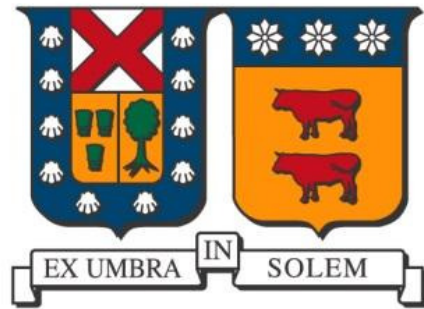


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
VALPARAÍSO - CHILE



“DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA PARA EL MANEJO DE DATOS ALTIMÉTRICOS PARA EL PROYECTO CHILE 3D”

MAXIMILIANO AUBEL OLIVARES

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Javier Cañas
Profesor Correferente: Luis Salinas

Agosto - 2023

DEDICATORIA

A mi familia, amigos, profesores,
especialmente a las personas que han creído en mí.
A Sysmic Robotics y a la Primera.

RESUMEN

Resumen— Los datos altimétricos son esenciales para la industria de las obras civiles, minería, estudios de desastres naturales, agronomía, entre otros. Esto los hace datos valiosos, los que requieren alta disponibilidad y precisión para un fluido desarrollo. No obstante, en Chile no existe una plataforma con dichas características, que presente y centralice los datos disponibles de elevación y de superficie al usuario. Chile 3D nace de la iniciativa de generar una plataforma de recolección, administración, formateo, y disponibilización de datos altimétricos de Chile. La presente investigación se enfocará en los aspectos de formateo de los datos, su respectivo manejo, y transformación de acuerdo a las necesidades.

Palabras Clave— Datos altimétricos; LiDAR; almacenamiento de datos; transformación

ABSTRACT

Abstract— Hypsometric data is essential for the civil engineering industry, mining, natural disasters, agronomy, and so on. This makes them highly valuable data, which requires good availability and precision for a fluid development of said areas of study. Nonetheless, there is no platform in Chile capable of meeting these criteria, which presents and centralizes the available elevation maps to the user. Chile 3D arises from the idea of generating a platform to collect, administrate, format, and make available altimetric data from Chile. The present investigation will focus on the format and management of the data, and transformation according to the needs.

Keywords— *Hypsometric data; LiDAR; data storage; transformation*

GLOSARIO

API: Application Programming Interface.
CCHEN: Comisión Chilena de Energía Nuclear.
CNIG: Organismo Autónomo Centro Nacional de Información Geográfica de España.
CONAF: Corporación Nacional Forestal.
CRUD: Create Read Update Delete.
CSN: Centro Sismológico Nacional.
DEM: Digital Elevation Model.
DGA: Dirección General de las Aguas.
DOH: Dirección de Obras Hidráulicas.
DSM: Digital Surface Model.
DTM: Digital Terrain Model.
DWT: Discrete Wavelet Transform.
ECW: ENhanced Compression Wavelet.
GIS: Geographic Information System.
GUI: Graphic User Interface.
IDE: Infraestructura de Datos Geoespaciales.
IFSAR: Interferometric Synthetic Aperture Radar.
LAS: LASer (formato para el intercambio y archivo de nubes de puntos LiDAR).
LAZ: LASzip (Algoritmos de compresión de LAS).
LCMP: LASCompression (algoritmo de compresión de LAS).
LiDAR: Light Detection and Ranging o Laser Imaging Detection and Ranging.
NEEA: Northwest Energy Efficiency Alliance.
NRC: National Research Council.
OOTT: Organismos Técnicos.
PDAL: Point Data Abstraction Library.
QL: Quality Level.
SERNAGEOMIN: Servicio Nacional de Geología y Minería.
SHOA: Servicio Hidrográfico y Oceanográfico de la Armada de Chile.
SIRGAS2000: Sistema de Referencia Geodésico para las Américas 2000.
WGS84: World Geodesic System 1984.

