. Los incendios forestales son a la intemperie, donde el oxígeno circula con libertad, llevando las chispas y otras partículas quemadas, cubriendo un área mayor de quema ([Comisión Nacional Forestal, 2010](#)).

3.3. Comportamiento y forma en la que se propaga el fuego

Las condiciones en las que se producen o se pueden llegar a producir los incendios forestales hablan mucho de la probabilidad de ocurrencia y como se mueve o moverá, para eso se presta atención a “La Gran Triada” mostrada en la Figura 3.2 ([Comisión Nacional Forestal, 2010](#)).

- Condiciones atmosféricas: Temperatura, humedad relativa, velocidad del viento, lluvias, entre otros ([Comisión Nacional Forestal, 2010](#)).
- Topografía: La manera en de la propagación del fuego (tipo de incendio) ([Comisión Nacional Forestal, 2010](#)).
- Combustible: El tipo y cantidad de vegetación que cuente el área ([Comisión Nacional Forestal, 2010](#)).

También se cuenta con otro elementos como:

- La temporada: La época de sequía es de los períodos de tiempo donde se tiene un mayor registro, para el caso de Yucatán es entre el mes de Enero y Junio ([Comisión Nacional Forestal, 2010](#)).
- Los asentamientos humanos: Como se mencionó con anterioridad, muchos de los incendios forestales son por causas humanas, entre mayor acceso estos tengan a las áreas forestales, hay una mayor probabilidad de ocurrencia ([Comisión Nacional Forestal, 2010](#)).

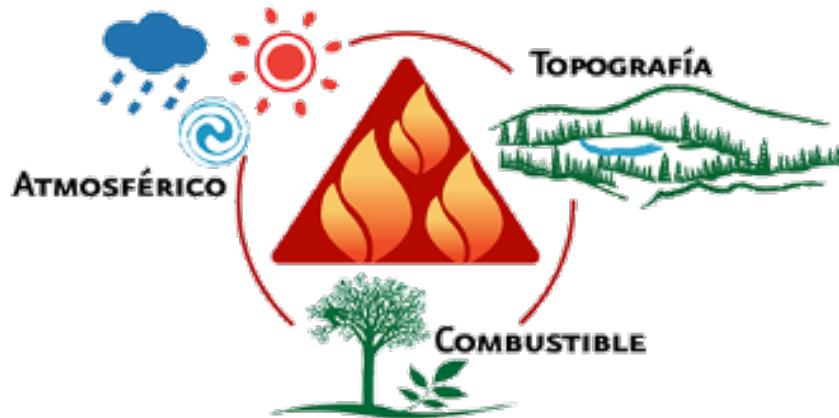


Figura 3.2: Comportamiento del fuego - La Gran Triada. Fuente: ([Comisión Nacional Forestal, 2010](#))

3.4. Tipos de incendios y ecosistemas sensibles al fuego

Como se mencionó anteriormente, el tipo de topografía influye en gran manera la propagación del fuego y es por lo que existen 3 tipos de incendios forestales:

- **Incendios superficiales:** Es cuando el fuego se propaga de manera superficial, alcanzando una altura promedio de 1.5 m de altura. Estos incendios afectan combustibles vivos y muertos tales como: pastizales, ramas, hojas, semillas o pequeños árboles ([Comisión Nacional Forestal, 2010](#)). Figura 3.3a
- **Incendios subterráneos:** Es cuando un incendio superficial llega a propagarse por el suelo. Los daños que pueden causar este tipo de incendios es la quema de materia orgánica acumulada y las raíces. Generalmente no producen llamas y emiten poco fuego ([Comisión Nacional Forestal, 2010](#)). Figura 3.3b
- **Incendios de copa o aéreos:** Este tipo de incendio es de los más destructivos, peligrosos y difíciles de controlar. Su nombre da una idea del alcance del fuego en este tipo de incendio, consume a toda la vegetación. Estos incendios inician de manera superficial y se propagan por continuidad vertical, es decir, escalan hacia arriba para usar la vegetación superior como combustible ([Comisión Nacional Forestal, 2010](#)). Figura 3.3c

El fuego igual viene con diferentes beneficios para la naturaleza al mantener la estructura y dinámica de los ecosistemas, algunas veces las áreas incendiadas propician a que ciertas semillas de las coníferas puedan florecer, tal es el caso de la especie *Pinus pseudostrobus* ([Ramos Reyes, 2015](#)); ayudan a la liberación e incorporación de nutrientes, saneamiento y/o control de enfermedades entre las plantas, reducción de competencia por malezas ([Comisión Nacional Forestal, 2010](#)), eliminar materia en putrefacción, plagas de insectos, animales u otra vegetación que ponga en peligro al ecosistema ([Wolters, 2020](#)). A la relación ecosistema-fuego se le llama régimen del fuego, relaciona diferentes variables del entorno que permite conocer el “patrón” en la naturaleza ([Comisión Nacional Forestal, 2010](#)). En el Cuadro 3.1 se observa la clasificación de ecosistemas de acuerdo con el régimen de fuego que poseen:



(a) Representación de incendio (forestal) superficial. Fuente: (Comisión Nacional Forestal, 2010)



(b) Representación de incendio (forestal) subterráneo. Fuente: (Comisión Nacional Forestal, 2010)



(c) Representación de incendio (forestal) de copa o aéreo. Fuente: (Comisión Nacional Forestal, 2010)

Figura 3.3: Ilustraciones representativas de tipos de incendios forestales

Tipo de ecosistema	Descripción	Relación con el fuego
Sensibles al fuego	Son bosques tropicales perennifolios o subperennifolios (que incluyen selvas altas y medianas perennifolias y subperennifolias), bosque tropical y selva baja (ambos) caducifolios, manglares, bosques de oyamel y bosques de encino	No es requerido para mantener el tipo de vegetación. Los incendios se presentan muy rara vez (cada varios siglos) y únicamente cuando se cuenta con una sequía extrema. Por lo mismo que cuando hay un incendio forestal resulta ser catastrófico. Para la recuperación de la vegetación deben transcurrir siglos
Dependientes del fuego	Encinos arbustivos y matorrales mantenidos por incendios relativamente poco frecuentes (entre 30 y 100 años). En México se cuenta con gran abundancia de estos ecosistemas como en la zona central y noroeste	Incendios superficiales cada dos a diez años. Tienen una buena relación con el fuego ya que los pinos se regeneran bien sobre sitios quemados, cuentan con corteza gruesa y aislante que protege al bosque de las altas temperaturas de los incendios.

Tipo de ecosistema	Descripción	Relación con el fuego
Independientes del fuego	Los ecosistemas de desiertos y tundras entran en esta categoría.	El fuego remueve materia orgánica acumulada y barreras de zacate para que la semilla pueda tener contacto con el suelo mineral. Muchas especies recuperan follaje cuando parte de su copa ha sido afectada por el incendio

Cuadro 3.1: Descripción y relación del fuego con las diferentes categorías de ecosistemas de acuerdo con el régimen de fuego (Comisión Nacional Forestal, 2010).

3.5. Manejo de la problemática

3.5.1. Estudios de incendios forestales (Nivel Nacional)

En el documento *Developing Models to Predict the Number of Fire Hotspots from an Accumulated Fuel Dryness Index by Vegetation Type and Region in Mexico* realizado por diferentes organismos educativos e instituciones de gobierno muestran un sistema de toma de decisiones operativo espacial y temporalmente explícito que pretende brindar información sobre potenciales áreas de peligro de incendio forestal en el país. Esta investigación toma imágenes satelitales MODIS (*Moderate Resolution Imaging Spectroradiometer*) para la predicción del índice de Densidad del Punto Caliente de Fuego (FHD, por su nombre en inglés). Para esta tarea se usan modelos (matemáticos) lineales y no-lineales. Los resultados arrojan una correlación de R2 entre 0.5 - 0.7, valores que se pueden considerar buenos para una predicción a 10 días (Vega-Nieva et al., 2018)

3.5.2. Programa de Manejo del Fuego (PMF) 2020 – 2024 (Nivel Nacional)

El Programa de Manejo del Fuego 2020 – 2024 (PMF) es derivado del Plan Nacional de Desarrollo (2019 – 2024), del Programa Sectorial de Medio Ambiente y Recursos Naturales (PROMARNAT) 2020 – 2024 y del Programa Nacional Forestal 2020 – 2024. En él se establecen las estrategias y líneas de acción que se impulsarán y fortalecerán el periodo dado. Estas acciones serán llevadas a cabo en coordinación con la administración federal, dueños y poseedores de recursos forestales, investigadores, dependencias de gobiernos de cada estado, entre otros (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021).

Con este plan se pretende mejorar la capacidad de respuesta de país para disminuir la alteración de los **regímenes de fuego** de los ecosistemas forestales (Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021).

El término **régimen del fuego** se define como el periodo natural de presencia de incendios en un ecosistema y los efectos que éstos ocasionan. Esta definición integra patrones espaciales y temporales de ocurrencia de incendios en un área, además de sus efectos ecológicos. Los regímenes de incendios son definidos de acuerdo con la frecuencia, severidad, estacionalidad, duración, extensión, distribución espacial, severidad, y tipo de incendio ([Coordinación General de Conservación y Restauración y Gerencia del Manejo del Fuego, 2021](#)).

3.5.3. Instituciones en Yucatán (Nivel Estatal)

De las dependencias e instituciones estatales encargadas de problemáticas ecológicas y del desarrollo sustentable, se encuentra la Secretaría de Desarrollo Sustentable de Yucatán (SDS), la cual esta conformada por diferentes Consejos y Comités.

El Comité de prevención y combate de incendios agropecuarios y forestales del Estado de Yucatán es el encargado de la planear, coordinar y ejecutar acciones permanentes, relacionadas con el objeto de la Ley de Prevención y Combate de Incendios Agropecuarios y Forestales del Estado de Yucatán ([Secretaría de Desarrollo Sustentable de Yucatán, s.f.](#)).

Algunas de las atribuciones más importantes son las siguientes:

- Establecer un Programa Estatal de Prevención y Combate de Incendios Agropecuarios y Forestales, en términos de la presente Ley y de conformidad con las prioridades señaladas en el Plan Estatal de Desarrollo ([Secretaría de Desarrollo Sustentable de Yucatán, s.f.](#)).
- Elaborar, coordinar y aplicar los programas en materia de riesgos por el uso y manejo del fuego en el Estado, vinculándolos con las actividades productivas en áreas rurales, agropecuarias, forestales ([Secretaría de Desarrollo Sustentable de Yucatán, s.f.](#)).
- Promover, en coordinación con la Federación, los programas y proyectos de educación, capacitación, investigación y cultura de prevención y combate oportuno de incendios, acordes con el programa nacional respectivo; así como su debida implementación de ellos ([Secretaría de Desarrollo Sustentable de Yucatán, s.f.](#)).
- Regular el uso del fuego en las actividades relacionadas con la producción agropecuaria y forestal ([Secretaría de Desarrollo Sustentable de Yucatán, s.f.](#)).

Capítulo 4

Marco Teórico

4.1. Imágenes satelitales

Los sistemas de imágenes satelitales permiten de manera fácil y efectiva la colecta de grandes cantidades de datos de regiones del planeta que antes era imposible obtener, causando una gran impacto en la escala y tiempo de los análisis de las variables del lugar (Borra et al., 2019).

4.1.1. Composición de las imágenes satelitales

Cuando se cuentan con **datos no estructurados**, es decir, que no cuentan con un formato específico así como los archivos de valores separados por comas (CSV), o tablas relacionales manejadas por SQL (Skiena, 2017), estos datos se tienen que transformar o dar una cierta estructura para poder extraer la mayor cantidad de información. Para el caso de las imágenes satelitales se pueden tratar como imágenes digitales, las cuales pueden ser estructuradas como vectores y matrices o arreglos de números.

Las imágenes digitales que son usadas dentro del campo de Visión por Computadora están compuestas por (normalmente) tres canales: **Rojo**, **Verde** y **Azul** (**RGB**, por sus siglas en inglés) como se puede observar en la Figura 4.1.

Cada canal tiene un arreglo de valores numéricos llamados píxeles o píxeles, estos son los elementos más pequeños que componen una imagen digital. Cada valor asociado a cada píxel representa la intensidad de ese color en ese píxel (Diccionario prehispánico de dudas, s.f.).

Para el caso de las imágenes satelitales, el análogo a los canales de una imagen digital serían las **bandas**. La bandas en las imágenes satelitales son las responsables de almacenar la información que se está colectando, esta puede ser de variables climatológicas, ambientales o “fotografías” de la superficie terrestre (Borra et al., 2019).

Los sistemas de imágenes satelitales obtienen los datos al captar y medir la **energía de los fotones** emitida o reflejada de los cuerpos terrestres (como se muestra en la Figura 4.2); estos datos son después almacenados en las bandas de las imágenes como pequeños puntos

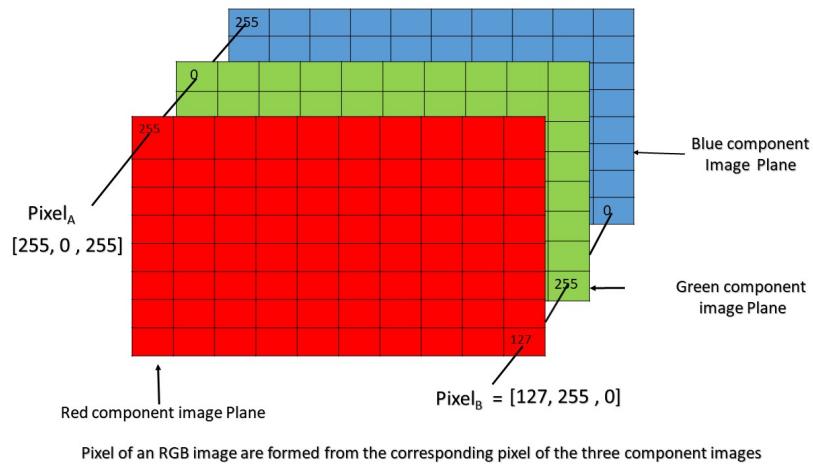


Figura 4.1: Representación de imágenes del canal **RGB** (*Red, Green,Blue*) Fuente: (Raj, 2018)

llamados **datos ráster** (análogos a los píxeles de una imagen digital), representando el valor de la variable del satélite encargado de captarla (Borra et al., 2019).

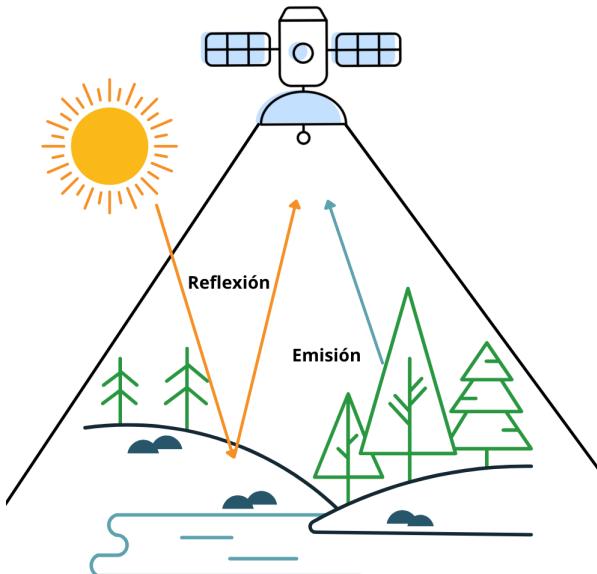


Figura 4.2: Representación de la captura de información para la generación de imágenes satelitales. Fuentes: (Borra et al., 2019)

El **espectro electromagnético** (EM) es la combinación de las diferentes longitudes de ondas, incluyendo las ondas de radio y televisión, la **luz visible** (de 400 a 700×10^{-9} m), radicación infrarroja y ultravioleta, los rayos X y los rayos Gamma, ver la Figura 4.3 (Young y Freedman, 2009).

Cada objeto en la superficie terrestre tiene diferentes valores de reflectancia (cantidad de luz que logra reflejar) a diferentes longitudes de onda. La magnitud de ese valor ayuda a determinar el tipo de objeto en la imagen satelital (Borra et al., 2019). La energía de los fotones puede ser expresada matemáticamente en términos de la longitud de onda (λ) o la frecuencia (ν) (Borra et al., 2019) como:

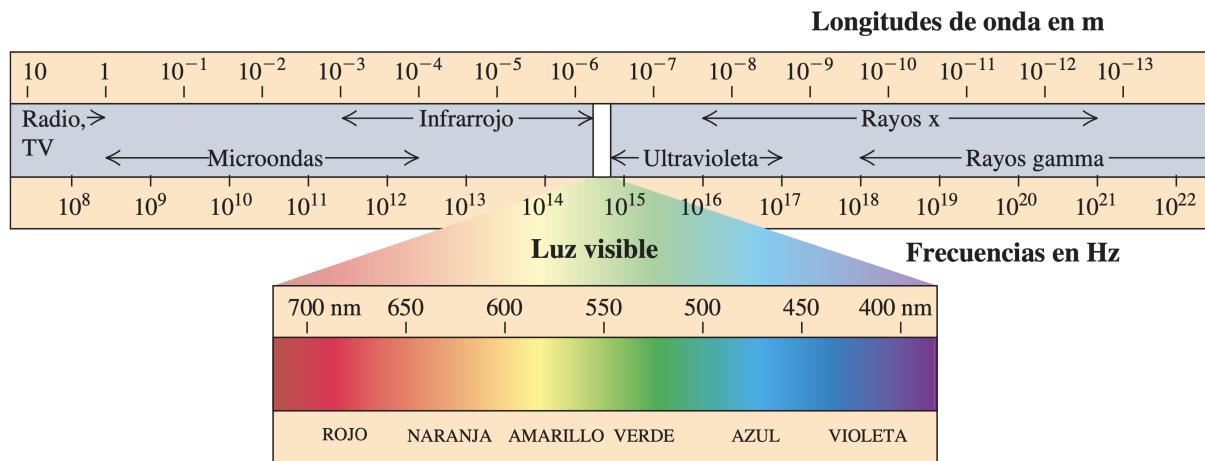


Figura 4.3: Para mostrar las bandas más importantes del espectro electromagnético, las frecuencias y longitudes de onda tuvieron que pasar por un escalamiento logarítmico. Las fronteras entre las bandas son un tanto arbitrarias. Fuente: ([Young y Freedman, 2009](#))

$$E = h\nu = \frac{hc}{\lambda} \quad (4.1)$$

donde c es una constante, conocida como la **velocidad de la luz en el vacío** y h es la **constante de Planck** ([Young y Freedman, 2009](#))

4.1.2. Tipos de imágenes satelitales

Tipo de sensor

Los **sensones ópticos** colectan datos de la magnitud de las longitudes de onda reseñadas en los cuerpos de la superficie terrestre dentro de ellos existen dos principales: de **luz visible** e **infrarrojos** ([Borra et al., 2019](#)).

Los **sensores ópticos de luz visible** sirven para entender la naturaleza de objetos como ríos, montañas, bosques, etc. Este tipo de sensores tienen un bajo desempeño en condiciones de poca luz (sea oscuridad o malas condiciones atmosféricas) ([Borra et al., 2019](#)).

En cambio, los **sensores ópticos infrarrojos**, los cuales captan las ondas infrarrojas emitidas por los cuerpos. Estos se desempeñan bien al captar las altas temperaturas de los cuerpos en la oscuridad ([Borra et al., 2019](#)).

Los **sensores de microondas** por otro lado, obtienen datos de las microondas reflejadas en los cuerpos de la superficie terrestre sin importar las condiciones atmosféricas. Son usados para la observación de valles, montañas, ríos y la condición de los hielos, como su espesor y temperatura ([Borra et al., 2019](#)).

Imágenes Pancromáticas y Multiespectrales

Las **imágenes pancromáticas** son aquellas que cubren un amplio espectro de longitudes de onda. Estas son eficientes al momento de la extracción de atributos y son buenas para

una interpretación fácil de los datos (Borra et al., 2019).

El caso de las **imágenes multiespectrales** son en realidad, un conjunto de imágenes adquiridas por múltiples sensores que operan en diferentes y cortos rangos de longitudes de onda (discretos o continuos) del EM. Este tipo de imágenes mejoran la obtención de información para casos en los que la vista humana esta limitada, tal como casos de condiciones de poca luz, ambientes opacos o la superficie terrestre (minas terrestres y minerales). La clasificación de píxeles en imágenes multiespectrales es una tarea compleja y demanda datos de validación de entrenamiento así como conocimiento de un experto del área (Borra et al., 2019).

Tipo de resolución

La resolución determina la habilidad de detectar, identificar y distinguir los diferentes objetos y sus propiedades. Esta corresponde a la cantidad de información y detalle que puede dar una imagen. Dentro de las imágenes satelitales existen cuatro tipos de resolución: **espacial, espectral, temporal y radiométrica** (Borra et al., 2019).

La **resolución espectral** corresponde al ancho, sensibilidad y ubicación de las bandas espectrales y su número. Una alta resolución espectral significa una alta habilidad de distinción. Los típicos satélites de **Sistema de Información Geográfica (GIS, por sus siglas en inglés)** capturan imágenes de 3 a 8 bandas espectrales (Borra et al., 2019).

Cuando se habla de la **resolución espacial** se refiere al área más pequeña que pueda cubrir un sensor. Entre mayor sea la resolución espacial, más pequeños son los píxeles lo que lleva a una mejor distinción de la superficie terrestre. Las resoluciones espaciales existentes están en el rango de los pocos metros hasta los kilómetros (Borra et al., 2019):

- Imágenes de baja resolución espacial: Arriba de 1 km.
- Imágenes de media resolución espacial: Entre 100 m y 1 km.
- Imágenes de alta resolución espacial: Entre 5 m y 100 m.
- Imágenes de muy alta resolución espacial: Menos de 5 m.

También se cuenta con la **resolución temporal**, correspondiente a los intervalos de tiempo en el que se “toman las imágenes” en ese mismo lugar, es decir, el periodo de tiempo en el que el satélite vuelve a visitar la región. Este tipo de resolución depende de la órbita del satélite y el ángulo de visión (Borra et al., 2019).

Finalmente se tiene la **resolución radiométrica**, esta última se refiere a la precisión de los números digitales usados para mostrar los datos en forma de imágenes. Depende del número de bits usados para representar cada píxel, entre mayor sea esta resolución menor es la variación numérica de la reflectancia. Sin embargo, esta resolución depende fuertemente del tipo de sensor usado (Borra et al., 2019).

4.2. Google Earth Engine

Google Earth Engine (GEE) es una plataforma en la nube que busca minimizar los obstáculos que se tienen cuando se tiene que trabajar o investigar con las imágenes satelitales al usar el poder de cómputo de la compañía Google. GEE también está diseñado para eliminar las barreras y facilitar a los usuarios investigadores, compartir su trabajo con otros colaboradores, organizaciones sin fines de lucro, gobierno e incluso público general (Gorelick et al., 2017).

GEE consiste en un **catálogo de datos** listos para ser analizados en conjunto con un servicio de poder de cómputo de alto rendimiento y cómputo paralelo. Para acceder a dichas herramientas y servicios se cuenta con una **interfaz de programación de aplicaciones (API)** y un **ambiente de desarrollo (IDE)** en internet, que proporciona el prototipado y visualización de los resultados (Gorelick et al., 2017)

4.2.1. Catálogo de datos

El catálogo de datos almacena una gran cantidad de datos geo-espaciales de manera pública. Contiene imágenes satelitales de alta definición (a partir de 10 m), observaciones en el espectro electromagnético visible y no visible, variables ambientales, climáticas, tipo de tierra, topografía y hasta conjuntos de datos con información socio-económica. Todos estos datos ya están listos para su visualización y exploración (Gorelick et al., 2017).

Los usuarios también pueden usar cualquiera de los conjuntos de datos que se encuentren en el catálogo así como pueden usar sus propios conjuntos de datos privados usando los operadores proporcionados por la API de GEE (Gorelick et al., 2017).

El catálogo se actualiza continuamente a un ritmo de casi 6000 escenas al día procedentes de misiones activas, con un desfase o retraso típico de unas 24 horas desde el momento de la adquisición de la imagen (Gorelick et al., 2017).

4.2.2. Objetos y operaciones

La API está compuesta por diferentes tipos de objetos quienes son las representaciones de los diferentes tipos y estructuras de datos (Google Developer Team, 2021h). Las principales y más relevantes son:

- **Imagen (ee.Image):** Es la principal representación de los datos ráster. Los objetos ee.Image están compuestos por una o más bandas, cada banda con su respectivo nombre, tipo de dato, escala, máscara y proyección. Los metadatos de cada imagen están almacenados como un conjunto de propiedades (Google Developer Team, 2021f)
- **Colección de Imágenes (ee.ImageCollection):** Es un grupo de objetos ee.Image. Las **Colecciones de Imágenes**, cuando son parte de un mismo satélite o proyecto y tienen un identificador propio (Google Developer Team, 2021e).

Por ejemplo, para extraer la Colección de Imágenes de Landsat 7 se escribe la operación: `ee.ImageCollection("LANDSAT/LE07/C02/T1_L2")`.

- **Geometría (ee.Geometry)**: GEE maneja datos vectoriales como los límites geográficos provincias, municipios, estados o países. Este tipo de objeto se encarga también de la representación de líneas, polilíneas, puntos y polígonos (Google Developer Team, 2021d).
- **Atributo (ee.Feature)**: Este tipo de objetos son la representación de los datos de atributos GeoJSON. Un objeto `ee.Feature` es aquel que tiene a una **Geometría** (o no) y un conjunto de propiedades almacenadas en un tipo de diccionario de otras propiedades (Google Developer Team, 2021c)
- **Colección de Atributos (ee.FeatureCollection)**: Análogamente a la **Colección de Imágenes** existen las **Colecciones de Atributos**. Estos tipos de objetos también llegan a tener un identificador dentro del catálogo de datos de GEE (Google Developer Team, 2021b).
- **Reductor (ee.Reducer)**: Los **Reductores** son la manera en la que resumimos o agrupamos datos a través del tiempo, espacio, bandas, arreglos o cualquier otro tipo de objeto o estructura de dato en GEE. Esta clase especifica como los datos serán resumidos, puede ser con estadísticas simples como la media, mediana o desviación estándar; a través de resúmenes más complejos como histogramas, listas, regresiones lineales, entre otros; reducciones de espacio (dando una región de reducción a través de una **Geometría**) o resumir una **Colección de Atributos** (Google Developer Team, 2021j).

Existen múltiples operaciones que se pueden hacer sobre cada uno de los objetos, desde aritmética básica (suma, resta, multiplicación y división) así como operaciones optimizadas entre bandas de imágenes como la Diferencia Normalizada entre bandas de imágenes (Gorelick et al., 2017).

4.2.3. Aplicaciones

Las aplicaciones donde se pueden encontrar los conjuntos de datos o el poder computacional que proporciona GEE cubren una amplia gama dentro del campo de las Ciencias de la Tierra.

Un reciente estudio muestra el incremento del uso de Google Earth (GE) y GEE. Dentro de los resultados más relevantes se revela que GE se usa diferentes áreas como Biología, Educación, Salud, Economía y Ciencias de la Información, por ende GE aparece en un mayor número de publicaciones. También GE se usa más como una herramienta de visualización mientras que en GEE existe un énfasis en el análisis de datos masivos (*Big Data*) y series de tiempo. Algo que comparten son las palabras claves en sus publicaciones, donde se destacan “cobertura de la tierra”, “agua”, “modelo”, “vegetación” y “bosques”, donde se puede inferir que son las áreas donde GE y GEE generan un mayor impacto. (Zhao et al., 2021).

Pérdidas forestales

En el año 2013, un grupo de investigadores, realizaron un análisis a nivel mundial sobre las pérdidas forestales. Para dicho análisis recurrieron a los datos públicos de **Landsat** dentro del catálogo de datos de GEE. En este trabajo se mapearon aproximadamente 2.3 millones de kilómetros cuadrados de pérdidas forestalal y 0.8 millones de kilómetros cuadrados de ganancia forestal del año 2000 al 2012. Los resultados y la actualización de las pérdidas y ganancias se pueden encontrar en el catálogo de datos de GEE bajo el nombre **Hansen Global Forest Change** ([Hansen et al., 2013](#)).

Mapa de la Vida

El mapa de la vida es un proyecto desarrollado por diferentes miembros de Departamentos de Ecología y Biología de diferentes universidades y centros de investigación de Estados Unidos y Canadá. Presentan un mapa interactivo que muestra la distribución de especies, señalando las ubicaciones de espesies en riesgo ([Jetz, McPherson, y Guralnick, 2012](#)). El mapa se encuentra en la siguiente liga: https://species.mol.org/species/map/Perdix_daurica

4.3. Herramientas computacionales

Para extraer información de las imágenes se necesita un proceso diferente a cuando se trabajan con datos tabulares y estructurados. Los datos estructurados pueden ser examinados por medio de programas conocidos y populares, tales como Microsoft Excel, STATA, Statgraphics y Tableau Public, siendo el último de licencia libre; las imágenes satelitales y los datos georeferenciados, como con los que se han estado trabajando a lo largo de este trabajo, necesitan otros tipos de programas, o en el caso de este trabajo, un lenguaje de programación.

4.3.1. Lenguaje de programación: Python

Python es un lenguaje de programación de propósito general creado alrededor del año 1990 por Guido Van Rossum en el CWI (*Centrum Wiskunde & Informatica*), Países Bajos ([Nosrati, 2011](#)).

Es flexible, dinámico y de alto nivel. Su popularidad radica principalmente en su sintaxis de alto nivel a comparación de otros lenguajes como **C** o **C#** ya que Python es explícito y claro, al grado de ser muy similar al lenguaje natural del ser humano (en inglés) ([Hill, 2020](#)), por ejemplo, para “imprimir” o mostrar la oración ‘Hola mundo’ en la consola basta con escribir el comando `print('Hola mundo')`. Algunas de las ventajas son las siguientes:

- Es un lenguaje de programación de código abierto u *open source*, como se le conoce a nivel mundial. Es gratis, se puede distribuir y contribuir al desarrollo de este, libremente ([Nosrati, 2011](#)). Así como que la ejecución puede ser a través de diferentes plataformas y sistemas (Windows, Mac OS y Linux)([Ceder, 2010](#)).

- Tiene un amplio catálogo de paquetes de funciones y operaciones, también conocidos como **librerías y módulos de Python**, que permiten a los usuarios realizar diferentes tareas de algún área específica, como: desarrollo web, creación de aplicaciones, cómputo científico, Inteligencia Artificial, entre otros (Ceder, 2010).

Con un enfoque más profundo en el último punto mencionado, se muestran las librerías de Python usadas en el trabajo:

- **Cómputo numérico y científico:**

- **NumPy:** El funcionamiento de NumPy esta cimentado en la manipulación de arreglos, su principal estructura de dato. La manipulación de arreglos es poderosa, compacta y rápida al momento de realizar cálculo numéricos. Las capacidad de NumPy se expresa al manipular los datos como vectores, matrices y arreglos multidimensionales. Es usado ampliamente en el análisis de los datos de diferentes ciencias, como biología, física, ingeniería, entre otras (Harris et al., 2020).
- **SciPy:** Construido sobre NumPy, SciPy es una librería que expande las capacidades de NumPy al integrar algoritmos científicos optimizados. Los submódulos en los que se divide cubren áreas como integración, agrupamiento, álgebra lineal, estadística, métricas (distancias), entre otros (Virtanen et al., 2020).

- **Manipulación de datos relacionales:**

- **Pandas:** Creada en el 2008 (The pandas development team, 2020), Pandas es una librería que al día de hoy es una herramienta relevante y constante en el análisis de datos. Aunque se pueden manipular los datos a través de arreglos de NumPy, Pandas tiene como principales estructuras de datos los `pandas.Series` y `pandas.DataFrame`, que proporcionan una mayor flexibilidad en la manipulación estadística, indexación y limpieza de los datos (Wes McKinney, 2010).
- **GeoPandas:** Cuando se cuentan con datos que tienen como propiedad, figuras con geo-referencias (puntos, líneas, polígonos o multipolígonos), es decir, son datos vectoriales, **GeoPandas** es la herramienta ideal. Esta librería es una extensión de **Pandas**, pero tiene la capacidad de adecuar las operaciones que se realizan en los `pandas.Series` y `pandas.DataFrame` a conjuntos de datos con atributos geométricos. Sus principales estructuras de datos son `geopandas.GeoSeries` y `geopandas.GeoDataFrame` (Jordahl et al., 2020).

- **Visualización de datos:** Parte de la exploración de los datos y la extracción de información relevante, es poder visualizar esta información para la comunicación de los hallazgos de manera directa y de fácil entendimiento. Estas tareas estan a cargo de las siguientes librerías:

- **Matplotlib:** Es de las principales herramientas de visualización de datos así como la base en muchas otras visualizaciones, por ejemplo, se encuentra integrada en Pandas y es la base en la que se construye Seaborn (Hunter, 2007).

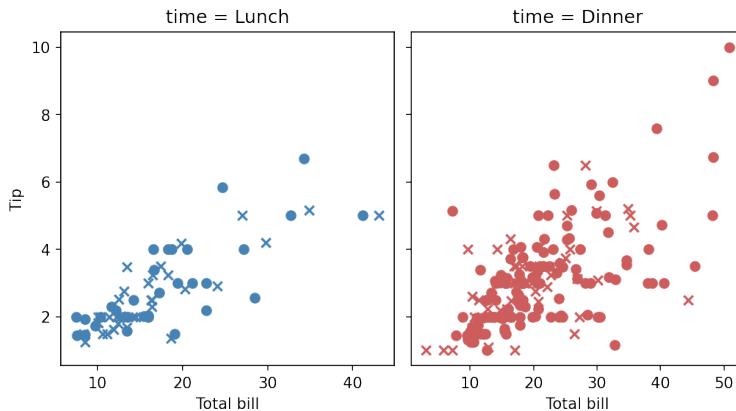


Figura 4.4: Gráfico de dispersión creado en Matplotlib, tomando inspiración de ([Waskom, s.f.](#))

- **Seaborn:** Se enfoca principalmente en gráficos estadísticos con una mejor estilización a comparación de Matplotlib. La integración con Pandas es mucho mejor al darle un conjunto de datos en forma de pandas.DataFrame puede “mapearlo” y al asignar las variables a visualizar, solo hace falta declarar el nombre de la columna en algún parámetro de visualización, dando como resultados visualizaciones claras con menos líneas de código. ([Waskom, 2021](#)).

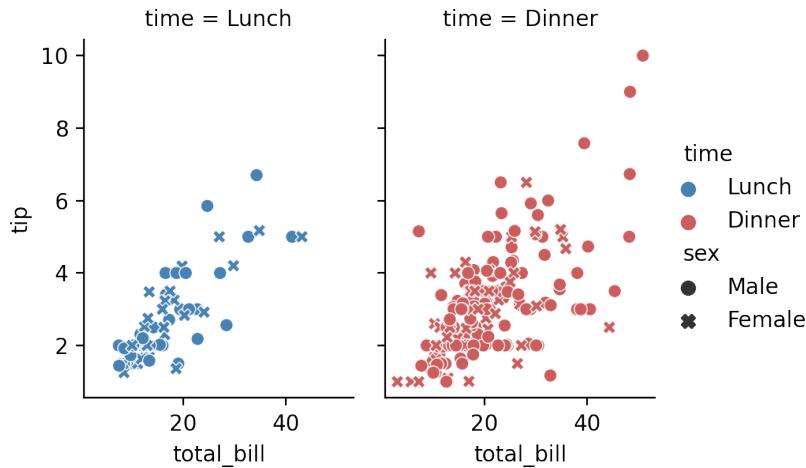


Figura 4.5: Gráfico de dispersión creado en Seaborn, tomando inspiración de ([Waskom, s.f.](#))

- **Folium:** Esta librería facilita la visualización de los datos geo-espaciales ya manipulados previamente (sea con Pandas o GeoPandas) para su representación a través de mapas interactivos. Se encuentra construido sobre la librería leaflet.js de JavaScript ([python visualization, 2021](#)). Un ejemplo se puede ver en la Figura 4.6
- **Documentación:** Para llevar un orden en los procedimientos, mejor visualización de la información, transparencia en los cálculos y reproducibilidad de código se usan **Jupyter Notebooks**. Este editor de código o ambiente de desarrollo es parte de los productos del **Projecto Jupyter** ([Project Jupyter](#)), los cuales son de amplio uso dentro de la

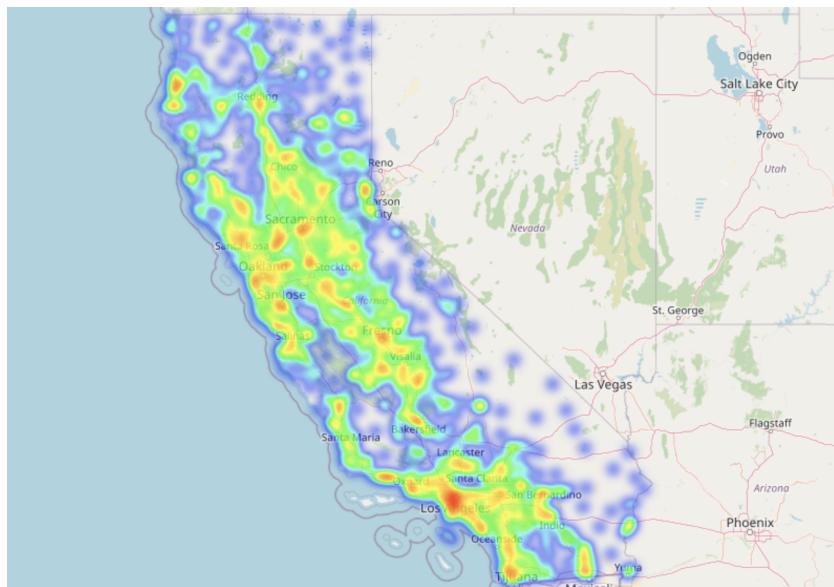


Figura 4.6: **Mapa de calor** hecho en **Folium** representando las posiciones y cercanías entre las residencias en el estado de California, EE. UU.

comunidad científica, y últimamente dentro de las áreas y trabajos involucrados en el campo de la Inteligencia Artificial y Ciencia de Datos. ([Kluyver et al., 2016](#))

Google Colaboratory o simplemente **Colab**, es un servicio de Google donde se pueden ejecutar Jupyter Notebooks de Python a través de los procesadores de Google ([Nelson y Hoover, 2020](#)), esta es una excelente alternativa para poder trabajar cuando no se tiene acceso a una maquina física, ya que gran parte de las librerías de Python del campo de Inteligencia Aritificial y Ciencia de Datos pueden ser usadas ahí mismo.

4.3.2. Lenguaje de apoyo: R

R es un lenguaje de programación con un fuerte énfasis el cómputo estadístico. Al igual que Python, es *open source* y con una sintaxis amable al usuario ([R Core Team, 2021](#)). Además de su uso en la academia, por sus paquetes estadísticos ya instalados por defecto, también tiene popularidad en el campo de Ciencia/Análisis de Datos principalmente por la colección de paquetes de **tidyverse**, dentro de los que destaca **ggplot2**. Este paquete de **R** produce visualizaciones de datos bajo la **Gramática de Gráficos**, que permite la gráfica con componentes independientes. Esta empieza con la premisa de las **capas** o *layers* y que un gráfico asigna los datos a los atributos estéticos (color, forma, tamaño) de los objetos geométricos (puntos, líneas, barras) ([Wickham, 2016](#)). A lo largo de este trabajo se tiene el apoyo de este lenguaje de programación en conjunto con la librería antes mencionada (**ggplot2**).

4.3.3. Uso del API de Google Earth Engine

Se mencionó con anterioridad en la Sección 4.2 sobre la plataforma de Google Earth Engine, los motivos de su creación, algunas de sus aplicaciones en diferentes investigaciones

y sobre el acceso a sus herramientas y bases de datos a través de su propia API.

Existe una serie de pasos que seguir para tener los permisos de usar la API, así como los lenguajes de programación y el ambiente en el que se manejan. El permiso de acceso a API de GEE es a través de su sitio web oficial <https://earthengine.google.com>, donde se llenan todos los campos del formulario y se espera a un correo de confirmación. Para el manejo de las herramientas de GEE existen dos lenguajes de programación con sus respectivos ambientes de trabajo:

- **JavaScript**: Es un lenguaje de programación ligero, interpretado y orientado a objetos, principalmente se enfoca en el desarrollo de páginas web ([MDN contributors, 2021](#)). Es el principal lenguaje de programación para la manipulación y visualización de datos en el **Editor de Código de GEE**, el principal IDE de GEE ([Google Developer Team, 2021a](#)). De las principales ventajas de su uso son:
 - El conocimiento de JavaScript no es alto, ya que entendiendo el flujo de trabajo, declaración de variables, funciones y estructuras de datos se pueden realizar una amplia variedad de tareas ([Google Developer Team, 2021g](#))
 - La visualización de mapas, la manipulación de ellos (como dibujar geometrías o cambiar configuraciones en las capas de información), acceso a la documentación, búsqueda e importación de los conjuntos de datos es bastante intuitiva dentro del Editor de Código de GEE ([Google Developer Team, 2021a](#)).
- **Python**: Las funcionalidades del Editor de Código de GEE se ven limitadas para este lenguaje de programación, principalmente las visualizaciones de los mapas y las gráficas. Sin embargo, existen diferentes alternativas que son de gran apoyo para el uso de la API. Estas están en la forma de librerías de Python:
 - **Earth Engine Python API**: Esta librería puede ser instalada y usada de manera local (en el dispositivo del usuario) o a través de la nube con Colab. ([Google Developer Team, 2021i](#))
 - **geemap**: A pesar de que el Equipo de Desarrolladores de Google proporciona una alternativa a la visualización de los mapas con Folium ([Google Developer Team, 2021i](#)), gracias a **geemap** se tiene una mayor flexibilidad. Esta librería no solo permite la manipulación de las capas de los mapas, también exportar regiones de interés como otro tipo de estructura de datos ([Wu, 2020](#))
 - **eemont**: Esta librería fue desarrollada para facilitar el pre-procesado y procesado de los objetos de GEE, así como expandir el número de opciones de tareas y flujos de trabajo que tanto en el Editor de Código así como con la API de GEE en Python, eran una extensa cantidad de líneas de código, con **eemont** se vuelven tareas cortas ([Montero, 2021](#)).

Cabe mencionar que el uso del Python API con apoyo de las librerías dedicadas a manipular datos de GEE, existe una gran interacción e integración de estos datos en flujos de trabajo de Ciencia/Análisis de Datos desarrollados en Python.

Capítulo 5

Metodología

5.1. Región de estudio

El estado de Yucatán se encuentra al sureste del país, colindando con el Golfo de México (noroeste), Quintana Roo (sureste) y Campeche (suroeste) (INEGI, 2017). La superficie del estado representa aproximadamente el 2% del total del territorio mexicano, siendo esto un área de 39 524.4 km² (INEGI, 2021b).

El uso que se le da a la superficie en un 96.1 % es para terrenos de agricultura, zonas urbanas, áreas sin vegetación, áreas con vegetación inducida o vegetación secundaria y cuerpos de agua (INEGI, 2017); es importante recalcar que la vegetación secundaria es la que se forma después de un disturbio natural o antropogénico (causas humanas, como agricultura) (Duran García y Méndez Gonzales, 2010). El porcentaje restante son regiones cubiertas por vegetación natural. Se puede observar la distribución de vegetación en la Figura 5.1.

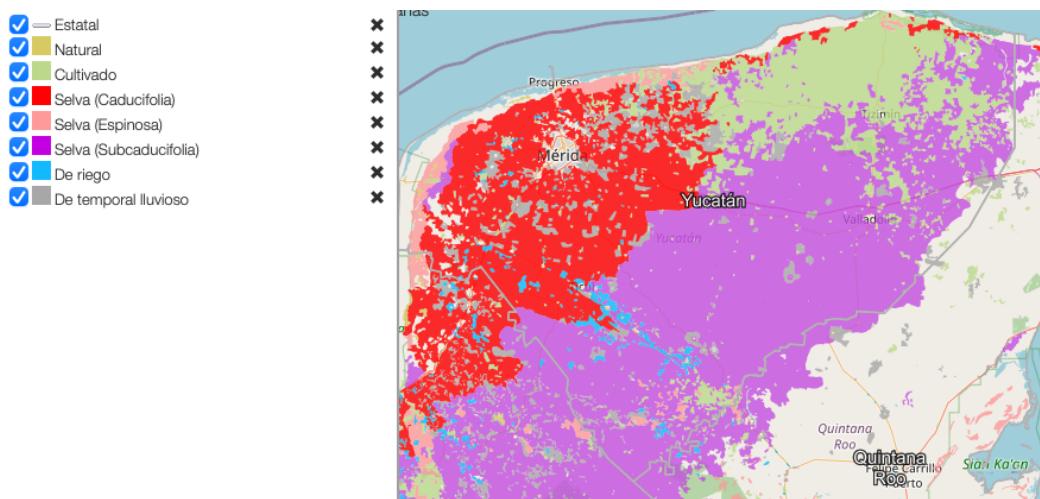


Figura 5.1: Captura de pantalla del mapa de vegetaciones del estado de Yucatán. Fuentes: (INEGI, 2021a)

El clima característico del estado de Yucatán es **cálido subhúmedo**, siendo el que más domina viendo el mapa de la Figura 5.2, sin embargo se cuenta con climas secos en las costas, en el norte del estado; siendo la temperatura anual promedio de 26 C° (INEGI, 2017)

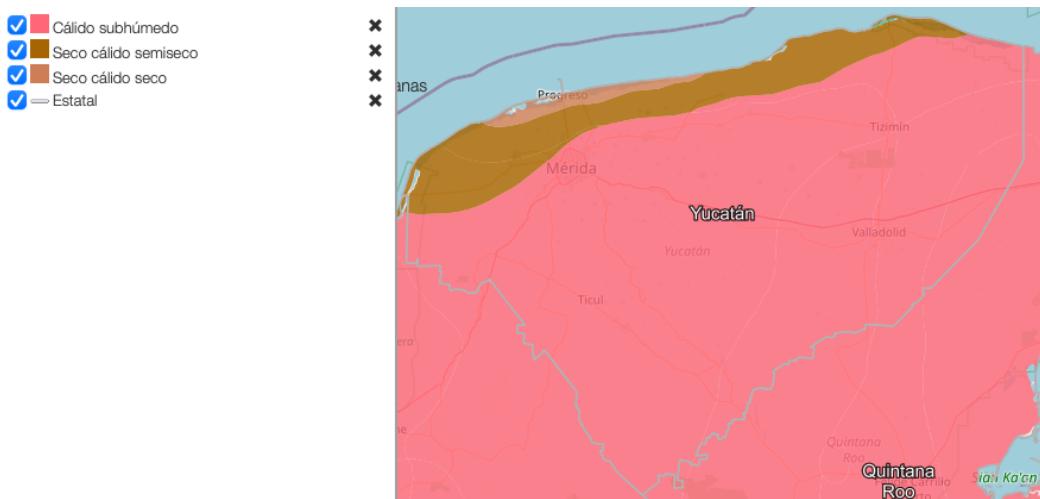


Figura 5.2: Captura de pantalla del mapa de climas del estado de Yucatán. Fuentes: (INEGI, 2021a)



Figura 5.3: Información sobre incendios forestales. Fuente: <https://datos.gob.mx/busca/organization/conafor>

5.2. Obtención de datos

5.2.1. Datos nacionales: CONAFOR

La **Comisión Nacional Forestal (CONAFOR)**, creada por decreto presidencial el 4 de abril del 2001, es un Organismo Público Descentralizado cuyo objetivo es desarrollar, favorecer e impulsar las actividades productivas, de conservación y restauración en materia forestal, así como participar en la formulación de los planes, programas, y en la aplicación de la política de desarrollo forestal sustentable (Gobierno de México, s.f.).

La obtención de la relación de incendios forestales se encontraron en la plataforma de **Datos Abiertos de México**¹. Se ingresó la institución encargada de esos datos (la CONAFOR) y se encontró el apartado **Incendios forestales** como se muestra en la Figura 5.3. Al descargar los archivos solamente se contaba con un archivo relevante llamado **Serie_histórica_anual_incendios_(2017).xlsx** con únicamente los registros de los incendios forestales del año 2017.

¹Sitio web oficial: <https://datos.gob.mx>

5.2.2. Datos de la NASA: FIRMS

La **Información sobre Incendios para el Sistema de Gestión de Recursos o FIRMS** (por sus siglas en inglés) distribuyen información de fuegos activos en Tiempo Cercaño al Real (NRT, por sus siglas en inglés) a través los satélites (Aqua y Terra) del sistema **MODIS** (*Moderate-Resolution Imaging Spectroradiometer*) y **VIIRS** (*Visible Infrared Imaging Radiometer Suite*) (NASA, s.f.).

Los datos que se pueden solicitar son las series históricas de puntos de calor en la superficie terrestre. La solicitud se hace a través de su sitio web oficial de **FIRMS**² dentro del apartado de *Archive Download*, en esta sección se hace una Petición (*Request*) especificando el área de interés, la fuente de la información (sistema de satélites), el periodo de tiempo, el tipo de archivo y un correo electrónico.

Se realizaron tres peticiones, todas compartían los siguientes campos:

- **Country:** Mexico
- **Fire Source:** MODIS
- **From:** 2000-01-01
- **To:** 2020-12-31

y diferían en el campo del tipo de formato de los datos, donde se solicitaron:

- Texto Separado por Comas (.csv)
- Shapefile (.shp)
- JSON (.json)

El tiempo de respuesta a las peticiones varía dependiendo de la cantidad de datos que se solicitan (que dependen de la región de interés, periodo de tiempo y la fuente de información) y formato del archivo, para este trabajo el tiempo de respuesta de cada una de las peticiones no pasaba de los 15 minutos. Los tres archivos contienen la misma cantidad de información y los mismo atributos, el archivo con carga de datos más rápida fue el archivo .csv.

5.2.3. Imágenes satelitales: Google Earth Engine Python API

Como se menciona en la Sección 4.2.1, GEE contiene un catálogo de información y datos de diferentes sistemas de satélites, tales como MODIS, Landsat, Sentinel, etc; así como diferentes tipos de variables ambientales, climatológicas y de vegetación.

Para este trabajo se toma como referencia la metodología así como los descubrimientos de trabajos como los de (Sulova y Jokar Arsanjani, 2021) y (Seydi, Akhoondzadeh, Amani, y Mahdavi, 2021), los cuales están relacionados a la exploración de las imágenes satelitales de

²Sitio web oficial: <https://firms.modaps.eosdis.nasa.gov>

la plataforma de GEE y su implementación en flujos de trabajo de Aprendizaje Automático (*Machine Learning*).

Los sistemas satelitales que se usarán serán los que proporciona **MODIS**, donde en el Cuadro 5.1 se especifica el nombre del conjunto de datos y las bandas donde se encuentran. Se destacan los **índices de vegetación** ya que han sido de amplio uso debido a que son conceptos simples y computacionalmente eficientes. En el Cuadro 5.2, se muestra la expresión matemática de los índices expresados con las bandas de las imágenes satelitales.

Siguiendo la línea de factores relevantes en la predicción de incendios forestales, variables como la temperatura, humedad e índice de sequía también han sido variables importantes en modelos predictivos de detección de incendios forestales ([Sulova y Jokar Arsanjani, 2021](#)).

Categoría	Nombre en el catálogo	Variable(s)	Banda(s) de datos	Resolución
Ambientales y climatológicas	MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1km	Temperatura máxima	LST_Day_1km	1000 m
Ambientales y climatológicas	TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho	Humedad del suelo e Índice de Gravedad de la Sequía de Palmer	soil y psdi	4638.3 m
Índices de vegetación	MOD13Q1.006 Terra Vegetation Indices 16-Day Global 250m	NDVI y EVI	NDVI y EVI	250 m
Registro mensual de incendios forestales	FireCCI51: MODIS Fire_cci Burned Area Pixel Product, Version 5.1	Área afectada (píxeles)	BurnDate	250 m

Cuadro 5.1: Listado de variables y conjuntos de datos extraídos del catálogo de datos de Google Earth Engine

Índice de vegetación	Usos	Fórmula (Landsat)
NDVI	El Índice de Vegetación de Diferencia Normalizado (<i>Normalized Difference Vegetation Index</i> , en inglés) se usa para detectar y monitorear vegetación sana y su extensión y densidad. Para su cálculo se usan las bandas Rojo (Red) e Infrarrojo Cercano (NIR)(Borra et al., 2019)	$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}$
EVI	El Índice de Vegetación Mejorado (<i>Enhanced Vegetation Index</i> , en inglés) es similar al NDVI y también se usa para cuantificar que tan sana o verde es la vegetación. Sin embargo, EVI hace correcciones para algunas condiciones atmosféricas y ruido del follaje de fondo y es sensible a áreas con vegetación densa. Incorpora nuevos parámetros como la “L” ³ para el ajustar el ruido, las “C” ⁴ para las condiciones atmosféricas y valores de la banda Azul (Blue)(U.S. Geological Survey, s.f.).	$\text{EVI} = 2.5 \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{C}_1 \cdot \text{Red} - \text{C}_2 \cdot \text{Blue} + \text{L}}$

Cuadro 5.2: Lista de Índices de Vegetación de Landsat 8

5.3. Filtro de datos

5.3.1. Rango temporal de estudio

Todos los conjuntos de datos cuentan con diferentes períodos temporales por lo que para tener la mejor calidad de datos es importante tener un rango de fecha donde los datos se encuentren y sean relevantes, dando una mayor prioridad a los datos extraídos de GEE.