ÍNDICE DE CONTENIDOS

RESUMEN III

ABSTRACT III

GLOSARIO IV

ÍNDICE DE FIGURAS VI

ÍNDICE DE TABLAS VII

INTRODUCCIÓN 1

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA 3

 1.1 Objetivo General 5

 1.2 Objetivos específicos 5

CAPÍTULO 2: MARCO CONCEPTUAL 6

CAPÍTULO 3: PROPUESTA DE SOLUCIÓN 10

 3.1 Descripción general de la plataforma Chile 3D 11

 3.2 Adquisición y almacenamiento de datos 12

 3.2.1 Almacenamiento de datos 13

 3.3 Submuestreo de datos altimétricos 17

 3.3.1 Submuestreo de Nubes de Puntos 17

 3.3.2 Submuestreo de Ráster 21

 3.4 Conversión de Nubes de Puntos LAS a Ráster TIFF 23

 3.5 Compresión de datos altimétricos 26

 3.5.1 Compresión de Nubes de Puntos LAS 26

 3.5.2 Compresión de Ráster TIFF 28

 3.6 INTEGRACIÓN CON API 32

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN 35

CAPÍTULO 5: CONCLUSIONES 42

REFERENCIAS BIBLIOGRÁFICAS 46

ÍNDICE DE FIGURAS

1	Extracción de información desde la captura de modelos tridimensionales. . . .	1
2	Comparación de resolución de datos de un DEM con resolución espacial de 30 arc-sec, 3 arc-sec, y 1 arc-sec.	4
3	Comparación entre modelos DEM y DSM (DTM).	7
4	Comparación entre un geoide y un elipsoide.	8
5	Modelo de dominio: vista funcional inicial Chile 3D.	10
6	Caso de ejemplo de uso.	11
7	Captura de nube de puntos de [USGS, 2017]	17
8	Muestreo de nubes de puntos usando celdas voxel [Wang <i>et al.</i> , 2018].	18
9	Efecto del método Sample que utiliza el submuestreo de Poisson [Smathermather, 2017].	19
10	Tiempo de cómputo	20
11	Tamaño de archivo de salida	20
12	LAZ de Cook County con un tamaño de celda de 19 pies	21
13	Ráster de la localidad de Paipote en Copiapó [Ministerio de Bienes Nacionales, 2017], sobrepuesto a un mapa de Openstreetmap.	22
14	Tiempo de cómputo	22
15	Tamaño de archivo de salida	22
16	Tiempo de cómputo en función de la resolución (tamaño de la celda) y el radio de búsqueda.	24
17	Celda: 1, Radio: 3	25
18	Celda: 19, Radio: 13	25
19	Celda: 1, Radio: 13	25
20	Celda: 19, Radio: 23	25

21 Tiempo de cómputo para reducción de celda de la Figura 13, en función de el porcentaje de escalado. 28

22 Acercamiento al gráfico de la Figura 21 desde escalados 0,05 a 0,2. 29

23 Tamaño de archivo de salida en disco para reducción de celda de la Figura 13, en función de el porcentaje de escalado. 30

24 Captura de las 30 celdas GeoTiff que componen el registro de la zona de Alberta. 36

25 Nube de puntos LAS original. 38

26 Ráster del proveedor. 38

27 Ráster obtenido con implementación. 38

28 LAS original 39

29 LAZ con celdas de 2 metro 39

30 LAS original 40

31 LAZ con celdas de 2 metro 40

32 GeoTiff original 41

33 GeoTiff escalado al 0.5 41

ÍNDICE DE TABLAS

1 Tamaño de archivos estimados para U.S.A. por nivel de calidad. 8

2 Escenarios de planificación por cada QL de la NEEA. 9

3 Densidad poblacional por región 13

4 Tamaño de archivo total estimado para Chile por nivel de calidad. 14

5 Comparación de desempeño entre algoritmos de compresión aplicados a archivos LAS “de naturaleza típica, experimental, e inusual” [Isenburg, 2013]. . 27

6 Comparación temporal de desempeño entre algoritmos de compresión aplicados a archivo LAS de Nueva York, Estados Unidos [USGS, 2014]. 27

7 Comparación espacial para un escalado de 0,5 del ráster de la Figura 13 31

8 Comparación Archivo LAS original y archivo TIFF procesado. 37

9 Comparación Archivo LAS original y archivo LAZ procesado. 40

10 Comparación Archivo GeotTIFF original y archivo procesado con escala 0.5 . . . 41

INTRODUCCIÓN

Un dato altimétrico indica la altura de un punto geográfico específico respecto a una georreferencia. La combinación de varios datos altimétricos ordenados permite generar mapas que cubren superficies de terreno. Estos modelos tridimensionales permiten realizar estudios en diversas disciplinas, tales como las obras civiles e.g. para el diseño de construcciones; minería, para estudios de terreno; agronomía; desastres naturales; entre otros. Por lo tanto, son datos altamente versátiles, cuya disponibilización significa un aporte en diversas materias.

Existen diversos métodos de captura de datos altimétricos, como por ejemplo captura de imágenes estereoscópicas, nubes de puntos a través de sensores LiDAR, etcétera. Con cada uno de estos métodos se logra obtener modelos tridimensionales que permiten realizar los estudios antes mencionados. En la Figura 1 podemos observar un ejemplo de extracción de información, que ocurre desde la entrada de datos (e.g. un sobrevuelo a la zona de interés con aviones equipados con sensores LiDAR), posterior análisis y extracción de objetos.

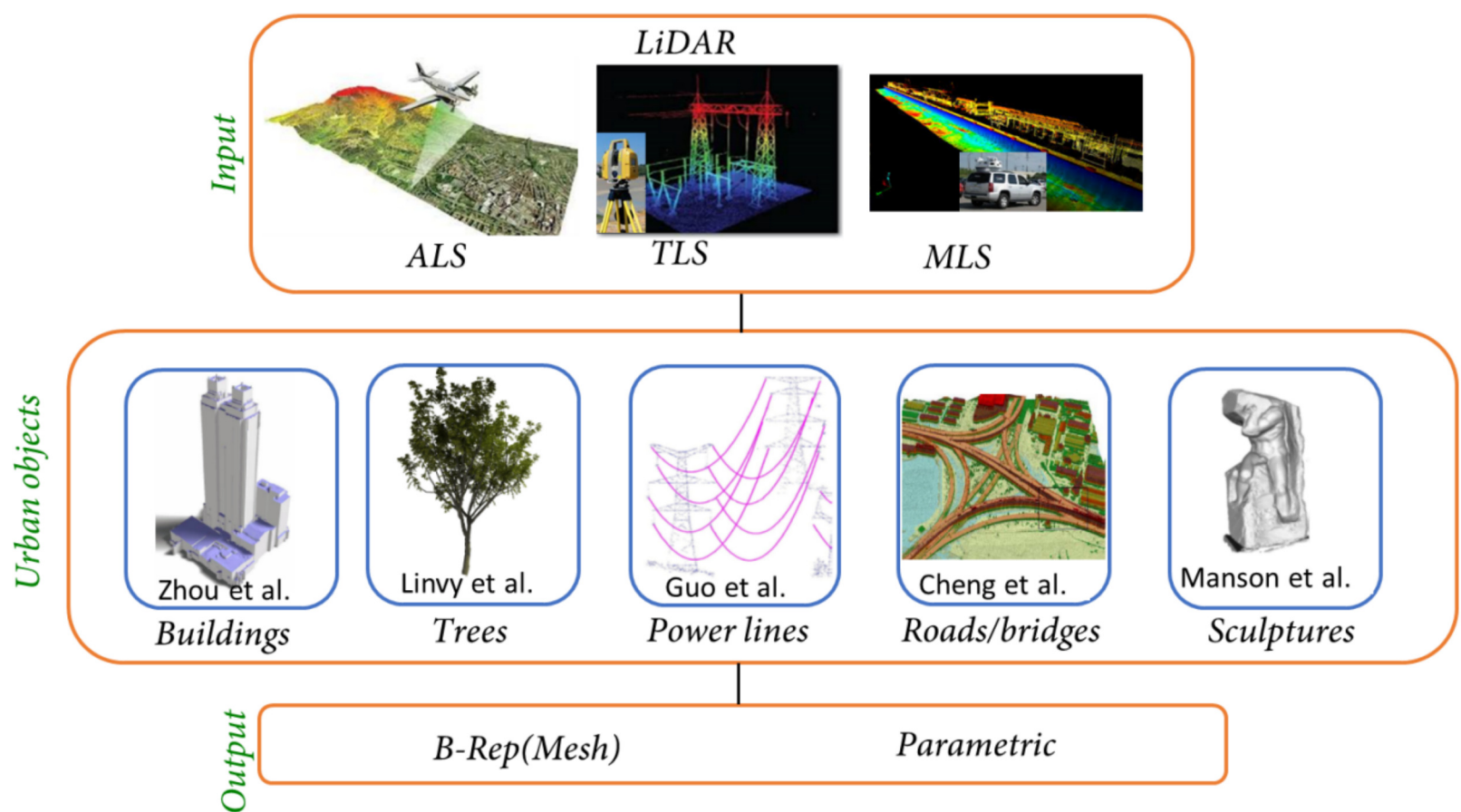


Figura 1: Extracción de información desde la captura de modelos tridimensionales.

Fuente: [Wang et al., 2018].