Se proponen tres diferentes rangos de trabajo para los diferentes tipos de datos:

- **De 2001-01-01 a 2020-12-31 (20 años):** El rango de tiempo más grande para analizar. En este periodo se pueden estudiar los registros proporcionados por FIRMS y

³Para el caso del sistema MODIS, L = 1

⁴Para el caso del sistema MODIS, C₁ = 6 y C₂ = 7.5

gran parte de los datos de registro de incendios forestales obtenidos en GEE, así como lo datos proporcionados por la CONAFOR.

- **De 2001-01-01 a 2019-12-31 (19 años):** Dentro de este periodo de tiempo se cuentan los registros de incendios forestales proporcionados por el conjunto de datos encontrado en GEE, por lo que también las variables ambientales y de vegetación obtenidos a través de la misma plataforma serán ajustados a este periodo de tiempo.
- **De 2017-01-01 a 2017-12-31 (1 año):** El periodo más corto con el que se trabaja. En este año se encuentran los registros de la CONAFOR.

5.3.2. Región de Interés

Tal como se menciona en la Sección 5.1, el estado de interés es Yucatán. Para que los datos puedan ser manipulados con la seguridad de que se encuentren dentro de esta superficie específica hace falta contar con los datos vectoriales de la división política de México, es decir, los límites de las entidades federativas.

En el sitio web oficial del **Instituto Nacional de Estadística y Geografía (INEGI)**⁵, en el apartado de **Datos**, en tema de **Marco Geoestadístico**, en la sección de **Descargas**, el título **Marco Geoestadístico. Censo de Población y Vivienda 2020** tiene dentro información de todas las entidades federativas y se selecciona descargar la información de Yucatán⁶. El archivo cuenta con diferentes carpetas: **catalogos**, **conjuntos_de_datos** y **metadatos**. La información del polígono del estado de Yucatán se encuentra dentro de la segunda carpeta bajo el nombre **31ent.shp**.

5.3.3. Selección de variables

Conjunto de datos de la CONAFOR

Las variables que se usan de este conjunto de datos son las siguientes ⁷:

- **Información geográfica:** ‘Grados’, ‘Minutos’, ‘Segundos’, ‘Grados.1’, ‘Minutos.1’, ‘Segundos.1’
- **Información de tiempo:** ‘Fecha Inicio’, ‘Duración días’
- **Información de la superficie afectada:** ‘Clave municipio’, ‘Estado’, ‘Municipio’, ‘ANP’, ‘Tipo de incendio’, ‘Tipo impacto’, ‘Tipo Vegetación’, ‘Causa’, ‘Causa específica’, ‘Superficie Salvada’ y ‘Total’.

⁵Sitio web oficial: <https://www.inegi.org.mx>

⁶Este proceso se lleva a cabo a la fecha de la escritura de este trabajo, Noviembre 2021.

⁷Los nombres son de las columnas del archivo CSV de la CONAFOR

Conjunto de datos de FIRMS

El catálogo o diccionario de atributos para este conjunto de datos se encuentra en el sitio web oficial del Programa de Sistemas de Datos de Ciencias de la Tierra de la NASA bajo la siguiente liga: <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/mcd14dl>. En el Cuadro 5.3 se encuentran los atributos que se usan así como sus nombres⁸:

Atributo	Columna	Descripción
Latitud	latitude	Centro del píxel de fuego de 1 km. Se pueden detectar más de un incendio dentro de un píxel, no necesariamente es la ubicación real.
Longitud	longitude	Centro del píxel de fuego de 1 km. Se pueden detectar más de un incendio dentro de un píxel, no necesariamente es la ubicación real.
Fecha de adquisición	acq_date	Fecha de adquisición
Hora de adquisición	acq_time	Hora de adquisición
Confianza	confidence	Confianza de que la observación haya sido detectada como un incendio o fuego. El rango es entre 0 y 100 %.
Temperatura de Luminosidad 21	brightness	Temperatura de la luminosidad del píxel de acuerdo con el Canal 21/21. Su unidad es Kelvin.
Temperatura de Luminosidad 31	brightness	Temperatura de la luminosidad del píxel de acuerdo con el Canal 31. Su unidad es Kelvin.
Potencia Radiante del Fuego	fpr	Es la Potencia Radiante del Fuego (FPR, por sus siglas en inglés) de la instancia. Su unidad es en Megawatts (MW).
Día o noche	daynight	Si la instancia fue detectada de día o de noche

Cuadro 5.3: Listado de atributos a usar y su descripción. Datos FIRMS. Fuente: ([Earth Science Data Systems Program, 2021](#))

Imágenes satelitales de GEE

Para el caso de las imágenes de GEE las variables fueron seleccionadas al momento de buscar y cargar los datos así como las bandas que se usarán cada conjunto. Esta información se puede ver en el Cuadro 5.1

⁸Los nombres de las columnas del archivo CSV de FIRMS

5.4. Carga, limpieza, transformación y procesamiento de datos

Cada tipo de conjunto de datos requiere una tratamiento diferente por lo que en esta sección se muestra una visión general de lo que se hace con cada uno de ellos.

- **Conjunto de datos de la CONAFOR:** Filtrar por entidad federativa, obtener atributos relevantes, revisar por valores nulos o extraños en los atributos, procesar unidades o tipos de datos de los atributos y agrupar los puntos por mes.
- **Conjunto de datos de FIRMS:** Filtrar por región de estudio, filtrar por fecha, obtener los atributos relevantes, revisar por valores nulos o extraños en los atributos, procesar unidades o tipos de datos de los atributos y agrupar los puntos por mes.
- **Imágenes satelitales de GEE:** Filtrar por región de estudio, filtrar por fecha, procesar unidades de los atributos, agrupar los registros por mes, transformar datos ráster a datos vectoriales.

La acción de “procesar” puede incluir tareas como la conversión de unidades, estandarización o normalización de valores y hasta la creación de nuevos atributos.

5.5. Análisis Exploratorio de Datos

Previo a la década de los 60, la estadística clásica se enfocaba primeramente en la **inferencia**, un área que a veces cuenta con procesos complejos para extraer conclusiones sobre poblaciones grandes basándose en muestras pequeñas. Sin embargo, en el año 1962, John W. Tukey publicó un artículo llamado *“The Future of Data Analysis”*, en donde propone una nueva disciplina científica llamada **Análisis de Datos** que incluía a la inferencia estadística como un componente. El campo de **Análisis Exploratorio de Datos** fue establecido tiempo después, en 1977 con su libro (ahora un clásico) *“Exploratory Data Analysis”* al presentar visualizaciones de datos simples (por ejemplo, gráfico de caja y bigotes, gráficos de dispersión, etc.) de la mano con métricas estadísticas (media, mediana, cuantiles, etc.), ayudan a dibujar la forma y estructura del conjunto de datos (Bruce, Bruce, y Gedeck, 2020).

5.5.1. Estimados de locación

También llamados **medidas de tendencia central**, son los primeros valores que se pueden extraer de los datos de manera directa y sencilla. Estos estimados reflejan donde se encuentran la gran mayoría de los datos o hacia donde tienden y son de ayuda para caracterizar el conjunto de datos.

Media

El estimado de locación más básico es la **media** o el valor *promedio*. Este estimado es la suma de todos los valores entre el número de valores.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5.1)$$

A pesar de que este estimado es de gran utilidad, es sensible a los **datos anómalos** o **anomalías**. Para estas situaciones existen otros estimados que son solución ante esta situación: **mediana** y **media truncada**.

Mediana

La **mediana** es el número del centro en una lista ordenada de datos. Si hay una cantidad par de datos, el valor del centro no se encuentra *dentro* del conjunto de datos, sino es el promedio de los dos valores que dividen los datos. A comparación de la media, la mediana depende únicamente de los valores en el centro de los datos ordenados (Bruce et al., 2020).

Media truncada

Una variación de la media es la **media truncada**. Este estimado “elimina” una porción de los datos. Si los datos ordenados son $x_1, x_2, x_3, \dots, x_n$ donde x_1 es el valor más pequeño y x_n es el más grande, la media truncada omite los p valores más pequeños y más grandes. La proporción de los datos omitidos para este cálculo es $p = 0.05$

$$\bar{x}_{\text{trim}} = \frac{1}{n - 2p} \sum_{i=p+1}^{n-p} x_i \quad (5.2)$$

5.5.2. Estimados de variabilidad

Mientras que los estimados de locación resumen la posición donde se encuentran las instancias de un conjunto de datos es solo una dimensión de caracterización. La segunda se da con los **estimados de variabilidad**, estos expresan la dispersión de los datos alrededor de los estimados de locación (Bruce et al., 2020).

Varianza y desviación estándar

Los estimados de variabilidad más usados están basados en las diferencias o *desviaciones* entre los estimados de locación y los datos del conjunto. Dentro de los más conocidos se encuentran la **varianza** y la **desviación estándar**. La varianza es el promedio de esas desviaciones y la desviación estándar es la parte positiva de la raíz cuadrada de la varianza (Bruce et al., 2020).

$$\text{Varianza} = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5.3)$$

$$\text{Desviación estándar} = \sqrt{s^2} = s$$

Correlación

Durante el Análisis Exploratorio de Datos en todas las ocasiones se trata de buscar relaciones entre las variables, sea entre atributos predictores o atributos predictores y atributos objetivo. Se dice que un par de atributos θ y φ están correlacionados cuando dada la información que proporcione uno se puede encontrar el del otro. Están correlacionados de manera positiva cuando altos valores de θ van con altos valores de φ y viceversa; y están correlacionados de manera negativa cuando altos valores de θ van con bajos valores de φ y viceversa.

El **coeficiente de correlación de Pearson** es un estimador de la correlación muy usado. Este coeficiente tiene un rango de 1 hasta -1, siendo 1 una correlación positiva perfecta, -1 una correlación negativa perfecta y 0 significa que no hay correlación alguna.

$$r = \frac{\sum_{i=1}^n (\theta_i - \bar{\theta})(\varphi_i - \bar{\varphi})}{(n-1)s_\theta s_\varphi} \quad (5.4)$$

Momentos estadísticos

Los momentos estadísticos métricas relacionadas a la forma de una distribución o en este caso, la forma del conjunto de datos ([Zwillinger, 2018](#)).

El k -ésimo momento (m_k) se define como:

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k \quad (5.5)$$

mientras que el k -ésimo **momento centrado** (a la media) (μ_k) se define como:

$$\mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad (5.6)$$

Coeficiente de asimetría

El **coeficiente de asimetría** también conocido como *skewness* mide la asimetría en el conjunto de datos. Si el valor del coeficiente es positivo, significa que hay muchos más datos dispersos hacia la derecha (cola de la distribución hacia la derecha), y si es negativo significa que hay muchos más datos dispersos hacia la izquierda (cola de la distribución hacia la izquierda) ([Zwillinger, 2018](#)).

Con los momentos estadísticos previamente definidos, el **coeficiente de asimetría** (γ_1) se define como:

$$\gamma_1 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} \quad (5.7)$$

5.5.3. Prueba de hipótesis

Como se mencionó con anterioridad, la Inferencia Estadística es un componente del Análisis Exploratorio de Datos, donde el término **inferencia** refleja la intención de aplicar los resultados de un experimento, donde se cuentan con datos limitados, a un proceso mayor o población (Bruce et al., 2020).

Las **Pruebas de hipótesis** también conocidas como pruebas de significancia estadística tienen como propósito ayudar a encontrar si efectos aleatorios son responsables de los efectos observados (Bruce et al., 2020).

A pesar de que existen fórmulas para llevar a cabo pruebas de hipótesis, existen procedimientos computacionales que son igual de útiles. Con estos procesos computacionales se evitan supuestos que pueden sesgar los resultados (suponer que se conocen uno o más estimados de locación y variabilidad poblacional).

Prueba de permutación

En este proceso, dos o más muestras están involucradas, normalmente en grupos A y B o alguna otra prueba de hipótesis. **Permutar** significa cambiar el orden de los valores por lo que el proceso involucra un remuestreo de los conjuntos de datos, el proceso es el siguiente (Bruce et al., 2020):

1. Combinar los conjuntos de datos (resultados, observaciones o valores) de ambos grupos en un solo arreglo de datos.
2. Mezclar las instancias y seleccionar de manera **aleatoria y sin repetición** un conjunto de datos del tamaño del grupo A.
3. De los datos sobrantes, del conjunto mayor, seleccionar de manera **aleatoria y sin repetición** un conjunto de datos del tamaño del grupo B.
4. Sea el caso de la métrica o el estimado calculado de las muestras originales, hacer el cálculo para estas nuevas muestras.
5. Registrar el valor, esto es una iteración.
6. Repetir los pasos 2 al 5 una N cantidad de veces para así crear un histograma de la distribución de los resultados de las iteraciones.

Con el histograma creado, encontrar la métrica o estimado del conjunto de datos originales. El **valor-p** de la prueba es la proporción de métricas generadas por la prueba de permutación que sean mayores o menores que, según sea el caso, la métrica original de observación. De manera visual, el valor-p es el área del histograma entre la métrica observada y el *final* del histograma (sea hacia la izquierda o derecha, según sea el caso).

Se tiene de ejemplo al grupo A con una media de 130 unidades y al grupo B con una media de 108 unidades. Se postulan las siguientes hipótesis:

- Hipótesis nula (H_0): El promedio del grupo A es mayor que al conjunto B. Es decir, $\text{media}(A) - \text{media}(B) > 0$
- Hipótesis alternativa (H_1): El promedio del grupo A es menor que al conjunto B. Es decir, $\text{media}(A) - \text{media}(B) \leq 0$

Después de realizar diez mil iteraciones de los pasos 2 al 5 de la lista anterior se obtiene la Figura 5.4. Donde el valor p es estadísticamente significativo a un nivel de 0.05 y por lo tanto, se tiene suficiente evidencia para rechazar la hipótesis nula.

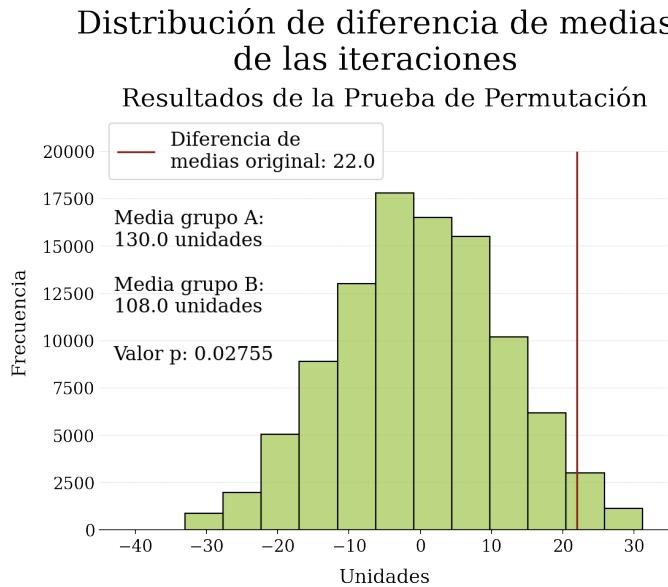


Figura 5.4: Ejemplo de Prueba de Permutación

5.5.4. Visualización de datos

La visualización de datos es la presentación de información (datos) en un contexto visual, la mayoría de las veces a través de un gráfico, por tanto, un gráfico siempre debe tener en cuenta el tipo y el propósito de los datos. Cierta información se adapta mejor a un tipo de gráfico que a otro: por ejemplo, un gráfico de barras en lugar de un gráfico circular. Pero con la mayoría de las herramientas, el usuario tiene una amplia variedad de opciones de herramientas visuales , desde representaciones comunes como gráficos de líneas y gráficos de barras hasta líneas de tiempo, mapas, diagramas, histogramas y diseños personalizados.

Histograma

Un histograma visualiza la distribución de los datos a lo largo de un intervalo continuo o un período de tiempo determinado. Cada barra en un histograma representa la frecuencia tabulada en cada intervalo/bin. El área total del histograma es igual al número de datos.

Los histogramas ayudan a dar una estimación de dónde se concentran los valores, cuáles son los extremos y si hay vacíos o valores inusuales. También son útiles para dar una visión

aproximada de la distribución de probabilidad. Se puede observar un ejemplo en la Figura 5.5.

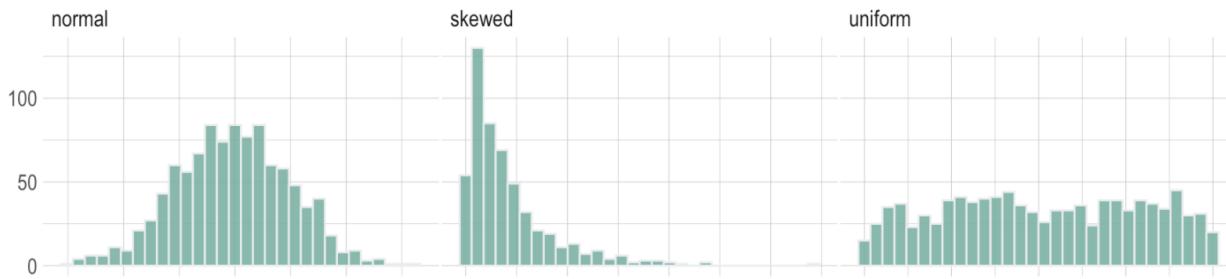


Figura 5.5: Ejemplos de distribuciones representadas en histogramas. Fuente: ([Holtz, s.f.-c](#))

Gráfico de densidad

Un gráfico de densidad visualiza la distribución de datos en un intervalo o período de tiempo continuo. Este gráfico es una variación de un Histograma que usa un suavizado computado por un kernel de densidad estimada, permitiendo distribuciones más suaves al suavizar el ruido ([Bruce et al., 2020](#)). Una ventaja que los gráficos de densidad tienen sobre los histogramas es que son mejores en la determinación de la forma de la distribución, al no verse afectados por el número de contenedores o barras usadas en los histogramas típicos, teniendo en cuenta que un histograma normal que conste de sólo 4 barras no produciría una forma lo suficientemente distinguible de la distribución como otro de 20 por ejemplo. Sin embargo con gráficos de densidad esto no es un problema. Se puede observar un ejemplo en la Figura 5.6.

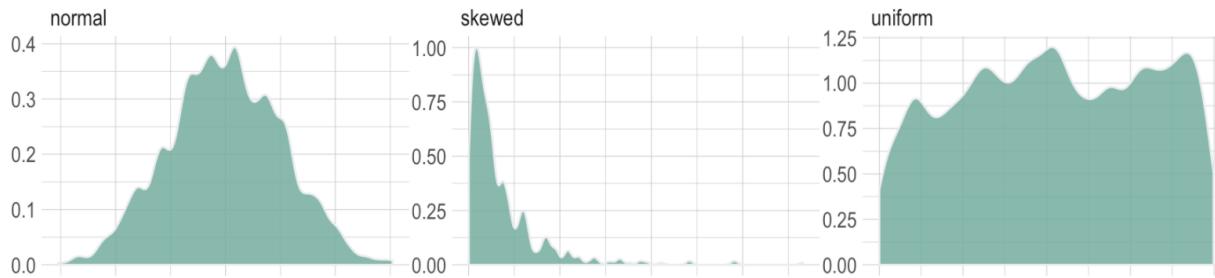


Figura 5.6: Representación de diferentes distribuciones a través de un gráfico de densidad. Fuente: ([Holtz, s.f.-b](#))

Ridgeline plot

Un gráfico Ridgeline o Joyplot muestra la distribución de un valor numérico para varios grupos. Pueden ser muy útiles para visualizar los cambios en las distribuciones a lo largo del tiempo o del espacio. La distribución puede representarse mediante histogramas o gráficos de densidad, todos alineados a la misma escala horizontal y presentados con un ligero solapeamiento. Las gráficas son especialmente interesantes cuando el número de grupos a representar es alto y, por lo tanto, una separación clásica de ventanas requeriría de mucho más espacio ([Holtz, s.f.-f](#)). Se puede observar un ejemplo en la Figura 5.7.

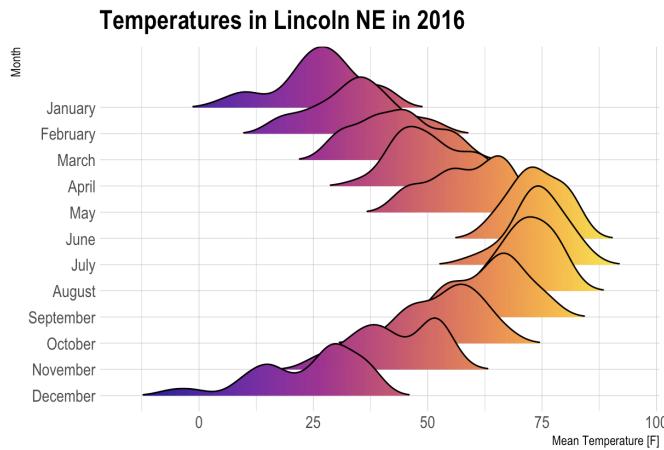


Figura 5.7: Ejemplo de un gráfico Ridgeline con las temperaturas mensuales de la ciudad de Lincoln, Nebraska, EE.UU. ([Holtz, s.f.-f](#))

Series de tiempo

Una serie de tiempo es una forma gráfica de mostrar una lista de eventos en orden cronológico. Algunas series de tiempo funcionan en una escala, mientras que otras simplemente muestran los eventos en secuencia. La función principal de una serie de tiempo es comunicar información relacionada con el tiempo, ya sea para el análisis o para presentar visualmente una historia o una visión de la historia. Si se basa en una escala, una línea de tiempo le permite ver cuándo ocurren o van a ocurrir las cosas, permitiendo al espectador evaluar los intervalos de tiempo entre los eventos. Esto permite al espectador ver cualquier patrón que aparezca en cualquier período de tiempo seleccionado o cómo se distribuyen los eventos en ese período de tiempo. Se puede observar un ejemplo en la Figura 5.8.

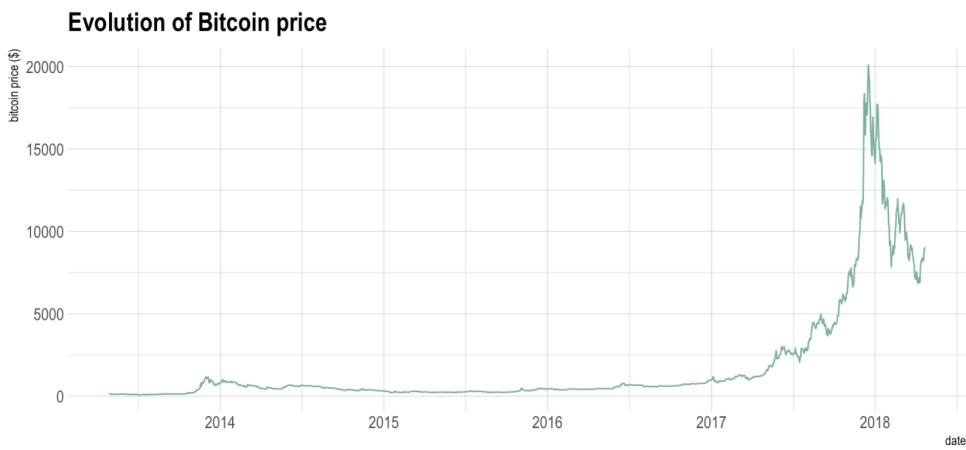


Figura 5.8: Ejemplo del valor del Bitcoin entre 2013 y 2018. Fuente: ([Holtz, s.f.-d](#))

Gráficos de barras

El gráfico de barras utiliza barras horizontales o verticales (gráfico de columnas) para mostrar comparaciones numéricas discretas entre categorías. Un eje del gráfico muestra las categorías específicas que se comparan y el otro eje representa una escala de valores discretos.

Los gráficos de barras se distinguen de los Histogramas, ya que no muestran desarrollos continuos durante un intervalo. Los datos discretos de los gráficos de barras son datos categóricos y por lo tanto responden a la pregunta de «¿cuántos?» En cada categoría.

Un fallo importante con los gráficos de barras es el etiquetado, que se convierte en problemático cuando hay un gran número de barras. Se puede observar un ejemplo en la Figura 5.9.

Diagrama *Lollipop* o *Lollipop plot*

Un diagrama Lollipop es básicamente una gráfica de barras, donde la barra se transforma en una línea y su extremo en un punto. Muestra la relación entre una variable numérica y una categórica. Sin embargo, es más atractivo y transmite mejor la información. Es especialmente útil cuando se tienen muchas barras (Holtz, s.f.-e). Se puede observar un ejemplo en la Figura 5.10.

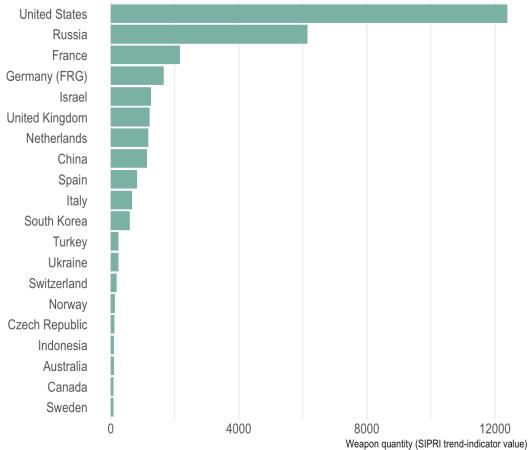


Figura 5.9: Representación de un conjunto de datos en una gráfica de barras (Holtz, s.f.-a)

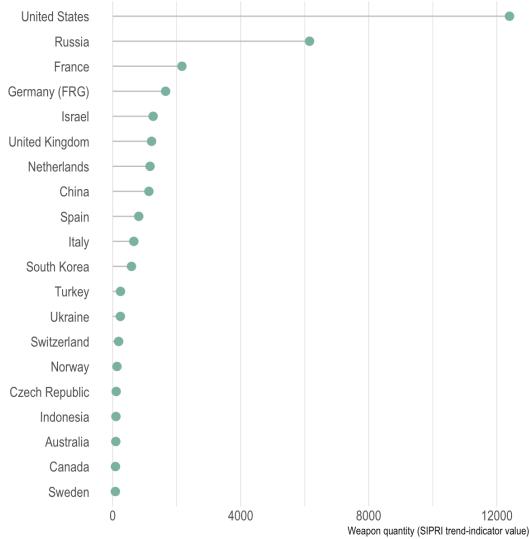


Figura 5.10: Representación de los datos previamente representados en una gráfica de barras en 5.9 a un un diagrama Lollipop (Holtz, s.f.-e)

Mapas de calor

Los mapas de calor visualizan los datos a través de variaciones en el color. Cuando se aplican a un formato tabular, los mapas de calor son útiles para el examinar cruces de variables en conjuntos multivariados, mediante la colocación de variables en las filas y columnas, y coloreando las celdas dentro de la tabla. Los mapas de calor son buenos para mostrar la varianza a través de múltiples variables, revelando cualquier patrón, mostrando si las variables son similares entre sí y para detectar si existen correlaciones entre ellas ([Ribecca, s.f.-c](#)).

Normalmente, todas las filas serían una categoría (las etiquetas se mostrarán a la izquierda o derecha) y todas las columnas asignadas serían otra (etiquetas mostradas en la parte superior o inferior). Las filas y columnas individuales se dividen en subcategorías, que coinciden entre sí en una matriz. Las celdas contenidas en la tabla contienen datos categóricos codificados por color o datos numéricos que se basan en una escala de color. Los datos contenidos dentro de una celda se basan en la relación entre las dos variables en la fila y columna de conexión ([Ribecca, s.f.-c](#)).

Se requiere una leyenda junto a un mapa de calor para que se pueda leer correctamente. Los datos categóricos están codificados por colores, mientras que los datos numéricos requieren una escala de color que se mezcla de un color a otro para representar la diferencia en valores altos y bajos. Puede utilizarse una selección de colores sólidos para representar un rango múltiple de valores (0-10, 11-20, 21-30, etc.) o una escala de gradiente para un solo rango (por ejemplo 0 - 100) mezclando dos o más colores juntos ([Ribecca, s.f.-c](#)).

Debido a su dependencia del color para comunicar valores, los mapas de calor son un gráfico más adecuado para mostrar una visión más generalizada de los datos numéricos, ya que es más difícil distinguir con precisión las diferencias entre tonos de color y extraer puntos de datos específicos (a menos que incluyan los datos sin procesar en las celdas) ([Ribecca, s.f.-c](#)). Se puede observar un ejemplo en la Figura 5.11.

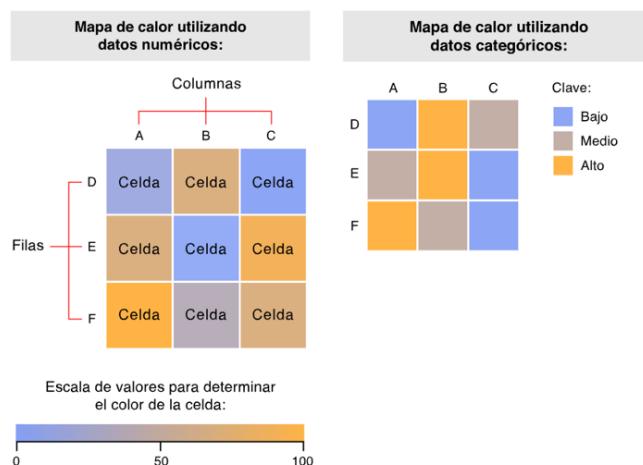


Figura 5.11: Elementos de un mapa de calor con datos numéricos y categóricos. Fuente: ([Ribecca, s.f.-c](#))

Diagramas de cajas y bigotes (*Box plot*)

Un diagrama de cajas y bigotes es una manera conveniente de mostrar visualmente grupos de datos numéricos a través de sus cuartiles. Las líneas que se extienden paralelas a las cajas se conocen como «bigotes», y se usan para indicar variabilidad fuera de los cuartiles superior e inferior. Los valores atípicos se representan a veces como puntos individuales que están en línea con los bigotes. Los diagramas de cajas y bigotes se pueden dibujar vertical u horizontalmente (Ribecca, s.f.-a).

Normalmente utilizado en estadísticas descriptivas, los gráficos de cajas y bigotes son una excelente forma de examinar rápidamente uno o más conjuntos de datos gráficamente. Aunque parezcan primitivos en comparación con un Histograma o un Gráfico de Densidad, tienen la ventaja de ocupar menos espacio, lo cual es útil cuando se comparan distribuciones entre muchos grupos o conjuntos de dato.

Aquí están los tipos de observaciones que uno puede hacer al ver un diagrama de cajas y bigotes:

- Cuáles son los valores clave, tales como: la mediana, el percentil 25 medio, etc.
- Si hay valores atípicos y cuáles son sus valores.
- Si los datos son simétricos.
- Que tanto se encuentran agrupados los datos.
- Si los datos están sesgados y si es así, en qué dirección.

Se puede observar un ejemplo en la Figura 5.12.

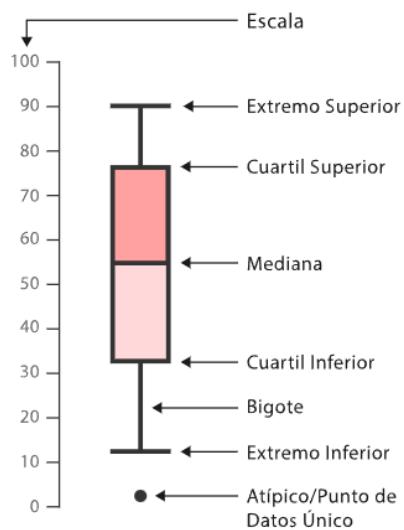


Figura 5.12: Elementos de un diagrama de cajas y bigotes. Fuente: (Ribecca, s.f.-a)

Mapa coroplético

Los mapas coropléticos muestran áreas geográficas divididas o regiones que están coloreadas, sombreadas o modeladas en relación con una variable de datos. Esto proporciona una forma de visualizar valores sobre un área geográfica, que puede mostrar variaciones o patrones en la ubicación visualizada ([Ribecca, s.f.-b](#)).

Estos mapas pueden abordar diferentes situaciones, desde mostrar una variable con determinado crecimiento de magnitud (paleta de colores monocromáticos) hasta la agrupación de instancias similares.

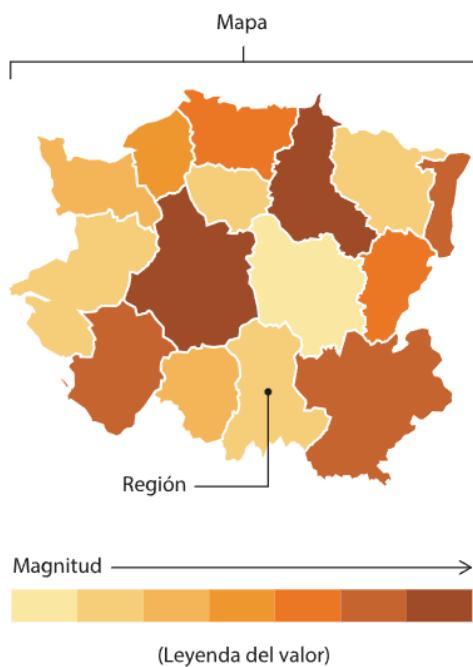


Figura 5.13: Elementos de un mapa coroplético. Fuente: ([Ribecca, s.f.-b](#))

Capítulo 6

Análisis Exploratorio de Datos

A continuación se mostrará el análisis realizado a tres diferentes fuentes de datos, los cuales son los siguientes:

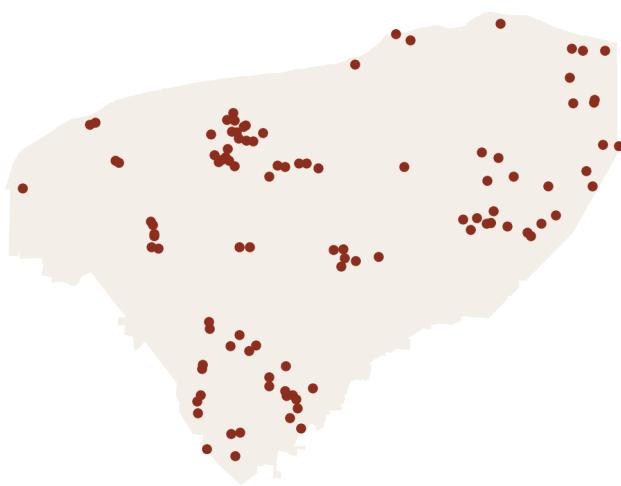
- **Datos de la CONAFOR**
- **Datos de FIRMS**
- **Imágenes satelitales de GEE**

El trabajo se centrará en el estado de Yucatán e inicialmente se mostrará el análisis para cada conjunto de datos y al final se mostrará una comparativa de algunas de las variables en común.

6.1. Datos de la CONAFOR

Este conjunto de datos es el de menor número de muestras entre todos (un total de 103 registros) así como el más limitado en el tiempo (solo cuenta con información del año 2017), sin embargo, es el conjunto de datos con mayor información fuera del aspecto ambiental o climatológico, por ejemplo las causas, tipo de vegetación y la magnitud del incendio forestal.

Ubicación de los incendios forestales en el estado de Yucatán durante el 2017



Registros de la CONAFOR para el estado de Yucatán en el año 2017.

Figura 6.1: **Mapa de puntos.** Ubicación de los incendios forestales registrados en el estado de Yucatán en el año 2017. Información del conjunto de datos de la CONAFOR.

6.1.1. Superficie afectada por los incendios forestales

Dentro de la información que se puede extraer de este conjunto de datos es las hectáreas afectadas por el incendio forestal.

En el Cuadro 6.1 y la Figura 6.2 se encuentran los estimados de locación y variabilidad así como la representación visual del área afectada por los incendios forestales.

Estimado	Valor
Media	106.878 ha
Mediana	20 ha
Media truncada	67.934 ha
Desviación estandar	217.972 ha
Coeficiente de asimetría (γ_1)	3.456 (asimetría por la derecha)

Cuadro 6.1: Estimados de locación y variabilidad del atributo ‘Total’ del conjunto de datos de la CONAFOR

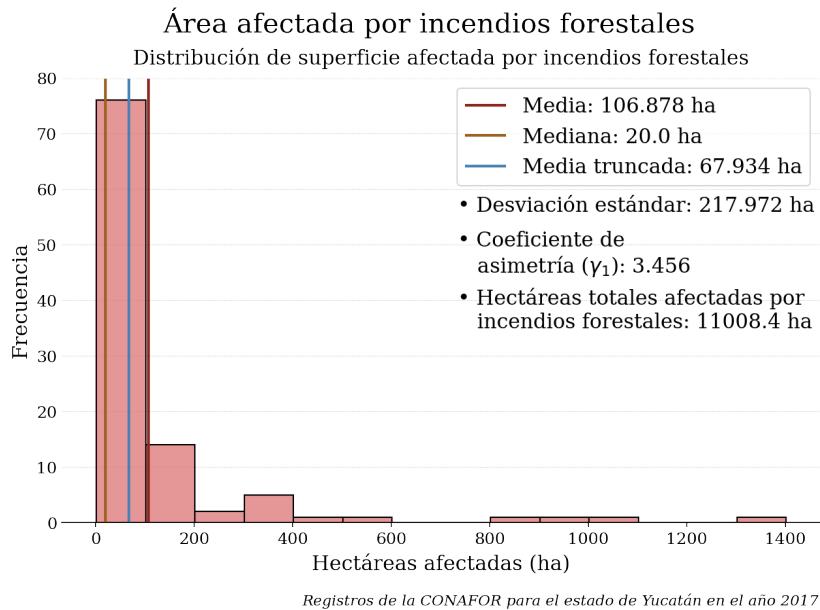


Figura 6.2: **Histograma.** Distribución de área afectada por incendios forestales. Se puede observar que la mayor concentración de datos se encuentran en el primer *bin* del histograma. (El ancho del **bin** es de 100 ha). Información del conjunto de datos de la CONAFOR.

También existen registros de ocurrencias en **Áreas Naturales Protegidas (ANP)** y para estos datos sus estimados se encuentran en el Cuadro 6.2:

Estimado	Valor
Media	528.6 ha
Mediana	213 ha
Desviación estándar	585.096 ha

Cuadro 6.2: Estimados de locación y variabilidad del atributo ‘Total’ para ANP del conjunto de datos de la CONAFOR

No se calcula la media truncada ya que solamente hay 5 instancias (ver Figura 6.3)

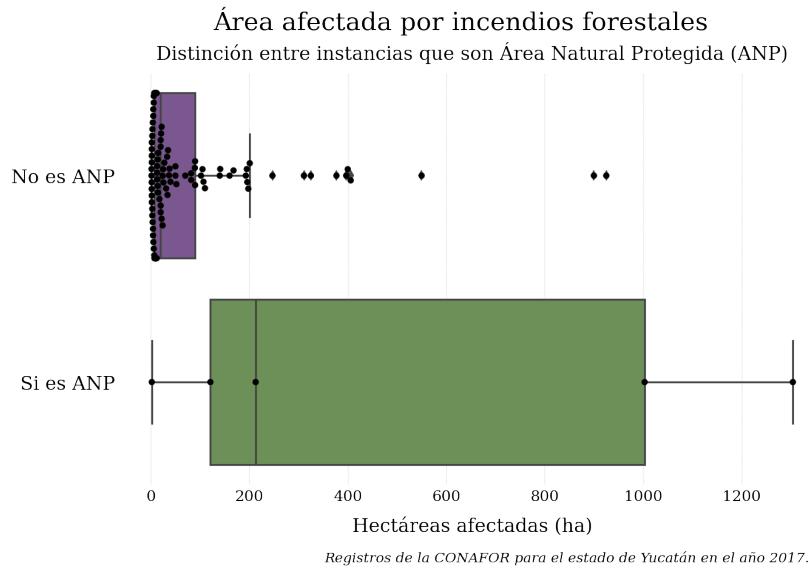


Figura 6.3: Box plot. Distribución de los datos del atributo ‘Total’ separados en dos grupos: si es ANP o no es ANP.

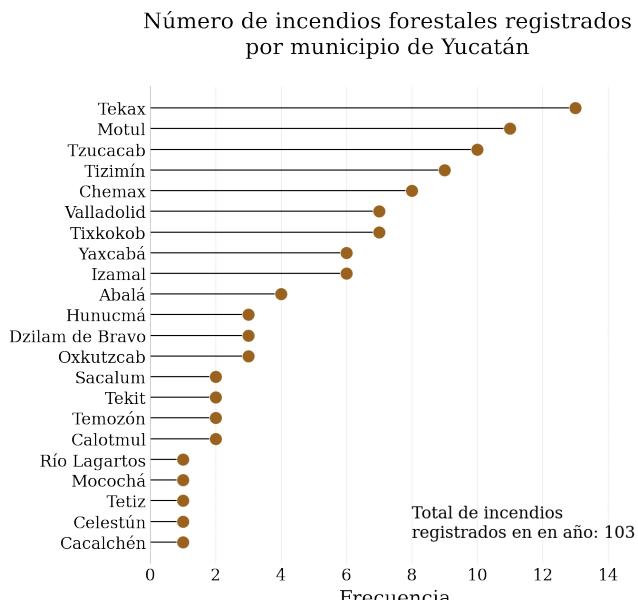
Se observa en la Figura 6.3 que el área superficial de mayor pérdida era una ANP (Ver Cuadro 6.3)

Atributo	Valor
Duración días	4 a 7 Días
Tipo de incendio	Superficial
Tipo impacto	Impacto Moderado
Causa	Otras causas
Causa específica	Desconocidas
Total	1304.0 ha
Municipio	Río Lagartos
ANP	Ría Lagartos

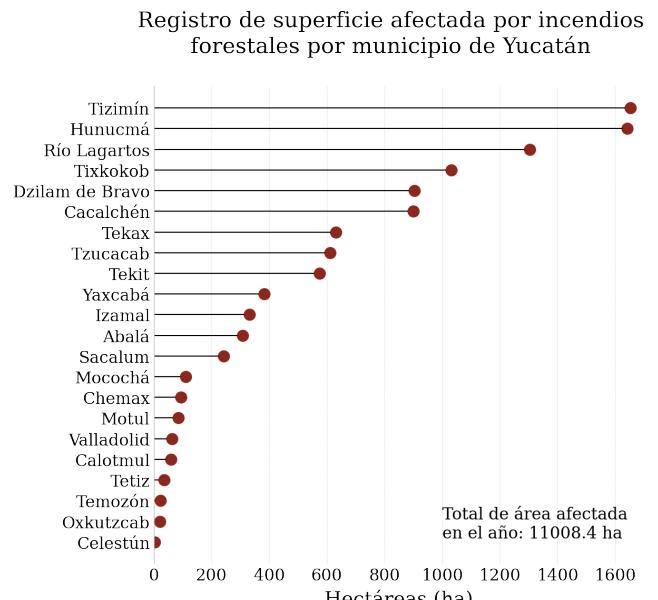
Cuadro 6.3: Atributos del área máxima afectada por incendios forestales del conjunto de datos de la CONAFOR

6.1.2. Municipios afectados

En la Figura 6.4 se puede observar que **Tekax** es el municipio con el mayor número de registro de incendios forestales en el 2017, con el 12% del nivel estatal. Sin embargo no fue el municipio con mayor número de hectáreas afectadas.



Registros de la CONAFOR para el estado de Yucatán en el año 2017.



Registros de la CONAFOR para el estado de Yucatán en el año 2017.

Figura 6.4: **Lollipop plot.** Se muestra el registro de instancias (incendios forestales) durante el 2017 en los municipios de Yucatán. Información del conjunto de datos de la CONAFOR.

Figura 6.5: **Lollipop plot.** Se muestra el registro de la extensión de superficie afectada por incendios forestales durante el 2017 en los municipios de Yucatán. Información del conjunto de datos de la CONAFOR.

Los municipios más afectados (en cuestión de hectáreas) fueron **Tizimín** y **Hununcmá** con 1653.6 y 1643 ha, respectivamente (ver Figura 6.5). Para una visualización espacial de los municipios y el total área afectada de cada uno de ellos, ver la Figura B.1 en el Apéndice B

6.1.3. Tipo de vegetación y área afectada

Otro aspecto a explorar es el tipo de vegetación afectada por los incendios forestales (ver Figura 6.6).

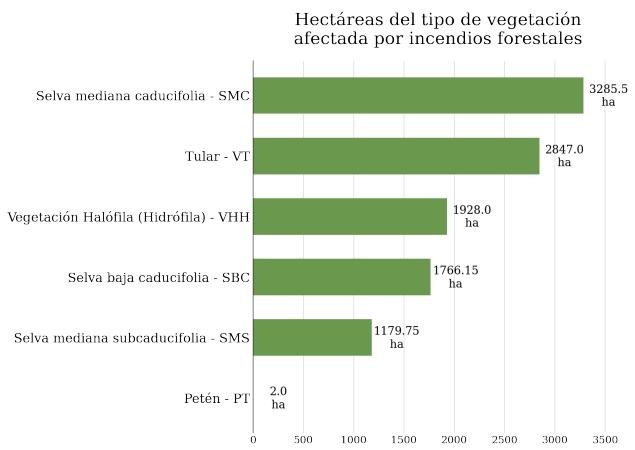


Figura 6.6: **Gráfica de barras.** Se muestra el registro de los tipos de vegetaciones afectadas por los incendios forestales y la extensión que estos últimos afectaron. Información del conjunto de datos de la CONAFOR.

Vegetación afectada por los incendios forestales en el estado de Yucatán durante el 2017

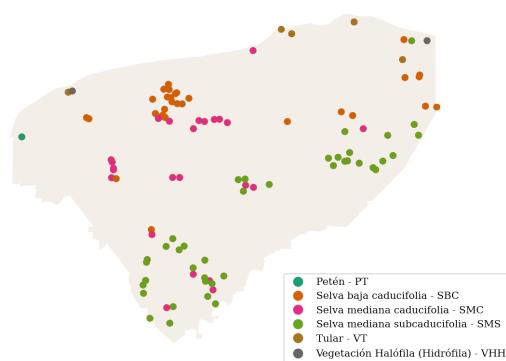


Figura 6.7: **Mapa de puntos.** Distribución del tipo de vegetación afectada por los incendios forestales en el estado. Información del conjunto de datos de la CONAFOR.

Se puede observar que entre la vegetación **Tular – VT** (comunidades de plantas acuáticas de 1 a 3 m de alto, de hojas angostas, características de climas templados subhúmedos ([Moreno-Casasola et al., 2010](#))) y **Selva mediana caducifolia – SMC** (vegetación alta y característica de Yucatán ([Zamora Crescencio, García Gil, Flores Guido, y Ortiz, 2008](#))) se cubre más del 50 % del área afectada en el estado.

6.1.4. Temporada de incendios

Al agrupar los datos por el mes de ocurrencia y la cantidad de superficie afectada, es claro el trimestre donde se registra un mayor número de incendios forestales: Marzo, Abril y Mayo, donde el mes de **Marzo** es el que cuenta con el mayor número de registros así como la mayor cantidad de hectáreas afectadas. Sin embargo, el mes de **Agosto** es el que menos registro de incendios forestales tuvo y supera al mes de **Abril** en superficie afectada (ver Figura 6.8).

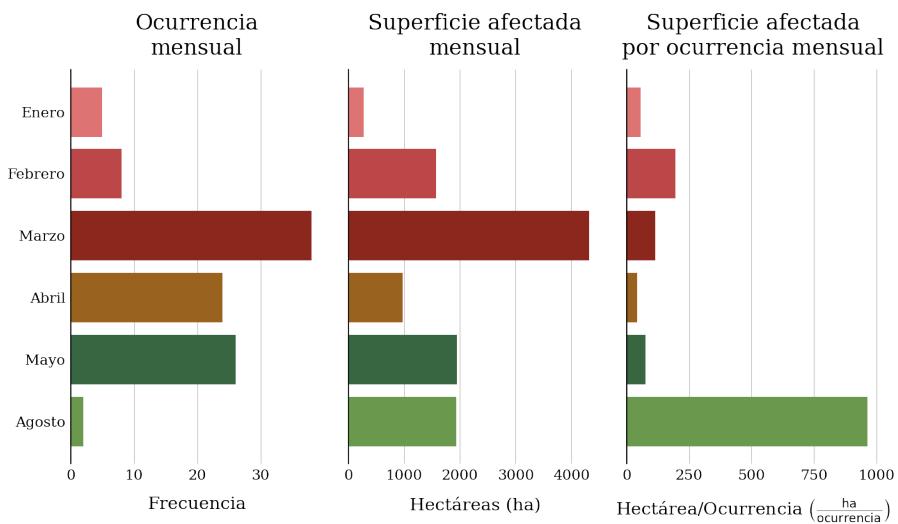
Al momento de encontrar los datos del mes de agosto se encontraron los siguientes registros (Ver Cuadro 6.4)

Duración (días)	Municipio	ANP	Tipo de impacto	Tipo de vegetación	Causa	Área afectada
Mayor a 7 Días	Hunucmá	No ANP	Impacto moderado	Vegetación Halófila (Hidrófila) – VHH	Cazadores	925 ha

Duración (días)	Municipio	ANP	Tipo de impacto	Tipo de vegetación	Causa	Área afectada
4 a 7 Días	Tizimín	Ría Lagartos	Impacto moderado	Vegetación Halófila (Hidrófila) – VHH	Cazadores	1003 ha

Cuadro 6.4: Atributos del área máxima afectada por incendios forestales del conjunto de datos de la CONAFOR

Incendios forestales a través del año 2017



Registros de la CONAFOR para el estado de Yucatán en el año 2017.

Figura 6.8: **Gráficas de barras.** Comparación de número de incendios y superficie afectada mensual. Información del conjunto de datos de la CONAFOR.

6.2. Datos de FIRMS

Este conjunto de datos es el que tiene la mayor cantidad de instancias, así como mayor rango de tiempo, ya que se cuentan con instancias desde el primer día del 2001 hasta el último día del 2020.

A diferencia del conjunto de datos de la CONAFOR, este conjunto no tiene una gran cantidad de atributos categóricos, la gran mayoría son atributos numéricos continuos, sin embargo se puede categorizar el atributo del tiempo, al agrupar información de manera anual y mensual.

6.2.1. Nivel de confianza en los puntos de calor y ubicación de puntos de calor.

De las características más relevantes dentro del conjunto de datos de FIRMS es que cada instancia tiene consigo misma un nivel de confianza que sirve para identificar si la observación es o no fuego.

Al tener un nivel de confianza, a cada instancia se le llama **punto de calor** y bajo este nombre se seguirá refiriendo a las observaciones. Los estimados del Cuadro 6.5 así como la Figura 6.9 revelan una distribución que podría asemejarse a distribución Gaussiana debido a una cercanía en los estimados de tendencia central y un coeficiente de asimetría que revela una ligera asimetría hacia la izquierda.

Estimado	Valor
Media	71.76 %
Mediana	73.00 %
Media truncada	72.56 %
Desviación estándar	17.46 %
Coeficiente de asimetría (γ_1)	-0.67 (asimetría hacia la izquierda)

Cuadro 6.5: Estimados de locación y variabilidad del atributo ‘confidence’ del conjunto de datos de FIRMS

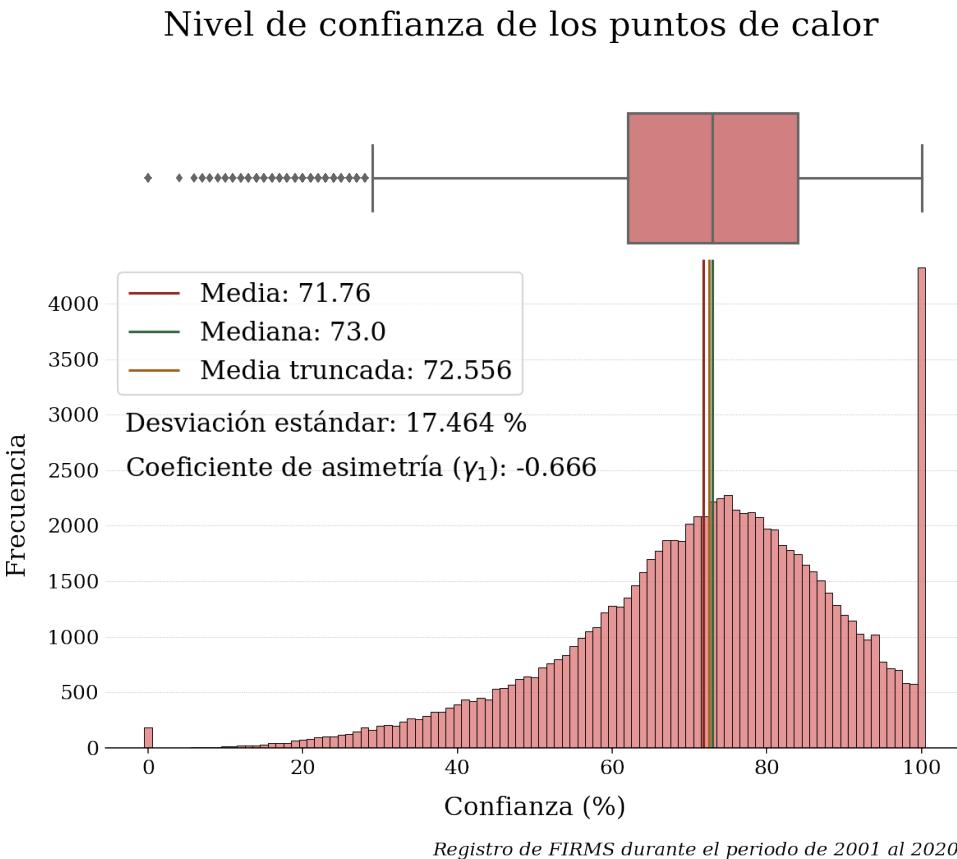


Figura 6.9: **Histograma y Box plot.** Distribución de niveles de confianza de los puntos de calor. Información del conjunto de datos de FIRMS.