No obstante, al día de hoy en Chile no se cuenta con un sistema de que almacene datos de calidad de toda la superficie nacional, de donde se identifican varias causas principales que serán ahondadas en el presente documento. Entre ellas se destaca la falta de datos con los que poblar dicho sistema, es decir, un proyecto de vuelos para poder capturar la completitud del territorio nacional con imágenes LiDAR de alta resolución.

El objetivo de esta memoria se centra en el diseño e implementación de un sistema para el

manejo de datos altimétricos de Chile para el proyecto Chile 3D. Este proyecto busca generar una plataforma de administración de datos altimétricos, desde la recolección, el manejo, hasta su disponibilización al usuario final.

En primera instancia se abordará el estado de arte de los datos altimétricos en Chile y el mundo, así como una contextualización de su importancia. Posteriormente, se presentará un estudio técnico de soluciones extranjeras para el problema en cuestión. Luego se estudiarán las consideraciones de diseño para implementar una solución a la realidad nacional, tras ello, se implementará la propuesta. Finalmente, se evaluará el desempeño y se integrará con las otras partes que constituyen el proyecto. Luego, se expondrá el trabajo a futuro.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

Se entiende por datos altimétricos aquellos que señalan la altura de un punto respecto a un plano de referencia [Rafferty, 2009]. Con ello es posible representar tanto el terreno como la superficie, útil en la industria de las obras civiles al permitir realizar levantamientos topográficos al momento del diseño y construcción de una edificación. En desastres naturales resulta importante para el estudio de avance de tsunamis, inundaciones, entre otros. Estos son breves ejemplos de lo amplio que es el campo de utilidad de dichos datos altimétricos.

En Chile no existe un levantamiento concreto de los requerimientos nacionales de datos altimétricos de alta resolución por los servicios públicos y/o la comunidad científica. No obstante, el proyecto Chile3D realizó en diciembre de 2021 una encuesta a investigadores y comunidad científica, cuyo objetivo era identificar los requerimientos de cobertura espacial, precisión, resolución y frecuencia de actualización, e identificar los beneficios futuros que significaría contar con dichos datos altimétricos de alta resolución. De los 77 encuestados, 39 de ellos con grado de doctor (50,2 %) y 18 % con grado de magíster (23,4 %) concordaron, que, contar con estos datos aportaría altos beneficios tanto en ahorro de tiempo como costos, generarían mejoras en los resultados de investigación, apoyarían la formulación de nuevos proyectos, y permitirían obtener altos beneficios sociales, en educación, divulgación, y seguridad pública¹.

Por otro lado, dos reportes del Consejo Nacional de Investigación (NRC) de Estados Unidos: “Datos de elevación para el mapeo de inundaciones” [Council, 2007]), y “Mapeando la zona: mejorando la precisión de los mapas de inundación” [Council, 2009], llegaron a la conclusión de que los datos topográficos precisos eran el factor más importante para el mapeo del riesgo de inundaciones, en la determinación del riesgo de inundaciones, en la determinación de la elevación de la superficie del agua, de la elevación base y en la extensión de la inundación [Dewberry, 2012].

Chile es un país de gran diversidad geográfica, que constantemente está afecto a eventos y fenómenos naturales con implicancias mayores en la calidad de vida de su población. Múltiples áreas de desarrollo requieren del uso de datos altimétricos para el análisis, planificación y monitoreo de diversas variables, dentro de las cuales se encuentran la gestión de desastres naturales y los mapas de riesgo, el monitoreo de estabilidad de faenas, la planificación urbana, catastros ambientales, forestales y agrícolas, entre otros. Si bien, existen fuentes de libre acceso que permiten descargar datos, estos se caracterizan por su baja resolución espacial, lo cual limita la detección de detalles en importantes análisis debido a la escala mínima cartografiable que su píxel permite, como se observa en la Figura 2, que ejemplifica el mismo proceso bajo 3 DEM de distinta resolución espacial (resolución de 30 arco segundos, 3 arc-

¹Ver iniciativa Proyecto Chile 3D, www.proyectochile3d.cl y resultados resumidos de la encuesta en <https://bit.ly/3tgkwBn>

sec, y 1 arc-sec), razón por la cual, el requerimiento de datos altimétricos de mejor calidad en cuanto a su detalle se hace indispensable para la obtención de mejores productos y análisis derivados de ellos.

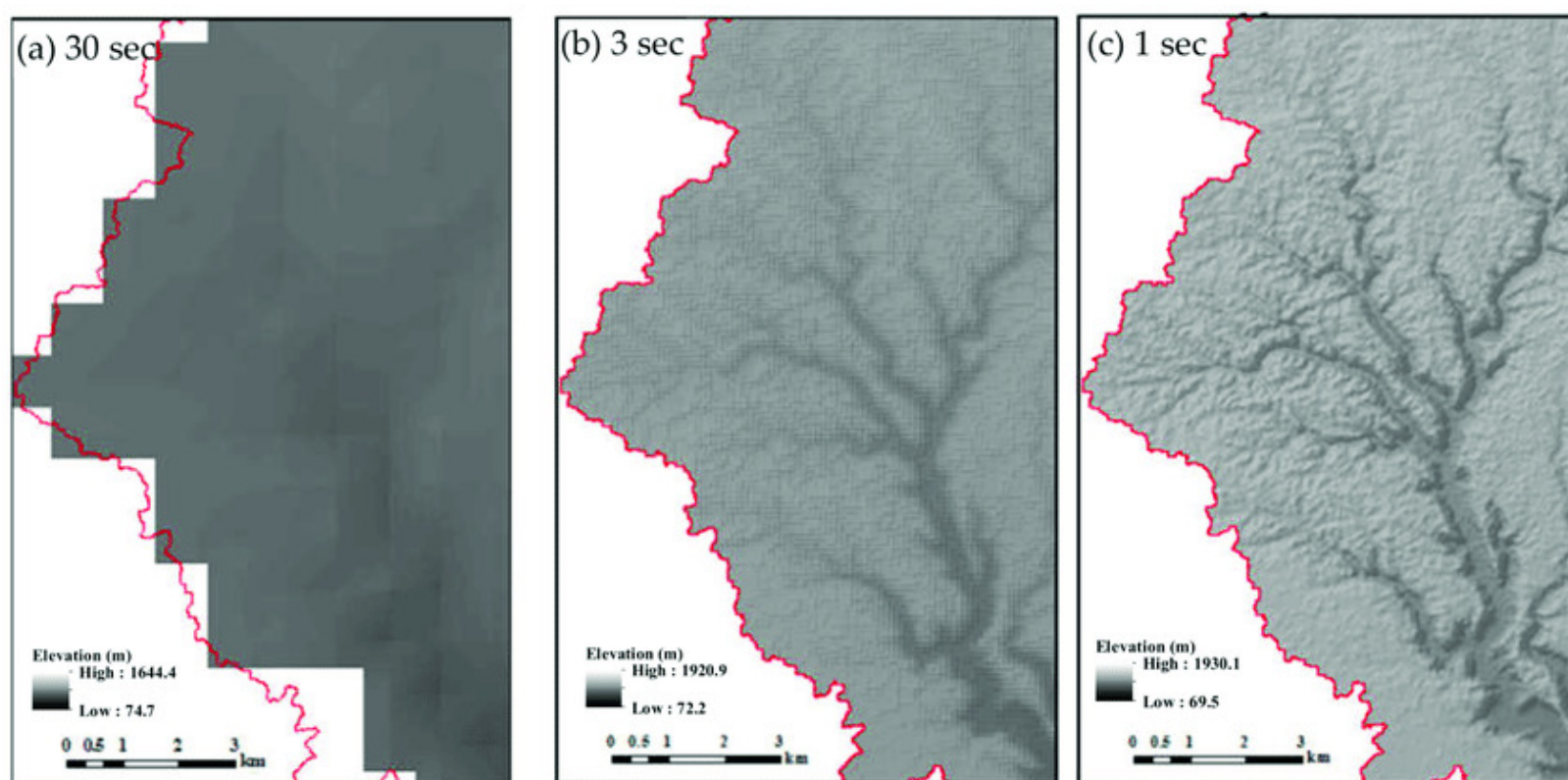


Figura 2: Comparación de resolución de datos de un DEM con resolución espacial de 30 arc-sec, 3 arc-sec, y 1 arc-sec.