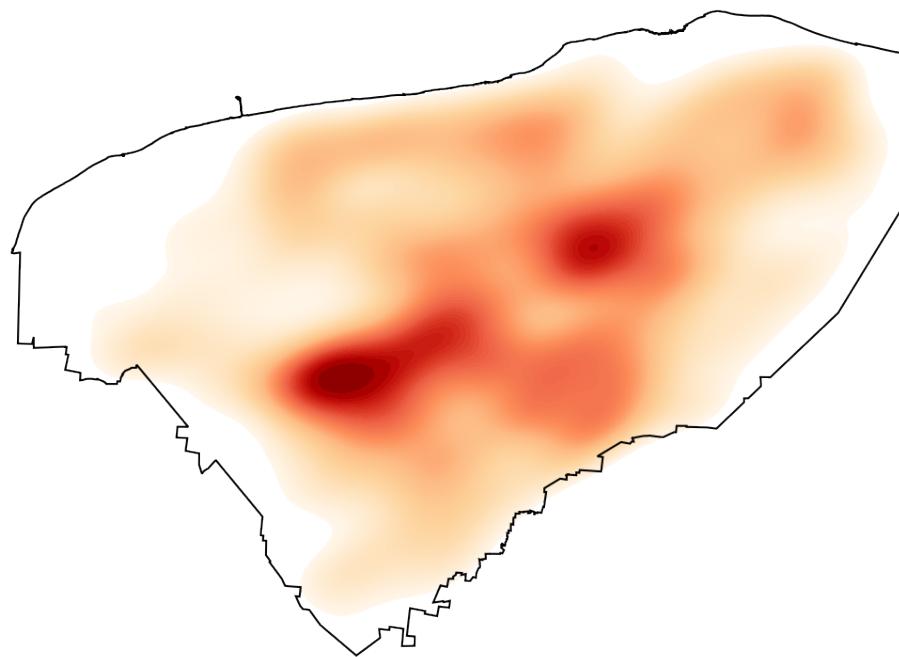
Como menciona el portal de FAQ de FIRMS en la pregunta *What is detection confidence?*¹, no hay una regla escrita para la selección de confianza con la que se desea trabajar, todo depende de la aplicación que se le esté dando a los datos.

Para el desarrollo de esta sección del análisis exploratorio de datos se usa una confianza mayor-igual (\geq) al tercer cuartil (**Q3**), es decir, trabajar con puntos de calor con un **nivel de confianza arriba del 84 %**.

Se puede observar en la Figura 6.10 que los puntos de calor se encuentran altamente concentrados en el centro y oriente del estado.

¹<https://earthdata.nasa.gov/faq/firms-faq#ed-confidence>

Concentración de puntos de calor en el estado de Yucatán durante el periodo 2001-2020



Puntos de calor con un nivel de confianza arriba del 84%. Registro de FIRMS durante el periodo de 2001 al 2020

Figura 6.10: Mapa de y gráfico de densidades en dos dimensiones. Información del conjunto de datos de FIRMS.

6.2.2. Registro de puntos de calor a través del tiempo

Se puede notar en la Figura 6.11 que hay una ventana de tiempo de aproximadamente 100 días (entre el día 50 y el día 150 del año) donde se cuenta con un alto número de registros. Esta ventana se puede ver en todos los años como una **variación estacional**, tal como se ve en la Figura 6.13.

El trimestre donde se registra un mayor número de registros es **Marzo, Abril y Mayo**, el mismo que muestra el conjunto de datos de la **CONAFOR**, con la diferencia en el mes con mayor número de registros. En el caso de la CONAFOR, el mes de Marzo fue el máximo y para el caso de los datos de FIRMS el mes máximo es Abril (ver Figura 6.12).

Registro de puntos de calor a través del tiempo

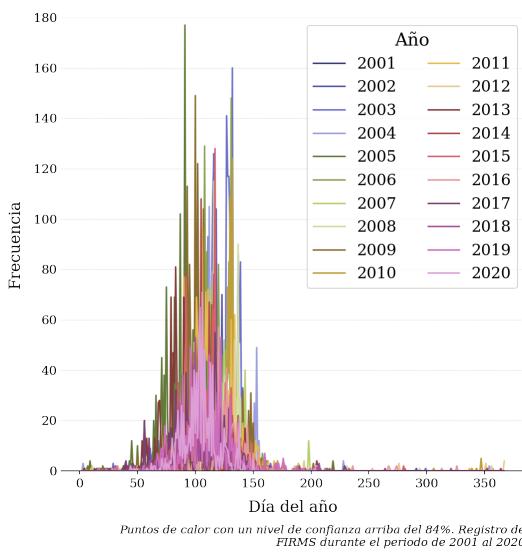


Figura 6.11: **Serie de tiempo.** Distribución del registro del número de puntos de calor con un nivel de confianza arriba del 84 % a través de año en el estado de Yucatán. Información del conjunto de datos de FIRMS.

Registro mensual de puntos de calor

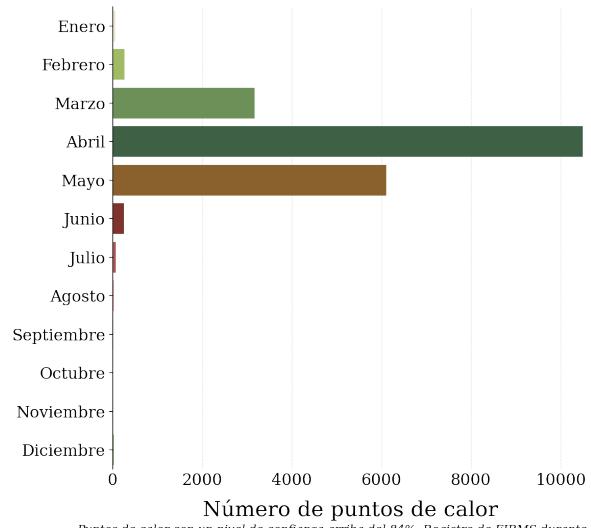


Figura 6.12: **Gráfica de barras.** Número de puntos de calor registrados con un nivel de confianza por arriba del 84 % por mes durante el periodo del 2001 al 2020, en el estado de Yucatán. Información del conjunto de datos de FIRMS.

Registro de puntos de calor a través de los años

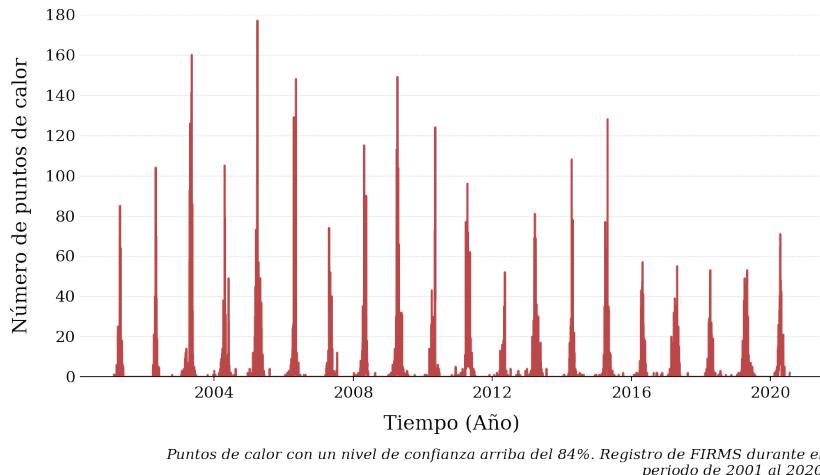


Figura 6.13: **Serie de tiempo.** Distribución del registro del número de puntos de calor con un nivel de confianza arriba del 84 % a través de los años del 2001 al 2020, en el estado de Yucatán. Información del conjunto de datos de FIRMS.

La correlación entre los años y el número de registros es negativa lo que se podría inferir que al paso de los años el número de puntos de calor va disminuyendo, sin embargo el valor de esta correlación es débil por lo que hace falta más información para dar un veredicto (ver Figuras 6.14 y 6.15).

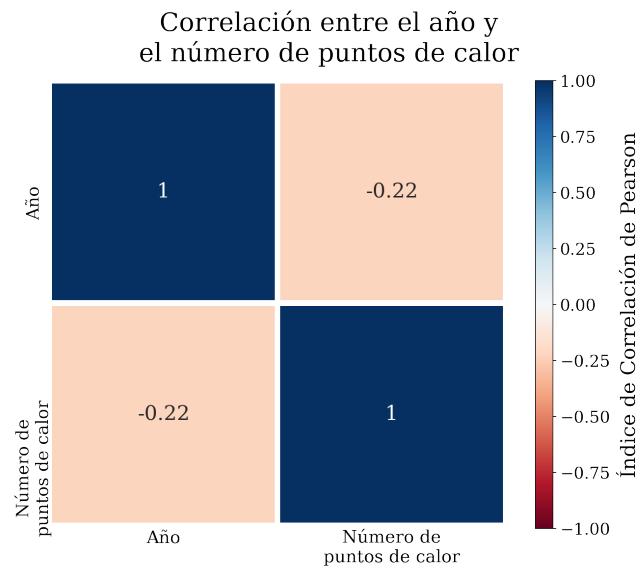


Figura 6.14: **Matriz de correlación.** Correlación entre el número de puntos de calor registrados por año y el año, en el estado de Yucatán. Información del conjunto de datos de FIRMS.

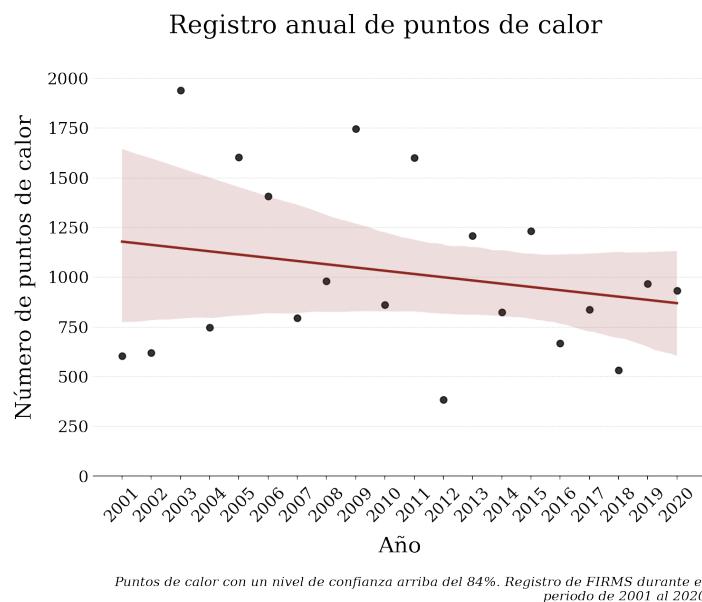
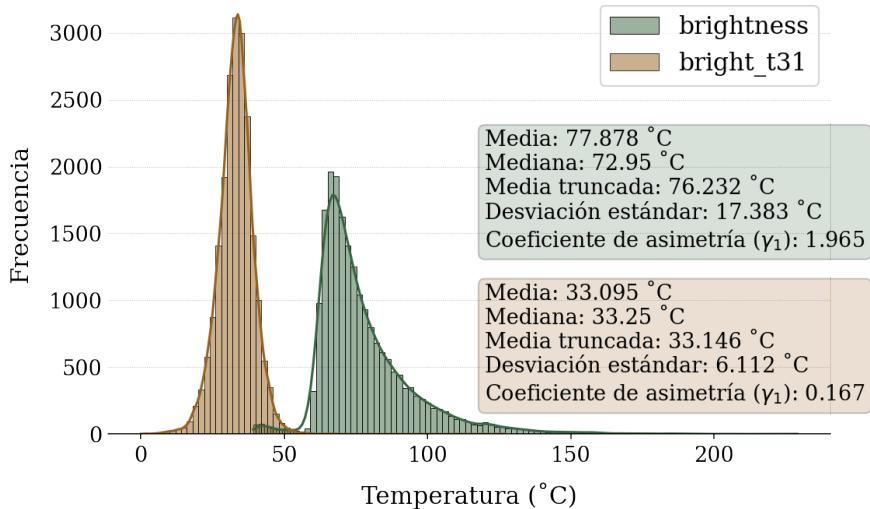


Figura 6.15: **Gráfico de dispersión y ajuste lineal.** Número de registros de puntos de calor a través de los años del 2001 al 2020. Información del conjunto de datos de FIRMS.

6.2.3. Temperaturas

Existen dos atributos similares los cuales son ‘brightness’ y ‘bright_t31’, que significan la temperatura de la observación. Ambos atributos están en escala Kelvin por lo que se hizo una conversión a grados Celsius ($^{\circ}\text{C}$). En la Figura 6.16 se observan ambas distribuciones así como sus respectivos estimados de locación y variabilidad.

Distribución de los atributos de temperatura



Puntos de calor con un nivel de confianza arriba del 84%. Registro de FIRMS durante el periodo de 2001 al 2020

Figura 6.16: **Histogramas.** Comparación entre los atributos `brightness` y `bright_t31`. Información del conjunto de datos de FIRMS.

Se observa que hay una gran diferencia entre las distribuciones de la temperatura. Para comprobarlo se realiza una Prueba de Hipótesis bajo las siguientes hipótesis:

- Hipótesis nula: $H_0 : A - B \leq 0$
- Hipótesis alternativa: $H_1 : A - B > 0$

Donde **A** es la **media muestral de ‘brightness’** y **B** es la **media muestral de ‘bright_t31’**.

Al realizar la prueba con la Prueba de Permutación se pudo observar que la diferencia de las medias muestrales de los atributos (44.783 °C) esta muy alejada de la distribución de la Figura 6.17, dando como resultado un valor-p ≈ 0 , siendo esto estadísticamente significativo, por lo que existe suficiente evidencia para rechazar la hipótesis nula H_0 . El atributo ‘brightness’ es en promedio, mayor que ‘bright_t31’.

Distribución de diferencia de temperaturas medias

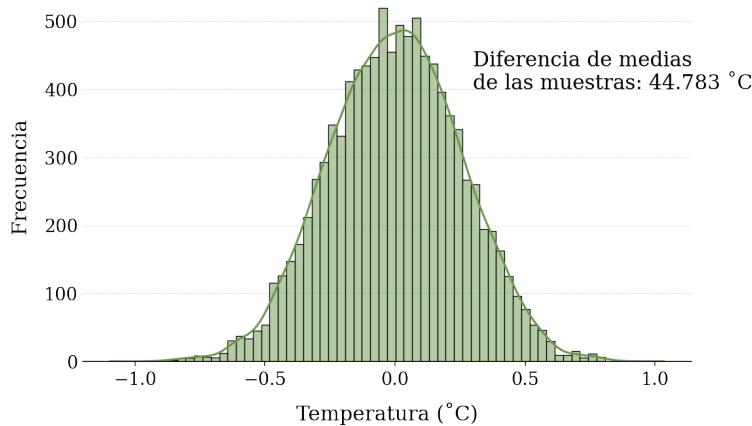


Figura 6.17: **Histograma.** Distribución de la diferencia de las medias muestrales tras la realización de una prueba de hipótesis con la Prueba de Permutación.

Tiene sentido físico que las temperaturas del atributo ‘brightness’ sean las de los puntos de calor ya que ‘bright_t31’ tiene registros de temperaturas cercanas a los 0°C .

La correlación entre los años y la temperatura media de los puntos de calor es negativa, lo que se podría inferir que al paso de los años la temperatura media de los puntos de calor va disminuyendo, sin embargo el valor de esta correlación es débil por lo que hace falta más información para dar un veredicto (ver Figuras 6.18 y 6.19).

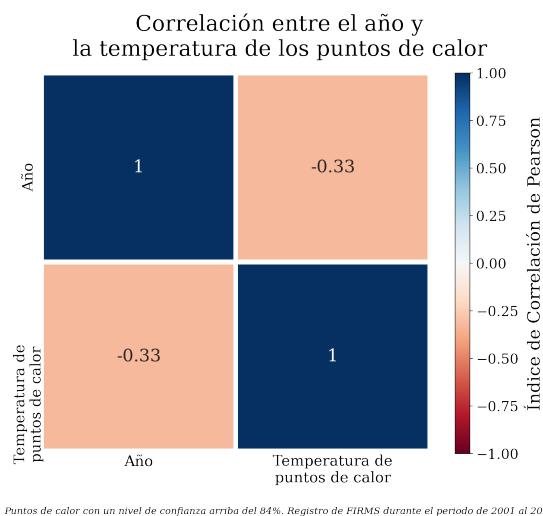


Figura 6.18: **Matriz de correlación.** Correlación entre la temperatura media anual de los puntos de calor y el año. Información del conjunto de datos de FIRMS.

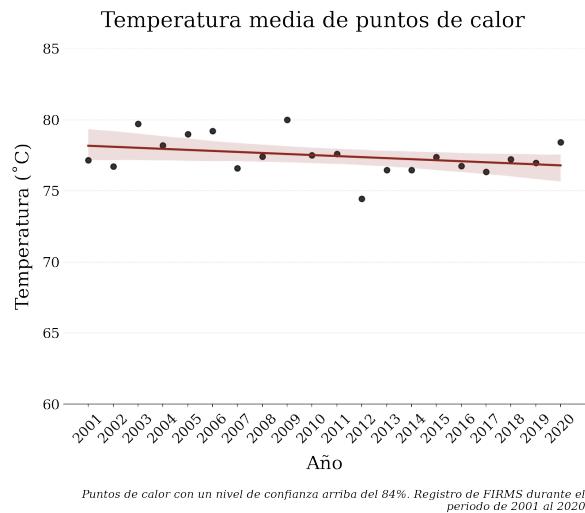


Figura 6.19: **Gráfico de dispersión y ajuste lineal.** Temperatura media anual de los puntos de calor a través de los años 2001 al 2020. Información del conjunto de datos de FIRMS.

6.2.4. Potencia Radiactiva del Fuego (FRP)

La Potencia Radiativa del Fuego (FRP, por sus siglas en inglés), es una métrica usada para el monitoreo remoto para cuantificar la biomasa quemada. FPR mide la energía radiativa liberada por unidad de tiempo por la vegetación quemada. Sus unidades están en MegaWatts (MW) ([Costa y Fonseca, 2017](#)).

Se puede explorar esta variable comparando su distribución con el atributo ‘`brightness`’ para poder encontrar la concentración de ambos dentro de todo el conjunto de datos. Donde se observa que existe una alta concentración de puntos de calor que emitieron menos de 250 MW y tenían una temperatura cercana a los 75°C (ver Figura 6.20).

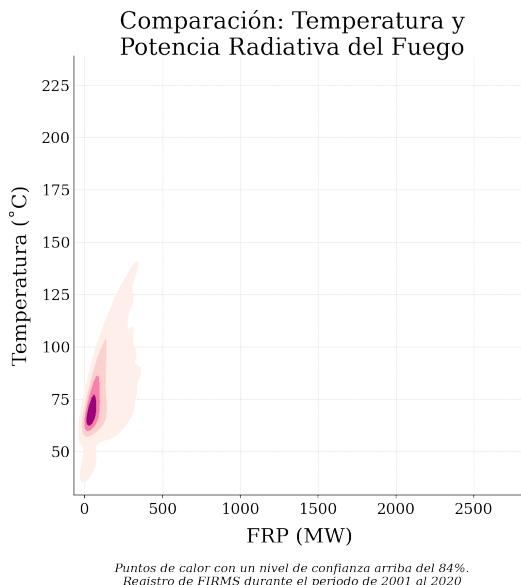


Figura 6.20: **Gráfico de densidades en 2 dimensiones.** Concentración y comparación de los datos en los atributos de FRP y Temperatura del punto de calor. Información del conjunto de datos de FIRMS.

6.2.5. Comparación con datos de la CONAFOR

Comparando de nuevo con el conjunto de datos de la CONAFOR, se encuentra el siguiente mapa (ver Figura 6.21) donde se muestran los registros de la CONAFOR durante el año 2017 en el trimestre más alto de registros (Marzo, Abril y Mayo) y las instancias del conjunto de FIRMS con la temperatura de los puntos de calor.

Se observa que hay registros de FIRMS que no coinciden con los registros de la CONAFOR, esto puede ser que estas observaciones hayan sido incendios controlados. Sin embargo, hay registros de la CONAFOR que no coinciden con los registros de FIRMS, esto puede suponer ha que hayan sido incendios muy pequeños y difíciles de distinguir, sea por la misma vegetación o condiciones atmosféricas.

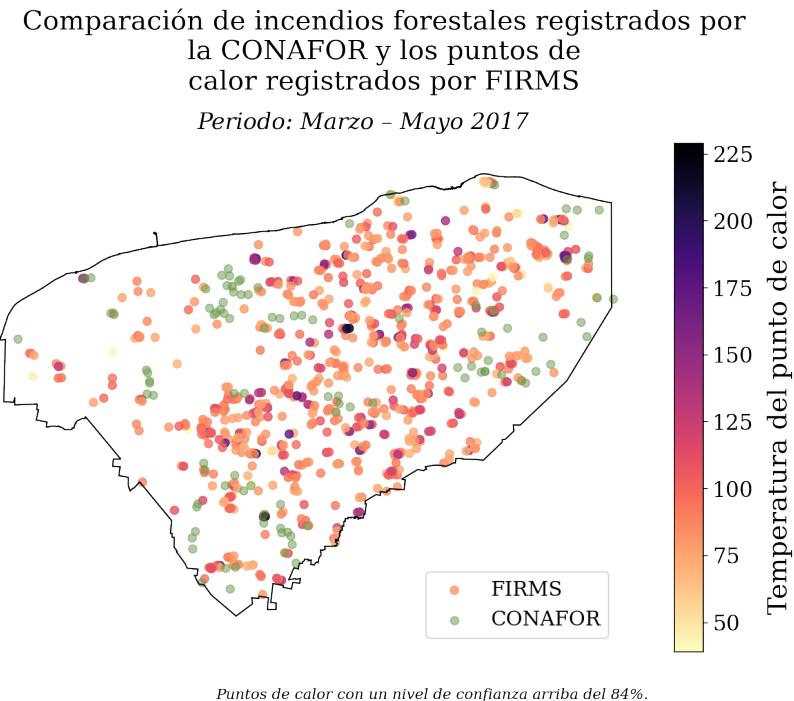


Figura 6.21: **Mapa de puntos.** Comparación incendios forestales registrados por la CONAFOR en el año 2017 y los puntos de calor registrados por FIRMS durante el periodo Marzo-Mayo del 2017.

6.3. Datos de Google Earth Engine

Los conjuntos de datos hasta ahora expuestos estuvieron enfocados en información sobre los incendios y sus alrededores. Para el caso de los datos de la CONAFOR se tenía acceso a la ubicación geográfica, municipios y estados donde ocurrieron los incendios, así como el tiempo que duró y el tipo de vegetación específica que fue afectada. Los puntos de calor del conjunto de datos de FIRMS son el resultado de observaciones de satélites; en este conjunto de datos se obtiene información como la temperatura de ese punto de calor, la ubicación geográfica y el nivel de confianza de que esa observación era un incendio forestal. Sin embargo, con esos conjuntos de datos no se puede caracterizar el área de ocurrencia o la región de estudio.

Los conjuntos de datos proporcionados por la plataforma Google Earth Engine (GEE) son de diferentes variables ambientales y de vegetación. Son los datos con el mayor registro histórico de sus respectivas variables así como su completa cobertura sobre toda la región de estudio.

6.3.1. Registro de incendios forestales por medio de MODIS (Moderate Resolution Imaging Spectroradiometer)

El conjunto de datos satelitales con la información mensual de los incendios forestales se encuentran en MODIS Fire_cci Burned Area pixel product version 5.1 (FireCCI51) ([Lizundia-Loiola et al., 2020](#)) a diferencia del conjunto de datos de FIRMS, la información se limita a

registros desde el año 2001 hasta el año 2019.

Un atributo que comparte el conjunto de GEE y FIRMS es la el nivel de confianza de un determinado incendio, por lo que se aplicará el mismo filtro únicamente para la identificación de valores arriba del 84 % de nivel de confianza.

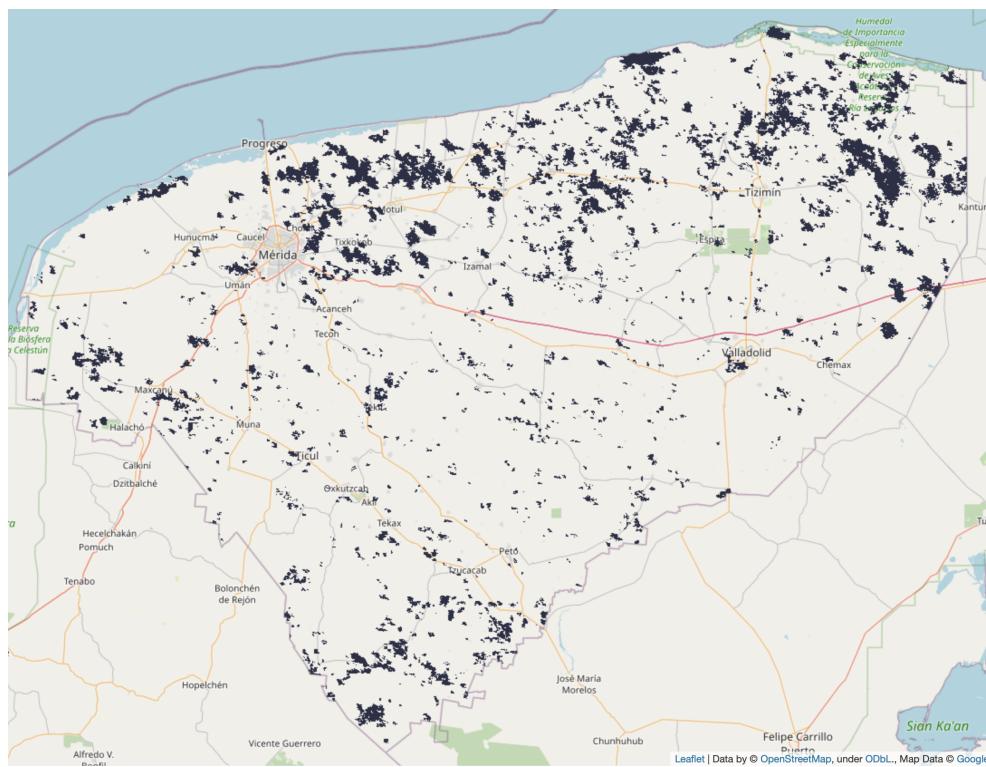


Figura 6.22: **Mapa.** Superficie afectada por incendios forestales en el estado de Yucatán durante el periodo de tiempo de 2001 a 2019 con un nivel de confianza mayor a 84 %. Fuente: ([Lizundia-Loiola et al., 2020](#)) vía Google Earth Engine

En la Figura 6.22 se puede ver de manera más clara las regiones que fueron afectadas durante el periodo de tiempo entre el inicio del 2001 hasta el fin del 2019. Se pueden ver afectaciones en todo el estado, principalmente en toda la zona norte central y oriente del estado; así como pequeñas concentraciones en la zona sur del estado.

6.3.2. Variables ambientales

En anteriores conjuntos de datos se pudo obtener la temperatura de los puntos de calor pero no se tiene acceso a las condiciones ambientales de la superficie terrestre tales como humedad del suelo, temperatura o nivel de sequía. Para este caso se usan los siguientes conjuntos de datos:

- **Temperatura:** MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1km ([Wan et al., 2015](#))
- **Humedad y sequía:** TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho ([Abatzoglou et al., 2018](#))

Variable	Media	Desviación estandar	Valor mínimo	Valor máximo
Temperatura de la superficie	29.384 °C	1.091 °C	25.25 °C	35.688 °C
Humedad del suelo	60.646 mm	51.162 mm	0.7 mm	280.85 mm
Índice de Sequía de Palmer (PDSI)	1.155	0.169	-2.404	1.576

Cuadro 6.6: Estimados de ubicación de las imágenes satelitales de las variables ambientales: Temperatura de la superficie, Humedad del suelo y PDSI durante el periodo de tiempo del 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.

Con el apoyo de la Figura 6.23 se puede observar que las zonas con mayor temperatura a la media (colores más cálidos) son también zonas donde se encuentra una mayor cantidad de incendios. Las temperaturas más altas se registran en el norte del estado, así como cerca de grandes ciudades como Mérida.

La humedad del suelo está en milímetros, entre mayor la cantidad de milímetros mayor es la cantidad de agua en la superficie (humedad). Observando el mapa se puede notar una región en el oriente del estado con altos niveles de humedad en su suelo, donde no se puede ver muchos registros de incendios forestales, sin embargo, para el resto del estado, se cuenta con una humedad baja y también es donde se encuentra una mayor concentración de registros de incendios forestales (ver Figura 6.24).

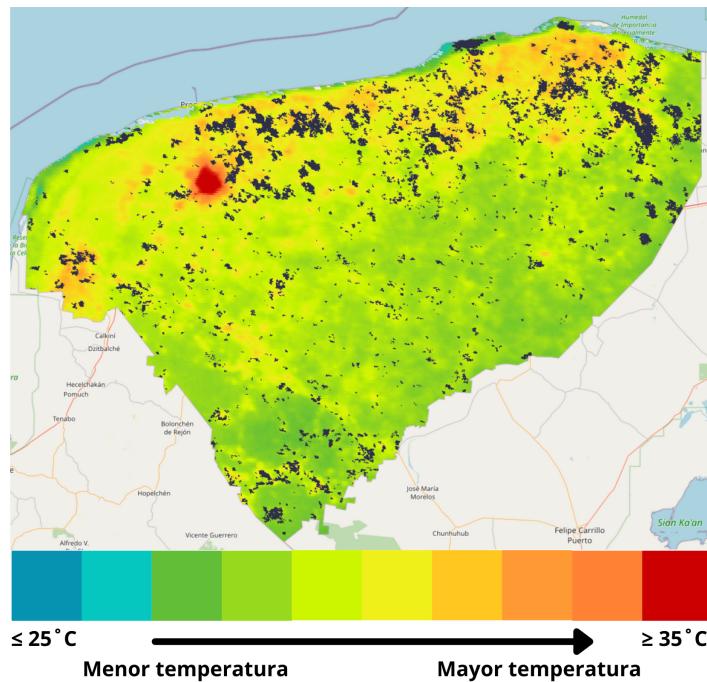


Figura 6.23: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84% y temperatura mediana de la superficie en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Wan et al., 2015](#)) vía Google Earth Engine

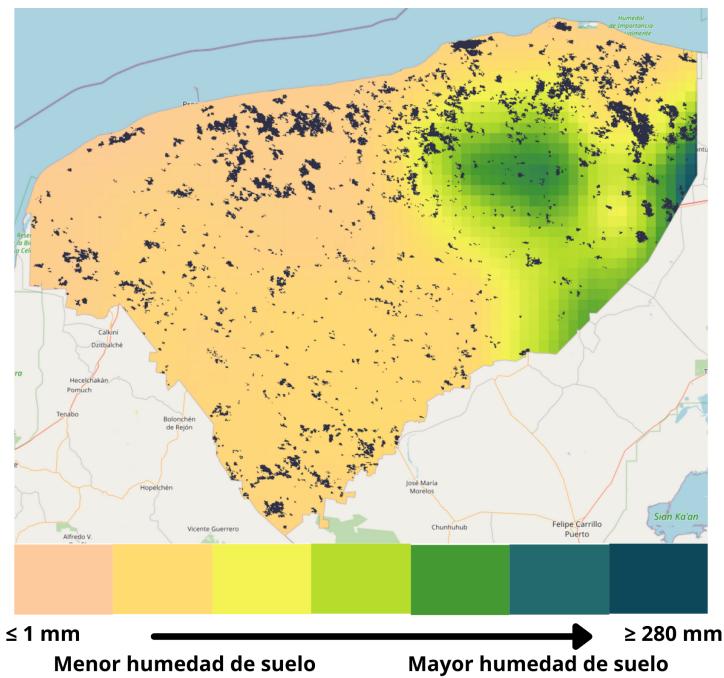


Figura 6.24: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84% y humedad del suelo mediana en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Abatzoglou et al., 2018](#)) vía Google Earth Engine.

Finalmente se cuenta con el **Índice de Sequía de Palmer (PDSI)**, en la práctica, este índice ronda entre los valores -4 y 4, donde menor o igual a -4 significa una región

extremadamente seca, con bajo número de lluvias; mientras que el caso contrario es una región con rachas de lluvias altas (Wells, Goddard, y Hayes, 2004). Se pueden ver que en promedio, una gran extensión del estado de Yucatán cuenta con extensiones territoriales con altas rachas de lluvia (color azul); no se puede encontrar a simple vista regiones específicas donde los incendios se han concentrado (ver Figura 6.25).

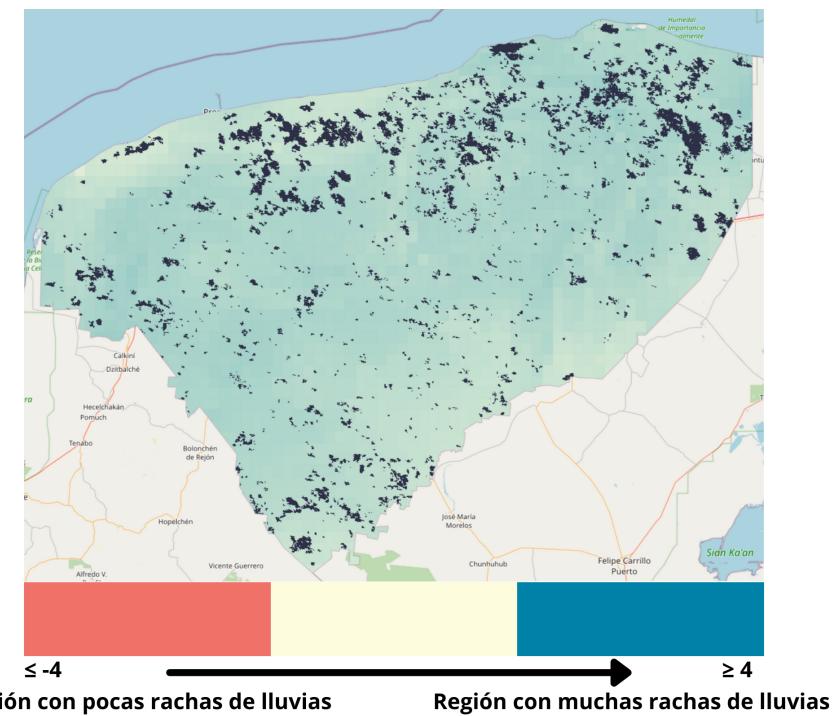
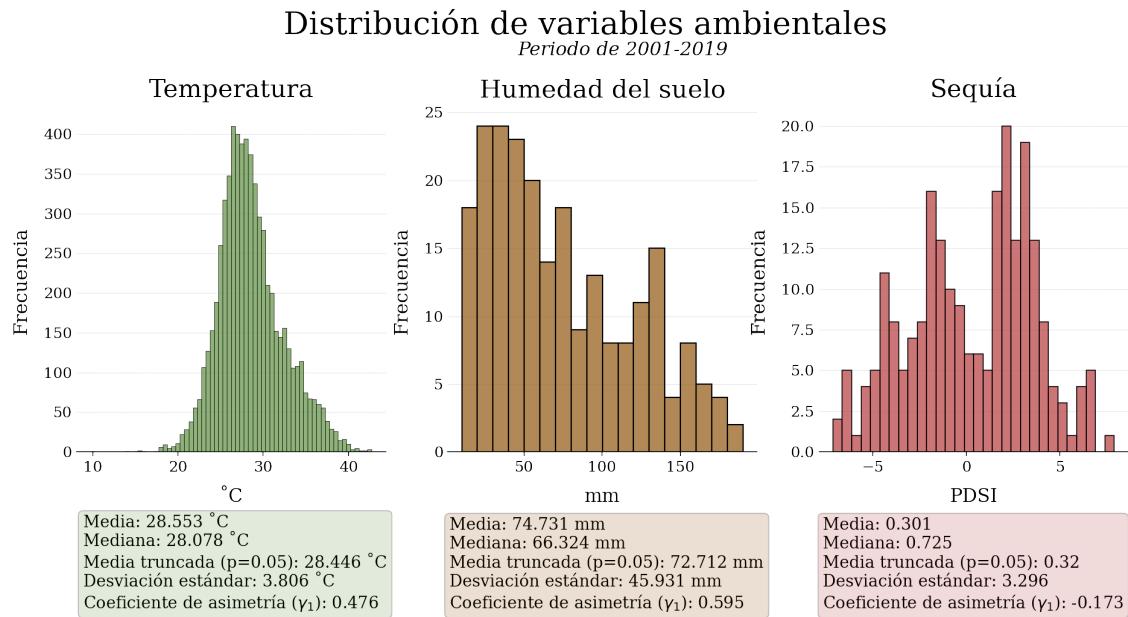


Figura 6.25: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI en el estado de Yucatán en el periodo de tiempo de 2001 a 2019. Fuente: (Lizundia-Loiola et al., 2020) y (Abatzoglou et al., 2018) vía Google Earth Engine.

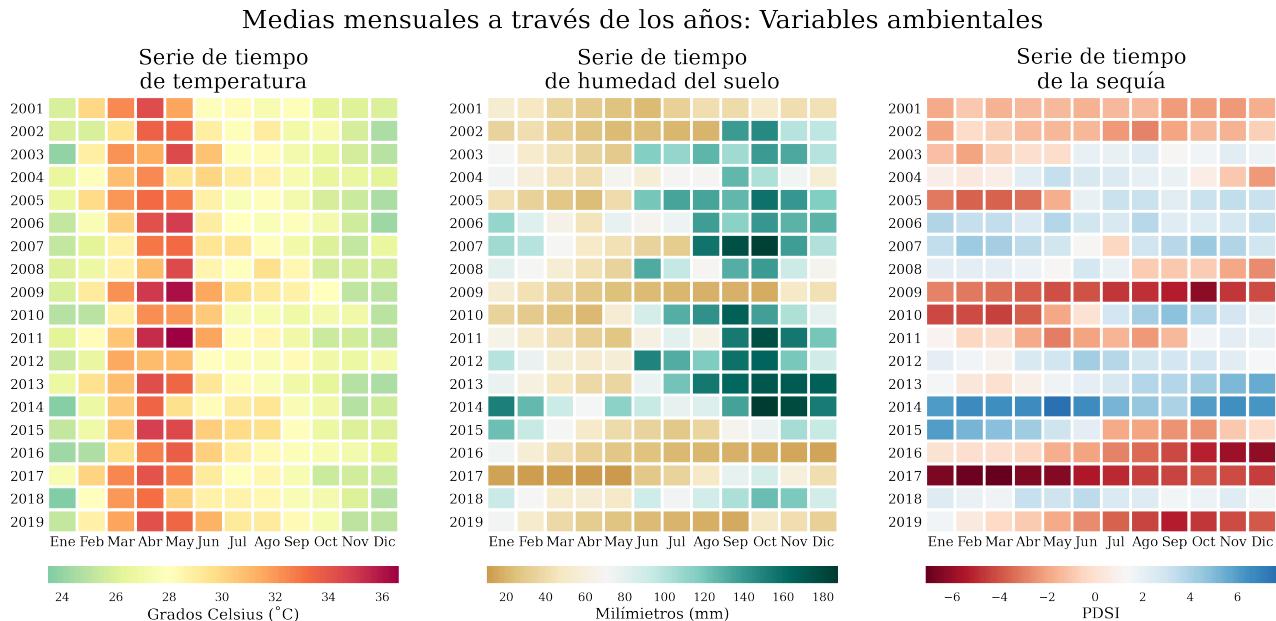
Cambios a través del tiempo

Los mapas muestran un resumen del periodo del tiempo, por lo que no están claras las variaciones estacionales de las variables ambientales, por lo que es importante identificarlas. En la Figura 6.27 se puede ver los años entre 2016 y 2018 fueron de baja humedad de suelo (con respecto a la mediana histórica) y secos (sin altas rachas de lluvias).



Fuente: MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1km y TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho via Google Earth Engine

Figura 6.26: **Histograma.** Distribuciones y estimados de locación y variabilidad de las variables ambientales: Temperatura de la superficie, Humedad del suelo y PSDI. Información obtenida por la plataforma Google Earth Engine



Las correlaciones que existen entre las tres variables ambientales se pueden observar en la Figura 6.30, donde el par Humedad del suelo y PDSI estan fuertemente correlacionados.

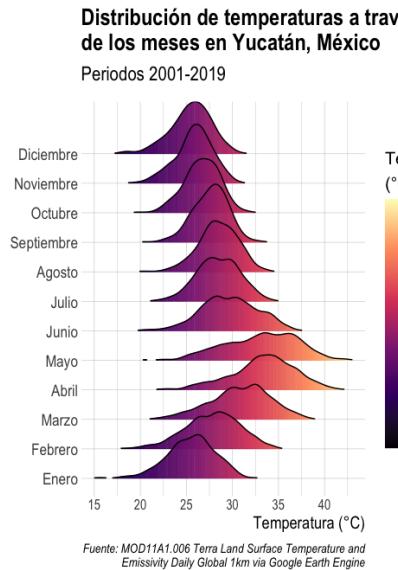


Figura 6.28: **Ridgeline plot.** Distribución de temperatura de la superficie a través de los meses en los años 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.

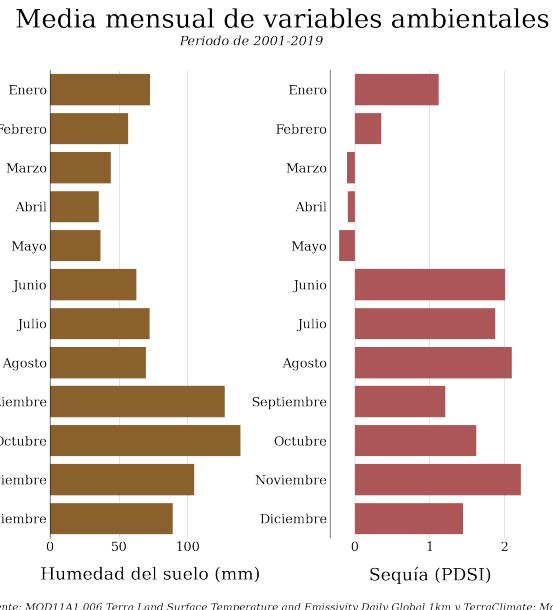


Figura 6.29: **Gráfico de barras.** Media mensual de las variables ambientales, Humedad del suelo y PDSI, en los años 2001 a 2019. Información obtenida por la plataforma Google Earth Engine.

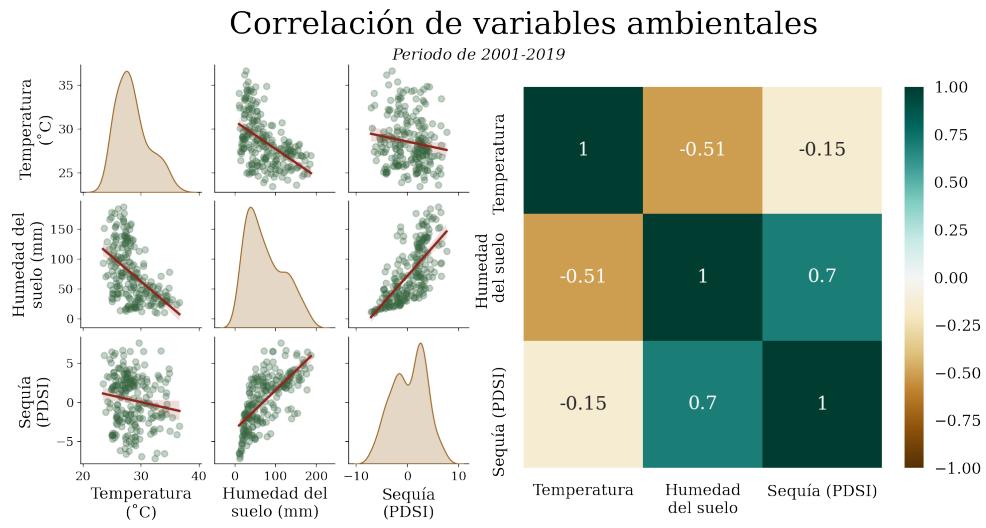
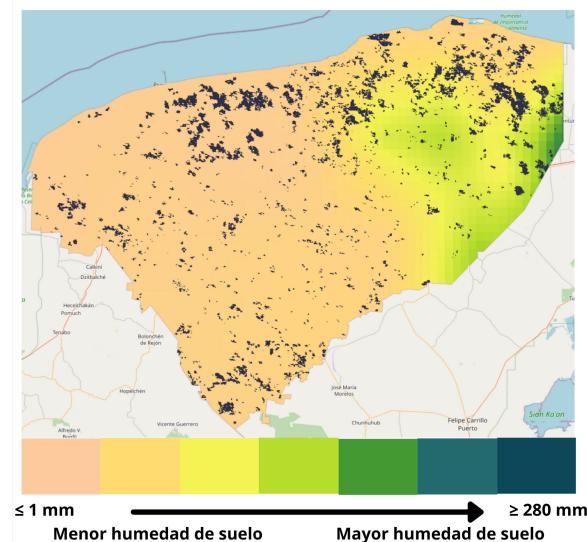
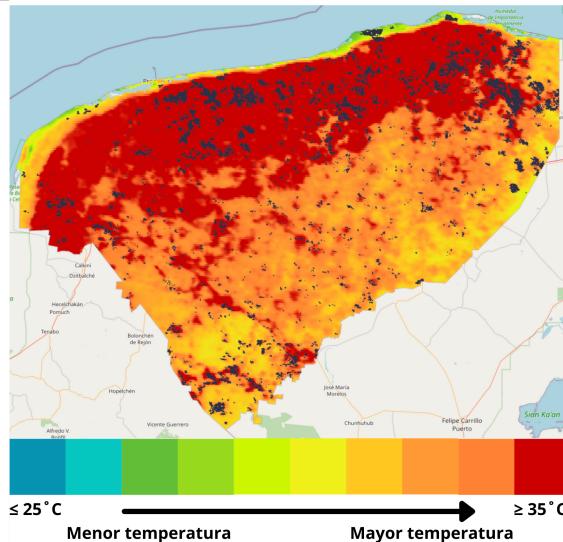


Figura 6.30: **Mapa de calor de matriz de correlaciones.** Matriz de correlaciones entre las variables ambientales: Temperatura de la superficie, Humedad del suelo y PSDI. Información obtenida por la plataforma Google Earth Engine.

Con la información mostrada se puede hacer un enfoque en las condiciones climáticas y las ocurrencias de fuego durante el trimestre de Marzo a Mayo en el periodo del 2001 al 2019.

Se conservan los parámetros de visualización de las métricas anuales para que se puedan observar de manera más clara el cambio de la temperatura durante ese trimestre del año. Es importante recalcar que el mapa que se esta observando es la (temperatura, humedad y PDSI) media de la mediana del trimestre antes mencionado, a través de todo el periodo del 2001 al 2019.

Para el caso de la temperatura se siguen viendo altas temperaturas en el norte del estado. Se puede notar que las áreas afectadas se encuentran dentro de las regiones de más altas temperaturas (ver Figura 6.31). El mismo comportamiento se observa cuando se mira a la humedad del suelo del estado, donde se puede ver de nuevo que el oriente del estado cuenta con altos niveles de humedad y un bajo registro de áreas afectadas (ver Figura 6.32). Los valores positivos del PDSI ya no se ven tanto en este trimestre a comparación de lo que se promedia a lo largo del año. La distribución de áreas afectadas por incendios forestales no parece ser altamente afectada por las zonas más secas (ver Figura 6.33).



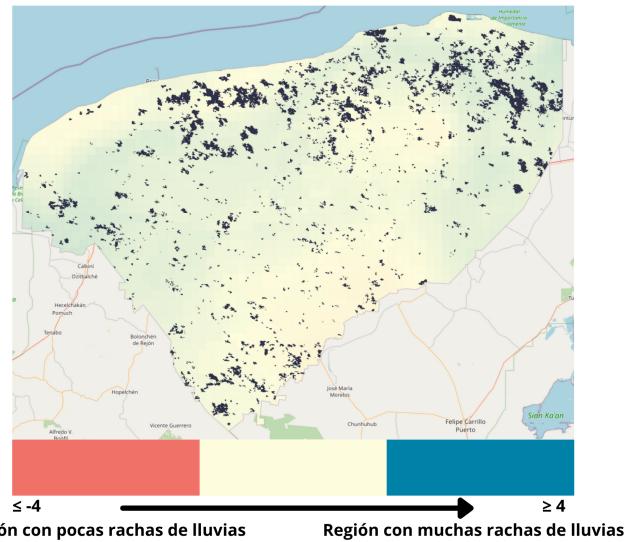


Figura 6.33: Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI en el estado de Yucatán en los meses Marzo, Abril y Mayo en los años 2001 a 2019. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Abatzoglou et al., 2018](#)) vía Google Earth Engine.

6.3.3. Índices de vegetación

Las variables ambientales no son las únicas que pueden caracterizar las zonas, también la vegetación y su cambio a través del tiempo.

Tanto el EVI así como el NDVI, indican los cambios en la vegetación. Valores positivos de ambos significa una vegetación más sana y verde, por lo contrario cuando la vegetación es mala, seca o menos verde, el valor de estos índices es negativo. El rango de ambos es [-1,1]; EVI es una versión mejorada de NDVI, no es sensible a anomalías debido a condiciones atmosféricas y es excelente para el monitoreo de vegetación tropical. Los cambios a través del tiempo en estos índices no son únicamente por incendios forestales, también son afectados por la agricultura y la deforestación.

De la Figura 6.34 hasta la Figura 6.39, se muestra el historial de cambios de los índices de vegetación a través del tiempo de NDVI y EVI, respectivamente. Ambos mapas son resultados de un periodo de tiempo base que es de Enero del 2001 hasta Diciembre del 2003. Entre mayor sea el descenso de la vegetación, mayor es la tonalidad roja en la región; y mientras más crezca la vegetación, tonalidades más verdes toma. Cada región varía en color dependiendo de los índices de vegetación.

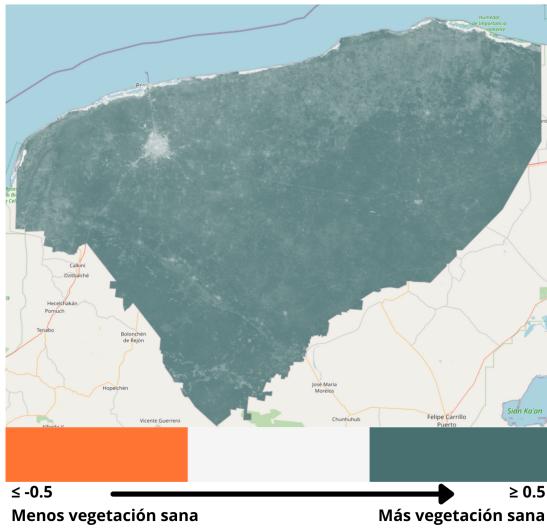


Figura 6.34: **Mapa.** Mediana de NDVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2001 a Diciembre 2003 (periodo de tiempo base). Fuentes ([Didan, 2015](#)) vía Google Earth Engine

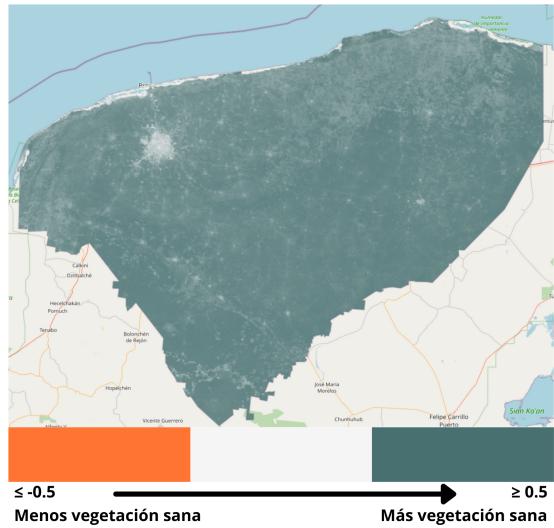


Figura 6.35: **Mapa.** Mediana de NDVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2004 a Diciembre 2019 (periodo actual). Fuentes ([Didan, 2015](#)) vía Google Earth Engine

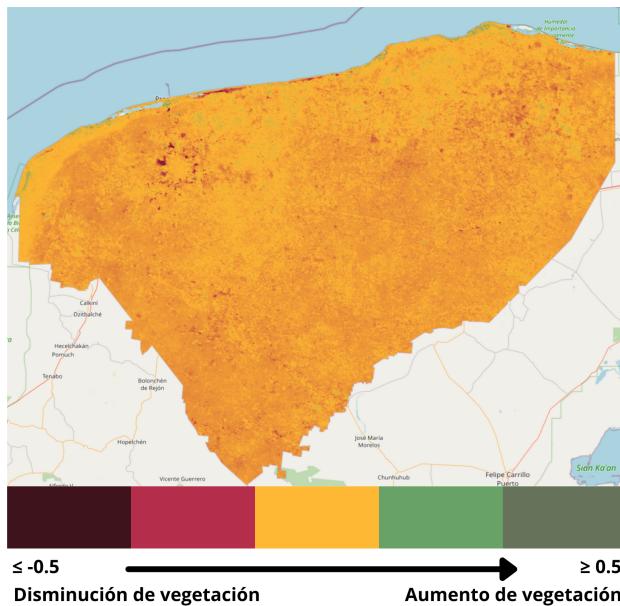


Figura 6.36: **Mapa.** Cambios en el NDVI entre el periodo base y el periodo actual. Fuentes ([Didan, 2015](#))

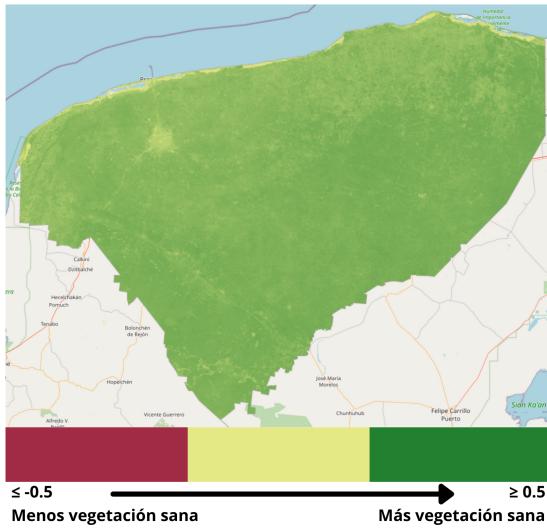


Figura 6.37: **Mapa.** Mediana de EVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2001 a Diciembre 2003 (periodo de tiempo base). Fuentes (Didan, 2015) vía Google Earth Engine

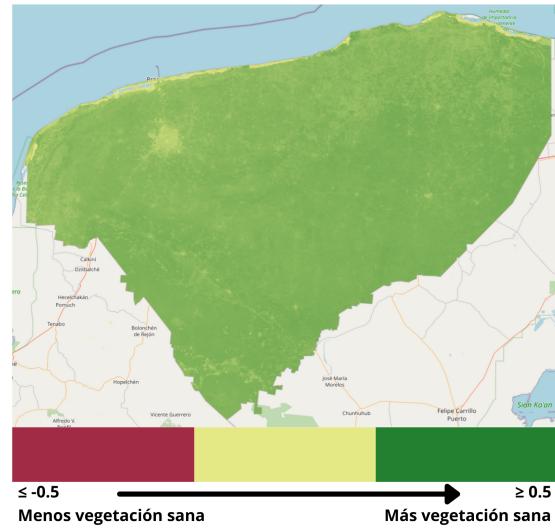


Figura 6.38: **Mapa.** Mediana de EVI en el estado de Yucatán durante en el periodo de tiempo de Enero 2004 a Diciembre 2019 (periodo actual). Fuentes (Didan, 2015) vía Google Earth Engine

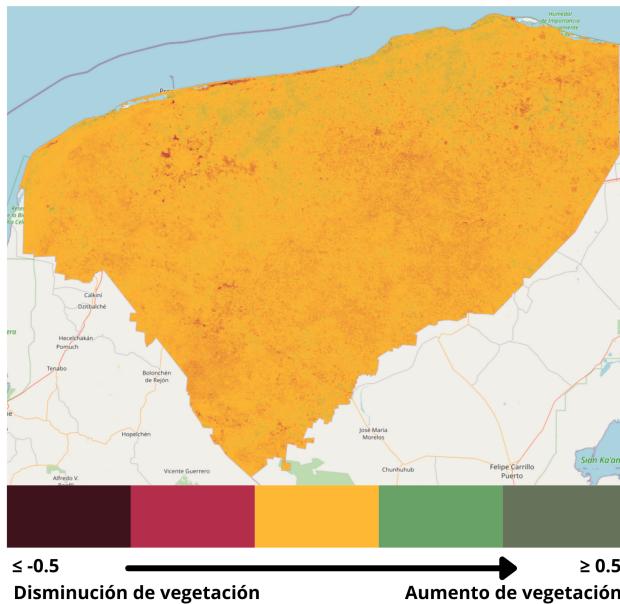


Figura 6.39: **Mapa.** Cambios en el EVI entre el período base y el período actual. Fuentes (Didan, 2015)

6.3.4. Comparación: Mes de Abril del año 2017

Con las variables obtenidas a través de GGE, se puede hacer un acercamiento a los incendios forestales que fueron registrados por la CONAFOR. En este caso se presta atención al mes de Abril del 2017, siendo el mes intermedio del periodo de altas temperaturas y registros de incendios forestales.

Se muestran mapas con las comparaciones de la temperatura mediana del mes (ver Figura 6.40), humedad del suelo mediana del mes (ver Figura 6.41), PDSI mediana del mes (ve Figura 6.42) y los cambios en el EVI (ver Figura 6.43) con respecto al mes anterior (marzo). Cada uno de ellos muestra tres tipos de puntos o pixeles:**Puntos cafés:** Registros de FIRMS; **Puntos verdes:** Registros de la CONAFOR; y **Pixeles azules:** Registros de FireCCI51.

Las variables ambientales como la temperatura, PDSI y humedad del suelo se muestran en sus extremos. Altas temperaturas, baja humedad y alto PDSI.

Gran extensión del territorio estatal muestra un nulo crecimiento del EVI. La zonas donde hay un crecimiento se encuentran en la frontera oriente y sur del estado, y las pocas regiones donde hay un descenso del EVI se encuentran en el norte. Se puede observar que en las zonas donde se registra un descenso del EVI se encuentran cercanas a los registros de los incendios forestales (ver Figura 6.43).

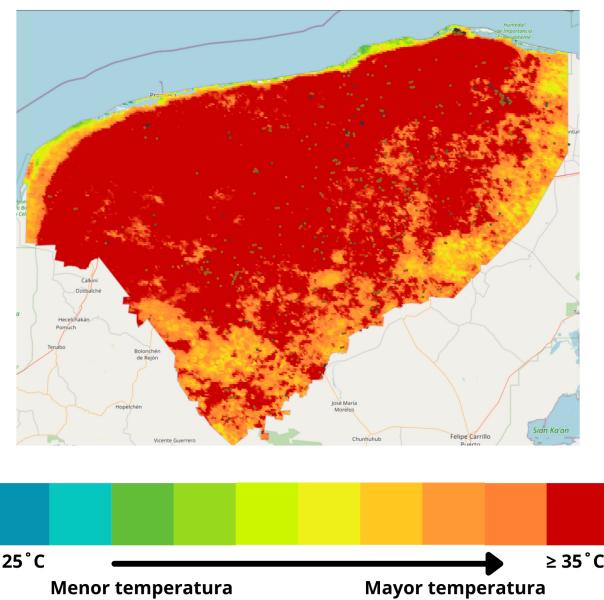


Figura 6.40: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y temperatura mediana de la superficie en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Wan et al., 2015](#)) vía Google Earth Engine

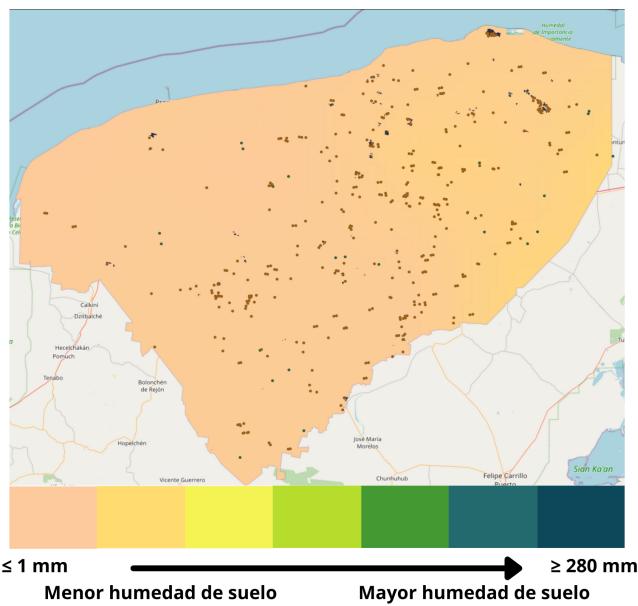


Figura 6.41: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y humedad del suelo mediana en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Abatzoglou et al., 2018](#)) vía Google Earth Engine

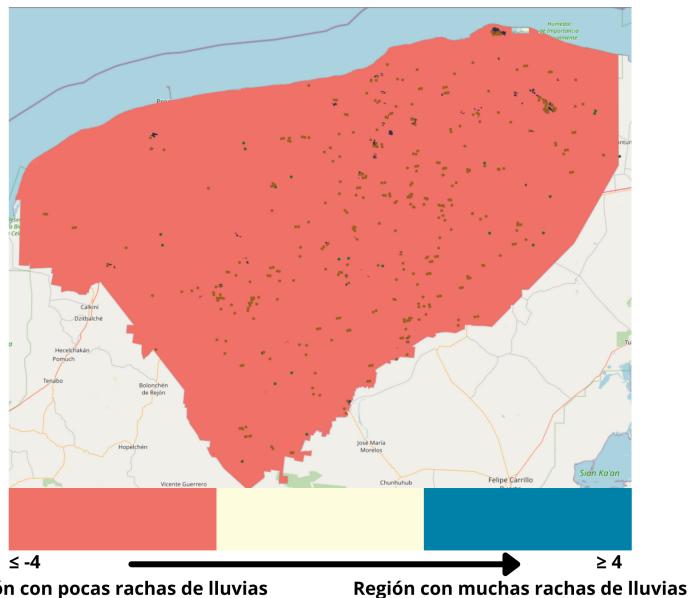


Figura 6.42: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y PDSI mediana en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Abatzoglou et al., 2018](#)) vía Google Earth Engine

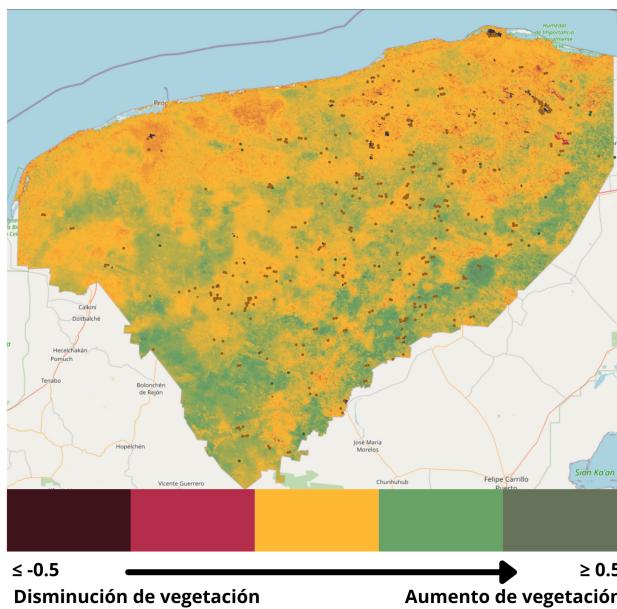


Figura 6.43: **Mapa.** Superficie afectada por incendios forestales con un nivel de confianza mayor a 84 % y mediana de cambios en el EVI (con respecto al mes anterior) en el estado de Yucatán en el mes de Abril del año 2017. Fuentes: ([Lizundia-Loiola et al., 2020](#)) y ([Didan, 2015](#)) vía Google Earth Engine

Algunos grupos de píxeles llegan a coincidir con registros de otros conjuntos de datos como los del conjunto de FIRMS, por ejemplo dentro del ANP de Ría Lagartos (ver Figura 6.44).

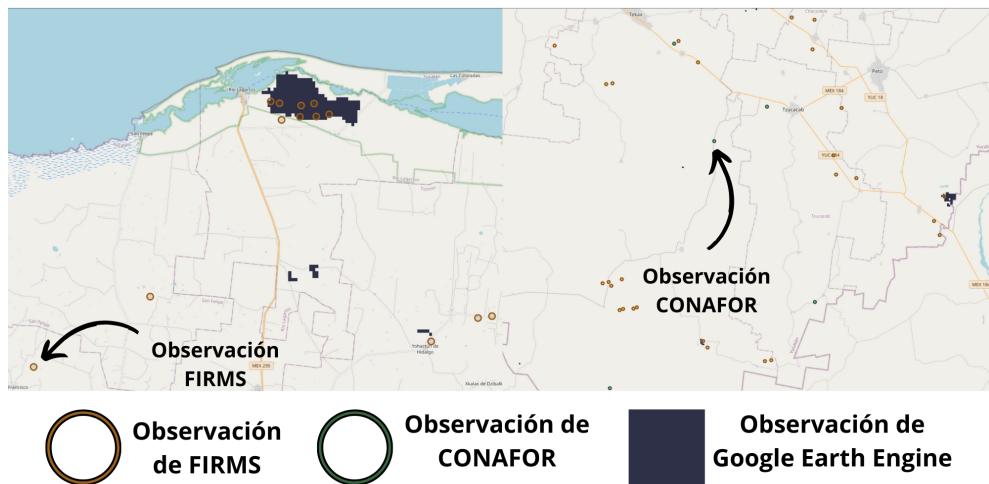


Figura 6.44: **Mapa.** Acercamiento a las observaciones de incendios forestales de la CONAFOR, FIRMS y Google Earth Engine.

Capítulo 7

Conclusiones

En este trabajo se exploraron tres conjuntos de datos con información histórica relacionada a los incendios forestales en el estado de Yucatán. En ellos se encontraron patrones y comportamientos similares, sin embargo, cada uno tuvo información diferente y relevante para el análisis exploratorio e histórico de los datos.

Tras la exploración de los datos de la CONAFOR se puede resaltar lo siguiente:

- Es el conjunto de datos más rico en información, ya que proporciona variables o atributos propios de la región tales como: tipo de vegetación, duración del incendio, el nombre del municipio donde ocurrió y las causas. Este tipo de información ayuda a contextualizar los incendios forestales.
- El atributo de causa es relevante ya que en el mes de Agosto se cuenta con un registro de incendio forestal que afectó 1000 ha, este fue causado por cazadores en un Área Natural Protegida (ANP); también dentro del ANP se tiene el registro de mayor extensión superficial afectada.

Para el caso del conjunto de datos proporcionado por FIRMS:

- Son datos flexibles gracias al atributo del nivel de confianza de las observaciones y de esta manera adecuarlos a las necesidades de cada proyecto.
- Los atributos de temperatura (**brightness**) y de Potencia Radiactiva del Fuego (**frp**) son relevantes ya que están completamente enfocados en dar características a los puntos de calor.
- En este conjunto de datos se obtuvo una baja correlación negativa entre el tiempo y el número de instancias registradas.
- Así como con los datos proporcionados por la CONAFOR para el año 2017, los meses con mayor número de registros de puntos de calor fueron Marzo, Abril y Mayo.

La plataforma de Google Earth Engine es amplia y diversa, contando con diferentes conjuntos de datos con variables diversas y de diferentes rangos temporales. Aunque los conjuntos de datos anteriores daban importante información sobre las instancias o incendios

forestales puntuales, no se contaba con datos sobre los alrededores o el comportamiento de la región en la que ocurrieron.

Con la información que proporcionan las imágenes satelitales se encontraron comportamientos históricos que ayudan a entender la dinámica ambiental y de vegetación de las diferentes regiones y del total del estado de Yucatán. Se pueden resaltar los siguientes aspectos:

- Los meses con temperaturas mayores a la mediana histórica (28°C , periodo 2001-2019) fueron Marzo, Abril y Mayo, los mismos en donde hay un mayor registro de incendios forestales. La región que tuvo una temperatura por arriba de la mediana fue la región norte de Yucatán.
- La humedad del suelo no tiene un comportamiento estacional como la temperatura pero se puede identificar una alta concentración de suelo húmedo en el oriente del estado, cercano a frontera con Quintana Roo.
- La sequía, así como la humedad del suelo, no cuenta con algún comportamiento estacional. Sin embargo, la alta correlación que tiene con la humedad del suelo muestra comportamientos históricos de períodos con un PDSI negativo junto con una humedad del suelo por debajo de la mediana histórica (66.324 mm, periodo 2001-2019) como en los años 2001, 2009, 2016 y 2017.
- Los índices de vegetación como el NDVI y el EVI muestran cambios a través del tiempo en diferentes regiones pero estas son también afectadas por el urbanismo y los comportamientos de la agricultura del estado. Esto no descarta que de manera visual se vea que las áreas afectadas por los incendios forestales se encuentren donde haya un descenso en ambos índices de vegetación.

Identificando el 2017 como uno año con alta sequía y baja humedad del suelo, se enfocó en el mes de Abril para la visualización del comportamiento de las regiones. En ella se observaron las altas temperaturas y el descenso del EVI en el norte del estado; y la alta sequía y baja humedad que cubrió el estado en ese mes.

Se pudo realizar un proceso estadístico y computacional para poder visualizar y explorar diferentes variables ambientales y de vegetación relacionadas a los incendios forestales. Esto no solamente se hizo para conjuntos de datos tabulares y estructurados, también a través de datos no estructurados como las imágenes satelitales, así como llevar la información que contienen (las imágenes satelitales) a un formato estructurado.

La plataforma Google Earth Engine es una herramienta rica en información y con un ambiente de desarrollo (el Editor de Código) amigable para usuarios sin un alto dominio de programación, así como para usuarios con experiencia en el área de Análisis y Ciencia de Datos, a través de su API para Python. Las oportunidades que ofrece esta plataforma van más allá de los incendios forestales, ya que se pueden cubrir diferentes problemáticas ambientales tales como la sequía, disponibilidad del agua, monitoreo de gases de efecto invernadero hasta

problemáticas sociales como los asentamientos humanos y como afectan el ecosistema en el que se desarrollan.

7.1. Trabajos a futuro en la plataforma GEE

Existen aún áreas de oportunidad para futuros trabajos en la plataforma de Google Earth Engine, se mencionan a continuación:

- El conjunto de datos de las áreas afectadas por los incendios forestales no cubre los años 2020 y 2021, por lo que es necesario encontrar otro conjunto de datos para juntar o complementar con los que se usaron en este trabajo.
- Los conjuntos de datos varían en resolución, por lo que resulta complicado poder crear un conjunto de datos vectoriales (shapefile o GeoJSON) usando como polígonos el perímetro de los pixeles de información.
- Es necesario poder complementar la información que se obtiene de esta plataforma con datos de campo para entender la relación entre las imágenes satelitales y la ecología.
- Este trabajo puede ser un primer paso para poder usar algoritmos de Aprendizaje Automático (*Machine Learning*) clásicos o algoritmos más complejos de Aprendizaje Profundo (*Deep Learning*) para la clasificación o segmentación de las áreas afectadas por incendios forestales.

7.2. Aprendizaje durante el desarrollo de la tesis

Durante la realización de este trabajo hubo un gran aprendizaje y mejora en las habilidades dentro de los siguientes rubros:

- **Programación con énfasis en Análisis Exploratorio de Datos:** Durante la realización de esta tesis se fue adquiriendo más conocimiento en la limpieza, procesamiento y descripción de datos estructurados y no estructurados. Las herramientas de estadísticas descriptivas y programación estructurada fueron cimiento para poder construir encima de ellas la metodología para la manipulación de datos a través del lenguaje de programación en Python. Simultáneamente se fue adquiriendo conocimiento para la visualización y comunicación efectiva de los datos dependiendo de su estructura y el impacto que se quiera transmitir. También es importante mencionar que para determinado tipo de gráficas, como por ejemplo las gráficas Ridgeline, se tuvieron que realizar en el ambiente de programación de **R**, ya que resulta más fácil su configuración y presentación.
- **Plataforma Google Earth Engine:** Aprender a manipular los objetos y estructuras de datos de la plataforma de GEE estuvo de la mano con aprender sobre las imágenes satélites. Mucho del trabajo fue facilitado gracias a librerías de Python especializadas en la manipulación, transformación y visualización de estos datos.

- **Aportes a la comunidad:** En el proceso de este trabajo se pudieron realizar aportes y guías en el uso de la plataforma de Google Earth Engine a través de publicaciones divulgativas así como la participación en el *9th International Fire Ecology and Management Congress*, esta evidencias se encuentra adjunta en el Apéndice C. El código de la metodología, el análisis exploratorio de datos y la creación de las visualizaciones se encuentran documentadas en un repositorio de GitHub adjunto en el Apéndice A.

Apéndice A

Repositorio

El código de la tesis puede encontrarse en el siguiente link: https://github.com/isaacarroyov/thesis_undergrad

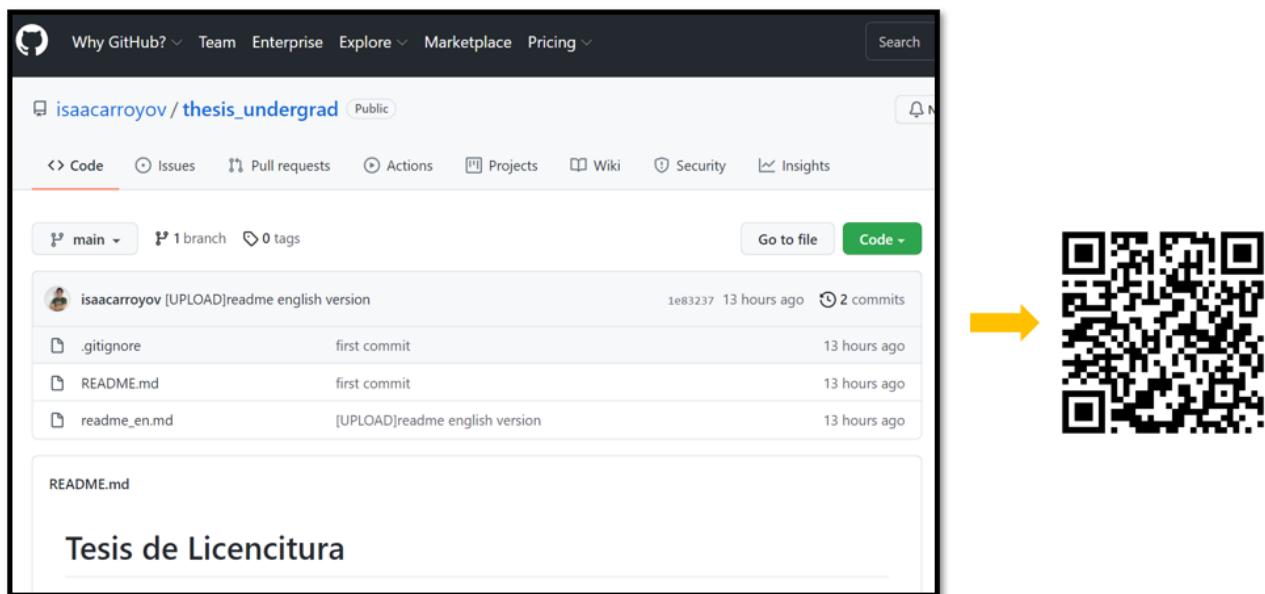


Figura A.1: **Repositorio de GitHub** con el código de programación utilizado en la tesis

Apéndice B

Visualizaciones extra

En este apéndice se comparten visualizaciones que fueron hechas durante el proceso de la tesis que resultan interesantes y no se querían excluir del trabajo. La manera en la que se hacen estas visualizaciones también se encuentran en el repositorio de GitHub de esta tesis (Apéndice A).

Mapas

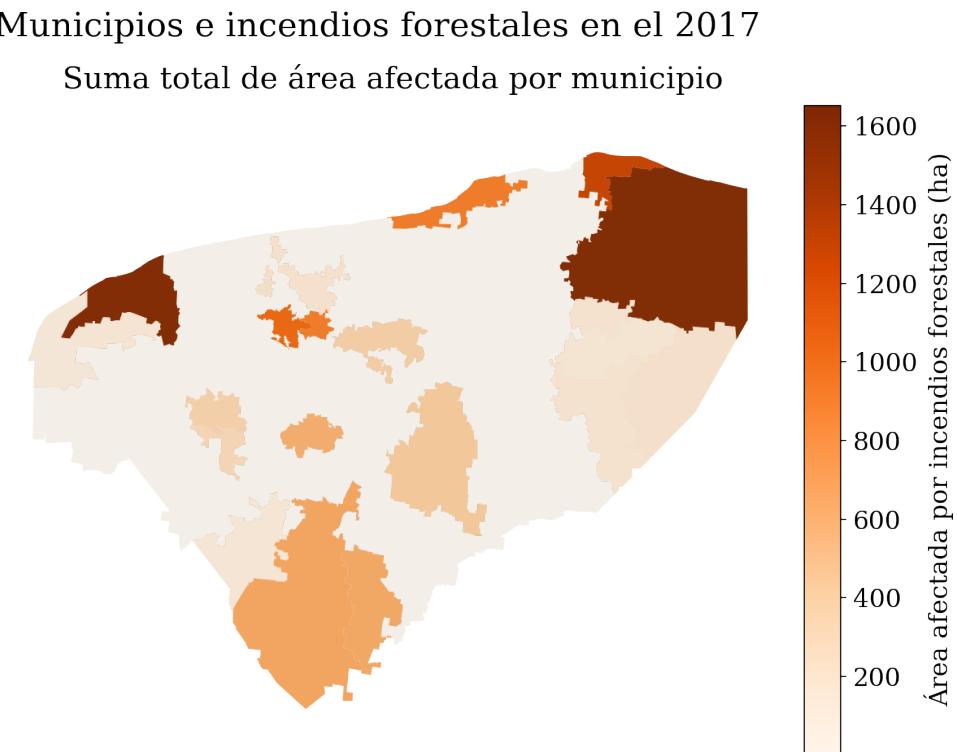
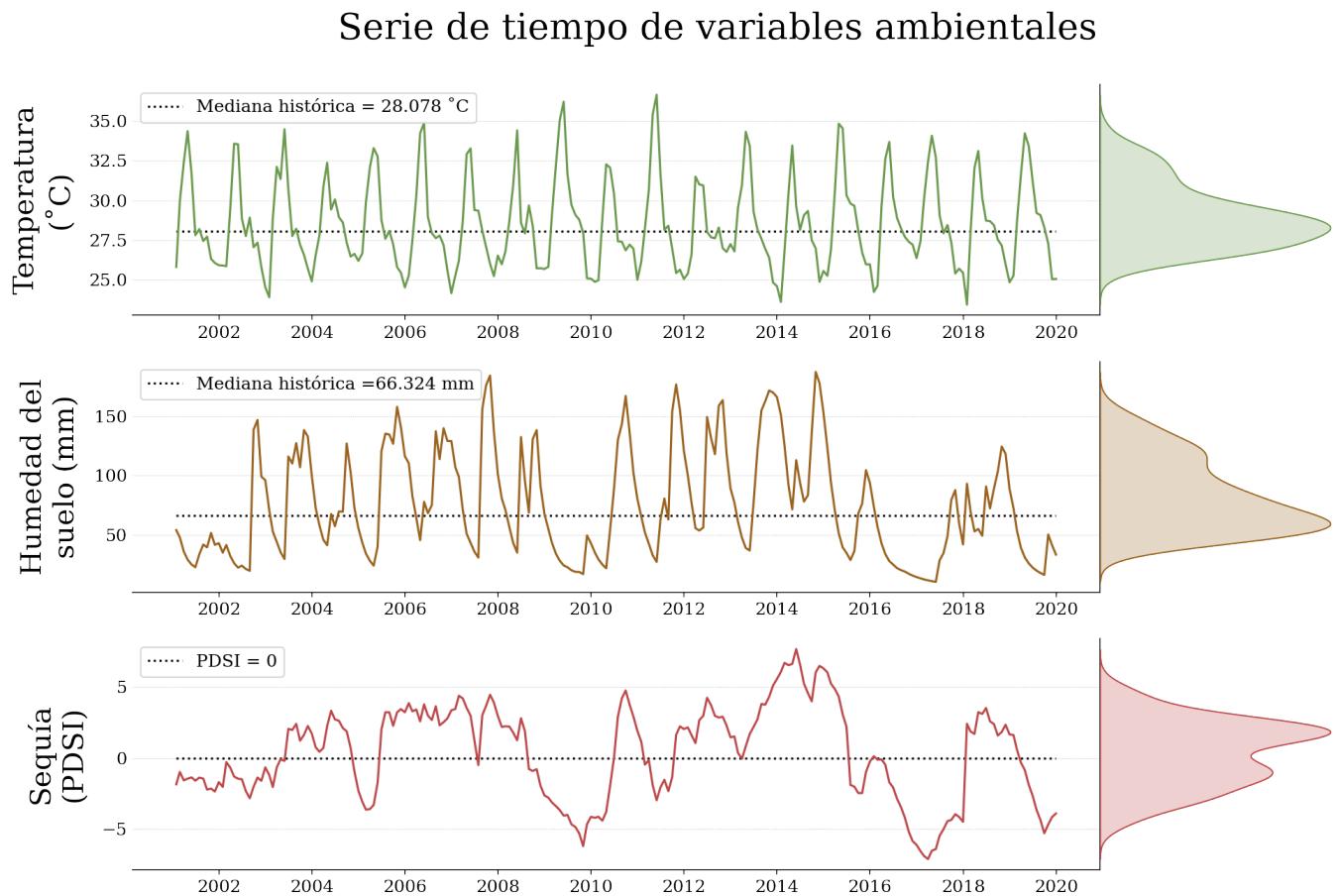


Figura B.1: **Mapa coroplético**. Distribución de área afectada en los municipios de Yucatán en el 2017. Información del conjunto de datos de la CONAFOR.

Serie de tiempo



Fuente: MOD11A1.006 Terra Land Surface Temperature and Emissivity Daily Global 1km y TerraClimate: Monthly Climate and Climatic Water Balance for Global Terrestrial Surfaces, University of Idaho via Google Earth Engine

Figura B.2: **Serie de tiempo y gráfica de densidad** Comportamiento de las variables ambientales: Temperatura de la superficie, humedad del suelo y PDSI. Información obtenida por la plataforma Google Earth Engine.

Diagrama de Sankey

Este tipo de visualización es muy útil cuando se cuentan con datos categóricos que siguen algún flujo, causalidad o son conjuntos relacionales. Son muy útiles para mostrar procesos de principio a fin o una evolución ([Holtz, s.f.-g](#)). Para esta visualización se usa con el objetivo de mostrar las proporciones de datos que hay en los dos atributos.

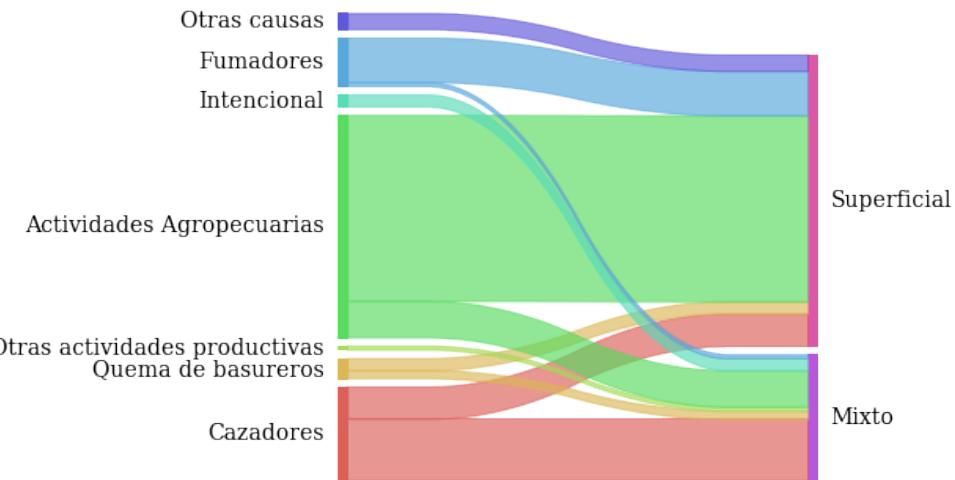


Figura B.3: **Diagrama de Sankey**. Relación entre las causas que ocasionaron el incendio forestal y el tipo de incendio en el que resultó en el estado de Yucatán en el 2017. Información del conjunto de datos de la CONAFOR

Apéndice C

Trabajos realizados durante la tesis

Plataforma Medium

Durante el desarrollo de la tesis los resultados fueron publicados como artículos de divulgación y compartidos con la comunidad global del área de ciencia de datos en la plataforma Medium. El primero se fue el análisis de los datos provenientes de la CONAFOR.

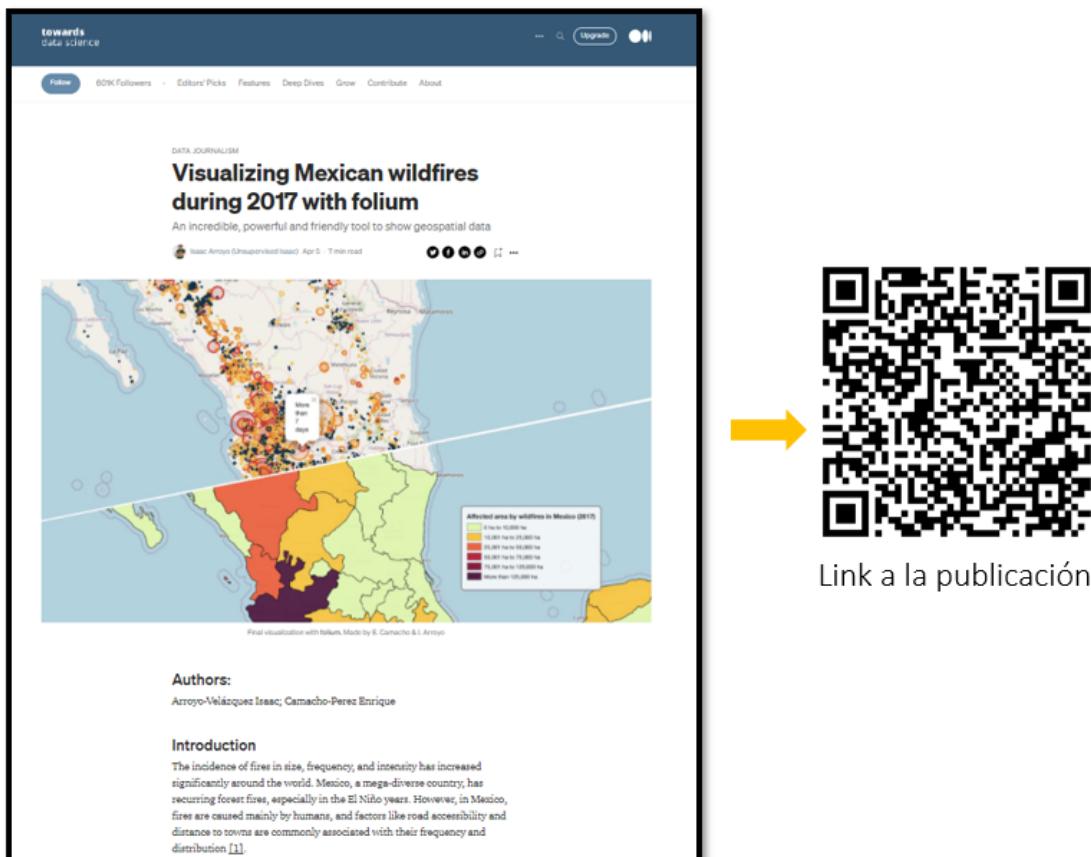


Figura C.1: Publicación en [towardsdatascience](#). Titulo: Visualizing Mexican wildfires during 2017 with folium

La segunda publicación es un tutorial sobre la manipulación y visualización de imágenes satelitales en la plataforma de Google Earth Engine (EE).

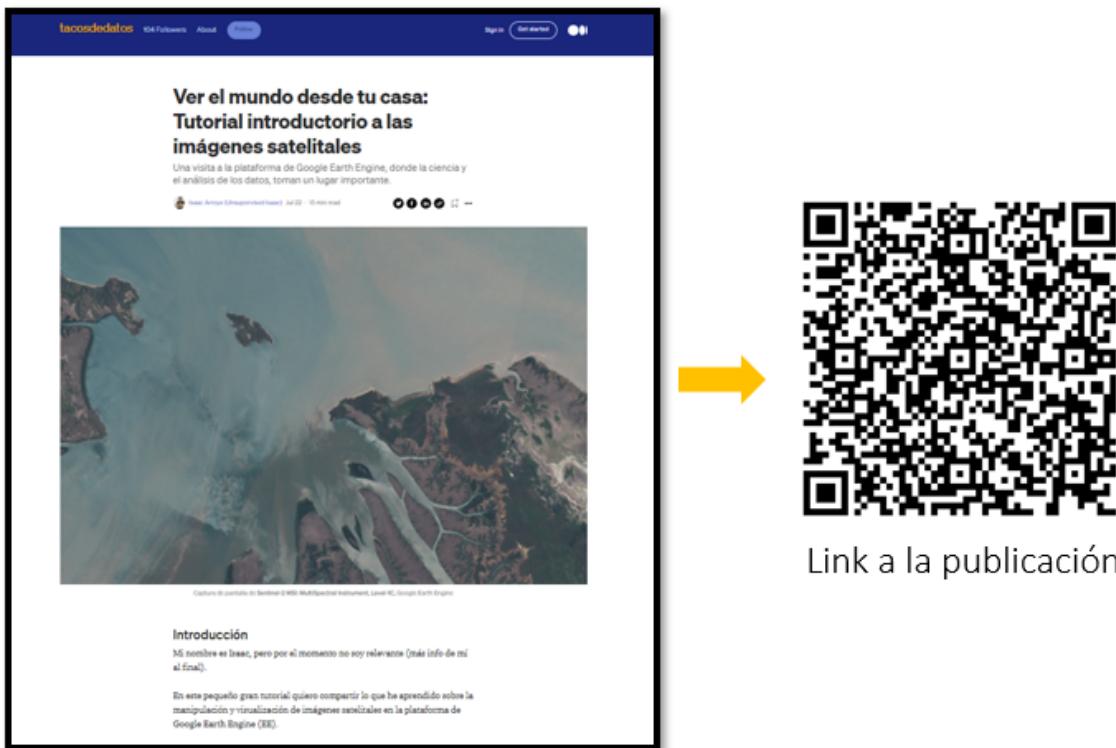


Figura C.2: **Publicación en tacosdedatos.** Titulo: Tutorial introductorio a las imágenes satelitales

9th International Fire Ecology and Management Congress

Una parte de los resultados de la tesis se presentaron en el “9th International Fire Ecology and Management Congress”, llevado a cabo del 30 de noviembre al 3 de diciembre en Florida, EEUU, con la presentación de un poster en formato virtual. El repositorio puede encontrarse en el siguiente link: https://github.com/isaacarroyov/2021_AFE_FireCongress09

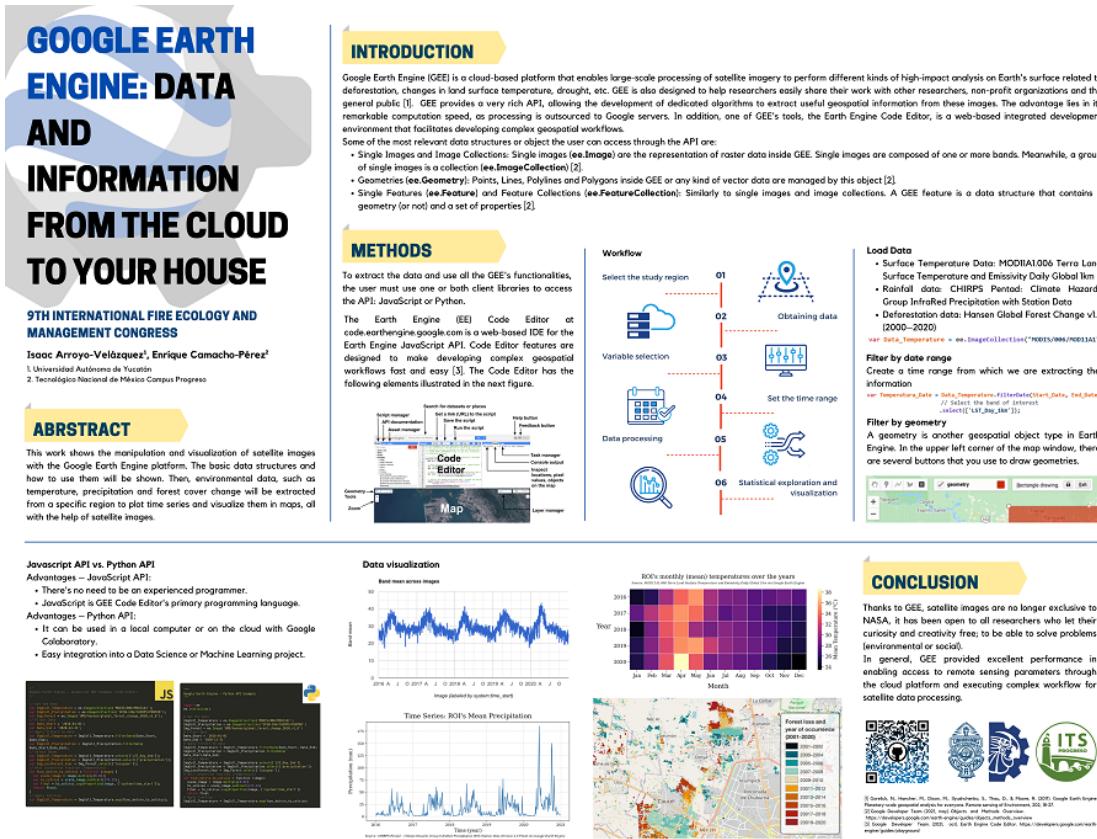


Figura C.3: Poster presentado en el “9th International Fire Ecology and Management Congress”

Referencias

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., y Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(1), 170191. Descargado de <https://doi.org/10.1038/sdata.2017.191> doi: 10.1038/sdata.2017.191 6, 7, 8, 69, 70, 71, 74, 75, 79
- Borra, S., Thanki, R., y Dey, N. (2019). *Satellite Image Analysis: Clustering and Classification*. Springer. 4, 23, 24, 25, 26, 38
- Bruce, P., Bruce, A., y Gedeck, P. (2020). *Practical Statistics for Data Scientist: 50+ Essential Concepts Using R and Python* (Second ed.). O'Reilly Media, Inc. 41, 42, 44, 46
- Ceder, N. (2010). *The Quick Python Book* (Third ed.). Manning Publications. 29, 30
- Comisión Nacional Forestal. (2010). Incendios Forestales: Guía práctica para comunicadores [Manual de software informático]. 4, 17, 18, 19, 20
- CONAFOR. (2016). *El combustible acumulado es el único elemento en el triángulo del fuego que puede ser controlado por el ser humano.* <https://twitter.com/CONAFOR/status/716671871463374848>. (Twitter) 4, 17
- Coordinación General de Conservación y Restauración, y Gerencia del Manejo del Fuego. (2021). *Programa de manejo del fuego 2020-2024*. Descargado de <https://snigf.cnf.gob.mx/incendios-forestales/> 4, 10, 11, 12, 13, 14, 15, 21, 22
- Costa, B. S. C. d., y Fonseca, E. L. d. (2017). The use of fire radiative power to estimate the biomass consumption coefficient for temperate grasslands in the atlantic forest biome. *Revista Brasileira de Meteorologia*, 32, 255–260. 66
- Diccionario prehispánico de dudas. (s.f.). *píxel*. Descargado de <https://www.rae.es/dpd/pixel> 23
- Didan, K. (2015). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006*. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD13Q1.006 7, 8, 76, 77, 80
- Duran García, R., y Méndez Gonzales, M. E. (Eds.). (2010). *Biodiversidad y Desarrollo Humano de Yucatán*. <https://www.cicy.mx/sitios/biodiversidad-y-desarrollo-humano-en-yucatan>: Centro de Investigación Científica de Yucatán. 34
- Earth Science Data Systems Program. (2021, abril). *MCD14DL*. Descargado de <https://earthdata.nasa.gov/earth-observation-data/near-real-time/firms/mcd14dl> 9, 40
- Gobierno de México. (s.f.). *¿Qué hacemos?* Descargado de <https://www.gob.mx/conafor/que-hacemos> 35
- Google Developer Team. (2021a, octubre). *Earth Engine Code Editor*. Descargado de <https://developers.google.com/earth-engine/guides/playground> 33

- Google Developer Team. (2021b, mayo). *FeatureCollection Overview*. Descargado de https://developers.google.com/earth-engine/guides/feature_collections 28
- Google Developer Team. (2021c, mayo). *Feature Overview*. Descargado de <https://developers.google.com/earth-engine/guides/features> 28
- Google Developer Team. (2021d, mayo). *Geometry Overview*. Descargado de <https://developers.google.com/earth-engine/guides/geometries> 28
- Google Developer Team. (2021e, mayo). *ImageCollection Overview*. Descargado de https://developers.google.com/earth-engine/guides/ic_creating 27
- Google Developer Team. (2021f, mayo). *Image Overview*. Descargado de https://developers.google.com/earth-engine/guides/image_overview 27
- Google Developer Team. (2021g, agosto). *JavaScript and Python Guides*. Descargado de <https://developers.google.com/earth-engine/guides> 33
- Google Developer Team. (2021h, mayo). *Objects and Methods Overview*. Descargado de https://developers.google.com/earth-engine/guides/objects_methods_overview 27
- Google Developer Team. (2021i, mayo). *Python Installation*. Descargado de https://developers.google.com/earth-engine/guides/python_install 33
- Google Developer Team. (2021j, mayo). *Reducer Overview*. Descargado de https://developers.google.com/earth-engine/guides/reducers_intro 28
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., y Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18–27. 27, 28
- Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... others (2013). High-resolution global maps of 21st-century forest cover change. *science*, 342(6160), 850–853. 29
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... others (2020). Array programming with numpy. *Nature*, 585(7825), 357–362. 30
- Hill, C. (2020). *Learning scientific programming with Python*. Cambridge University Press. 29
- Holtz, Y. (s.f.-a). *Barplot*. Descargado de <https://www.data-to-viz.com/graph/barplot.html> 5, 48
- Holtz, Y. (s.f.-b). *Density*. Descargado de <https://www.data-to-viz.com/graph/density.html> 5, 46
- Holtz, Y. (s.f.-c). *Histogram*. Descargado de <https://www.data-to-viz.com/graph/histogram.html> 5, 46
- Holtz, Y. (s.f.-d). *Line Plot*. Descargado de <https://www.data-to-viz.com/graph/line.html> 5, 47
- Holtz, Y. (s.f.-e). *Lollipop chart*. Descargado de <https://www.data-to-viz.com/graph/lollipop.html> 5, 48
- Holtz, Y. (s.f.-f). *Ridgeline Plot*. Descargado de <https://www.data-to-viz.com/graph/ridgeline.html> 5, 46, 47
- Holtz, Y. (s.f.-g). *Sankey Diagram*. Descargado de <https://www.data-to-viz.com/graph/sankey.html> 88

- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. doi: 10.1109/MCSE.2007.55 30
- INEGI. (2017). *Conociendo Yucatán*. <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825097349>. 34
- INEGI. (2021a). *Espacio y datos de México*. Descargado de <https://www.inegi.org.mx/app/mapa/espacioydatos/> 5, 34, 35
- INEGI. (2021b). *Panorama sociodemográfico de Yucatán. Censo de Población y Vivienda 2020* (Reporte). <https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=702825198046>: Instituto Nacional de Estadística y Geografía. 34
- Jetz, W., McPherson, J. M., y Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: toward a global map of life. *Trends in ecology & evolution*, 27(3), 151–159. 29
- Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., ... Leblanc, F. (2020, julio). *geopandas/geopandas: v0.8.1*. Zenodo. Descargado de <https://doi.org/10.5281/zenodo.3946761> doi: 10.5281/zenodo.3946761 30
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... development team, J. (2016). Jupyter notebooks ? a publishing format for reproducible computational workflows. En F. Loizides y B. Schmidt (Eds.), *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87–90). IOS Press. Descargado de <https://eprints.soton.ac.uk/403913/> 32
- Lizundia-Loiola, J., Otón, G., Ramo, R., y Chuvieco, E. (2020). A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sensing of Environment*, 236, 111493. 6, 7, 8, 67, 68, 70, 71, 74, 75, 78, 79, 80
- MDN contributors. (2021, noviembre). *Acerca de JavaScript*. Descargado de https://developer.mozilla.org/es/docs/Web/JavaScript/About_JavaScript 33
- Montero, D. (2021). eemont: A python package that extends google earth engine. *Journal of Open Source Software*, 6(62), 3168. 33
- Moreno-Casasola, P., Cejudo-Espinosa, E., Capistrán-Barradas, A., Infante-Mata, D., López-Rosas, H., Castillo-Campos, G., ... Campos-Cascaredo, A. (2010). Composición florística, diversidad y ecología de humedales herbáceos emergentes en la planicie costera central de veracruz, méxico. *Boletín de la sociedad botánica de México*(87), 29–50. 57
- NASA. (s.f.). *Fire Information for Resource Management System*. Descargado de <https://firms.modaps.eosdis.nasa.gov> 36
- Nelson, M. J., y Hoover, A. K. (2020). Notes on using google colaboratory in ai education. En *Proceedings of the 2020 acm conference on innovation and technology in computer science education* (pp. 533–534). 32
- Nosrati, M. (2011). Python: An appropriate language for real world programming. *World Applied Programming*, 1(2), 110–117. 29
- python visualization. (2021, enero). *Folium*. Descargado de <https://python-visualization.github.io/folium/> 31
- R Core Team. (2021). R: A Language and Environment for Statistical Computing [Manual de software informático]. Vienna, Austria. Descargado de <https://www.R-project.org/> 32
- Raj, H. (2018). *MATLAB — RGB image representation*. Descargado de <https://www>

- .geeksforgeeks.org/matlab-rgb-image-representation/ 4, 24
- Ramos Reyes, J. C. (2015). *Análisis del crecimiento, composición y estructura de un bosque de coníferas: relación del fuego y variabilidad climática en el cerro el potosí* (Tesis Doctoral no publicada). Universidad Autónoma de Nuevo León. 19
- Ribecca, S. (s.f.-a). *Box and Whisker Plot*. Descargado de https://datavizcatalogue.com/methods/box_plot.html 5, 50
- Ribecca, S. (s.f.-b). *Choropleth Map*. Descargado de <https://datavizcatalogue.com/methods/choropleth.html> 5, 51
- Ribecca, S. (s.f.-c). *Heat map*. Descargado de <https://datavizcatalogue.com/methods/heatmap.html> 5, 49
- Secretaría de Desarrollo Sustentable de Yucatán. (s.f.). *Comité de prevención y combate de incendios agropecuarios y forestales del estado de yucatán*. Descargado de <https://sds.yucatan.gob.mx/consejos-comites/incendios.php> 22
- Seydi, S. T., Akhoondzadeh, M., Amani, M., y Mahdavi, S. (2021). Wildfire damage assessment over australia using sentinel-2 imagery and modis land cover product within the google earth engine cloud platform. *Remote Sensing*, 13(2), 220. 36
- Skiena, S. S. (2017). *The data science design manual*. Springer. 23
- Sulova, A., y Jokar Arsanjani, J. (2021). Exploratory Analysis of Driving force of Wildfires in Australia: An Application of Machine Learning within Google Earth Engine. *Remote Sensing*, 13(1), 10. 36, 37
- The pandas development team. (2020, febrero). *pandas-dev/pandas: Pandas*. Zenodo. Descargado de <https://doi.org/10.5281/zenodo.3509134> doi: 10.5281/zenodo.3509134 30
- U.S. Geological Survey. (s.f.). *Landsat Enhanced Vegetation Index*. Descargado de https://www.usgs.gov/core-science-systems/nli/landsat-landsat-enhanced-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con 38
- Vega-Nieva, D. J., Briseño-Reyes, J., Nava-Miranda, M. G., Calleros-Flores, E., López-Serrano, P. M., Corral-Rivas, J. J., ... others (2018). Developing models to predict the number of fire hotspots from an accumulated fuel dryness index by vegetation type and region in Mexico. *Forests*, 9(4), 190. 21
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... others (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3), 261–272. 30
- Wan, Z., Hook, S., y Hulley, G. (2015). *MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006*. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD11A1.006 6, 7, 8, 69, 70, 74, 78
- Waskom, M. L. (s.f.). *seaborn.replot*. Descargado de <https://seaborn.pydata.org/generated/seaborn.replot.html#seaborn.replot> 4, 31
- Waskom, M. L. (2021). *seaborn: statistical data visualization*. *Journal of Open Source Software*, 6(60), 3021. Descargado de <https://doi.org/10.21105/joss.03021> doi: 10.21105/joss.03021 31
- Wells, N., Goddard, S., y Hayes, M. J. (2004). A self-calibrating palmer drought severity index. *Journal of climate*, 17(12), 2335–2351. 71
- Wes McKinney. (2010). Data Structures for Statistical Computing in Python. En Stéfan

- van der Walt y Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (p. 56 - 61). doi: 10.25080/Majora-92bf1922-00a [30](#)
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. Descargado de <https://ggplot2.tidyverse.org> [32](#)
- Wolters, C. (2020, julio). *Here's how wildfires get started—and how to stop them.* <https://www.nationalgeographic.com/environment/natural-disasters/wildfires/>. (National Geographic) [17](#), [18](#), [19](#)
- Wu, Q. (2020). geemap: A python package for interactive mapping with google earth engine. *Journal of Open Source Software*, 5(51), 2305. [33](#)
- Young, H. D., y Freedman, R. A. (2009). *Física Universitaria con Física Moderna* (12th ed.). Pearson Educación. [4](#), [24](#), [25](#)
- Zamora Crescencio, P., García Gil, G., Flores Guido, J. S., y Ortiz, J. J. (2008). Estructura y composición florística de la selva mediana subcaducifolia en el sur del estado de yucatán, méxico. *Polibotánica*(26), 39–66. [57](#)
- Zhao, Q., Yu, L., Li, X., Peng, D., Zhang, Y., y Gong, P. (2021). Progress and trends in the application of google earth and google earth engine. *Remote Sensing*, 13(18), 3778. [28](#)
- Zwillinger, D. (2018). *CRC Standard Mathematical Tables and Formulas* (33th ed.). Chapman and Hall/CRC. [43](#)