Fuente: [Abdel-Fattah *et al.*, 2017]

Sin embargo, el requerimiento de estos datos de alta resolución como insumos se volverá más importante al momento de que entre en operación el reglamento que establece la nueva Ley N° 21.364 [Ministerio del Interior y Seguridad Pública, 2021]. Esta ley establece que los organismos técnicos (OO.TT.), que correspondan según sus competencias, serán los responsables de elaborar, validar y actualizar los “Mapas de Amenaza”, que son instrumentos de gestión definidos en la ley y que además serán utilizados “para la elaboración de los instrumentos de planificación territorial, además de la Planificación del Borde Costero, el Ordenamiento Territorial y el Manejo Integrado de Cuencas” (Art. 35). Los organismos técnicos para el monitoreo de las amenazas que nombra la ley son: la Dirección Meteorológica de Chile, el Servicio Hidrográfico y Oceanográfico de la Armada (SHOA), el Servicio Nacional de Geología y Minería (SERNAGEOMIN), la Corporación Nacional Forestal (CONAF) o su sucesor legal, el Centro Sismológico Nacional (CSN), la Dirección General de Aguas (DGA), la Dirección de Obras Hidráulicas (DOH), Bomberos de Chile, la Comisión Chilena de Energía Nuclear (CCHEN), y otros que pudiera señalar el reglamento (Art. 38). Pese a lo descrito anteriormente, no existe una plataforma que satisfaga lo indicado en Chile.

1.1. Objetivo General

Construir un sistema de almacenamiento, manejo, y formateo de datos altimétricos para el proyecto Chile 3D.

1.2. Objetivos específicos

- Implementar algoritmos de cambio de formato para modelos georreferenciados de altimetría.
- Construir herramientas centradas en la reducción de resolución y submuestreo de datos georreferenciados.
- Desarrollar funcionalidades fácilmente integrables con las otras partes del sistema del proyecto Chile 3D.

CAPÍTULO 2

MARCO CONCEPTUAL

Se dividen los datos altimétricos en dos grandes grupos: Modelos Digitales de Elevación (DEM), y Modelos Digitales de Superficie. DEM corresponde a la elevación del terreno desnudo, mientras los Modelos Digitales de Superficie (DSM) incluyen tanto el terreno como aquellas características ajenas al mismo (vegetación, edificaciones, etc.). Sumado a lo anterior, existen tipos de datos geospaciales de dos familias: datos vectoriales, y datos rasterizados. Cada uno con sus ventajas y usos específicos. Cada modelo identifica características que permiten una georreferenciación, así como una yuxtaposición de capas para un análisis que lo requiera. Respecto a la fuente de datos, estos pueden originarse desde radares satelitales, fotogrametría, estereoscopia, interferometría, láser, entre otros. En este último es posible identificar la tecnología LiDAR, que mediante pulsos de luz genera un nube de puntos desde donde se puede identificar y aislar el terreno de la vegetación luego de ser procesados los datos.

En la Figura 3 vemos una comparación entre DEM (en esquema se utiliza el análogo *Digital Terrain Model* (DTM)) con modelos DSM. Cuando un sensor LiDAR realiza una lectura, algunos pulsos de luz son capaces de pasar a través de los espacios libres de vegetación. De esta forma, podemos generar una nube de puntos, teniendo datos tanto de la superficie terrestre, como de vegetación a diferentes niveles de altura.

A priori, lo importante de los modelos tridimensionales a lo que se le deberá poner atención al momento de su almacenaje y manejo son la georreferencia y el sistema de coordenadas. Para su análisis, un modelo será posicionado sobre o junto a otro modelo, por ello, es menester tener las referencias que le permitirán al usuario saber precisamente cómo ubicar estas imágenes en un Software GIS (*Geographic Information System*) por ejemplo. Algunos sistemas de referencias geodésicos (o *datum*) frecuentemente usados son WGS84 o su alternativa creada para América del Sur, SIRGAS 2000.

Con el objetivo de ejemplificar en qué consisten estos modelos, se describirá WGS84. Este es capaz de representar cualquier punto en el planeta mediante 3 coordenadas x, y, z, mediante una proyección del globo a un elipsoide centrado en su centro de masa. El elipsoide posee un semieje mayor y menor que permiten acercarse a las dimensiones reales de la Tierra. No obstante, zonas diferentes deben ser representadas por sistemas que se aproximan de mejor manera a las características de su terreno. Es por ello que surgen modelos como el antes mencionado SIRGAS 2000, calculado inicialmente para América del Sur, hoy usado en las Américas. En la Figura 4 vemos una comparación entre la topografía real, una proyección de un geoide y un elipsoide. Esto es un ejemplo de la importancia de que el datum sea el indicado para la zona geográfica que se está trabajando.

En el reporte que analiza la *Northwest Energy Efficiency Alliance* (NEEA) elaborado por la consultora Dewberry [Dewberry, 2012] para documentar los requerimientos nacionales pa-

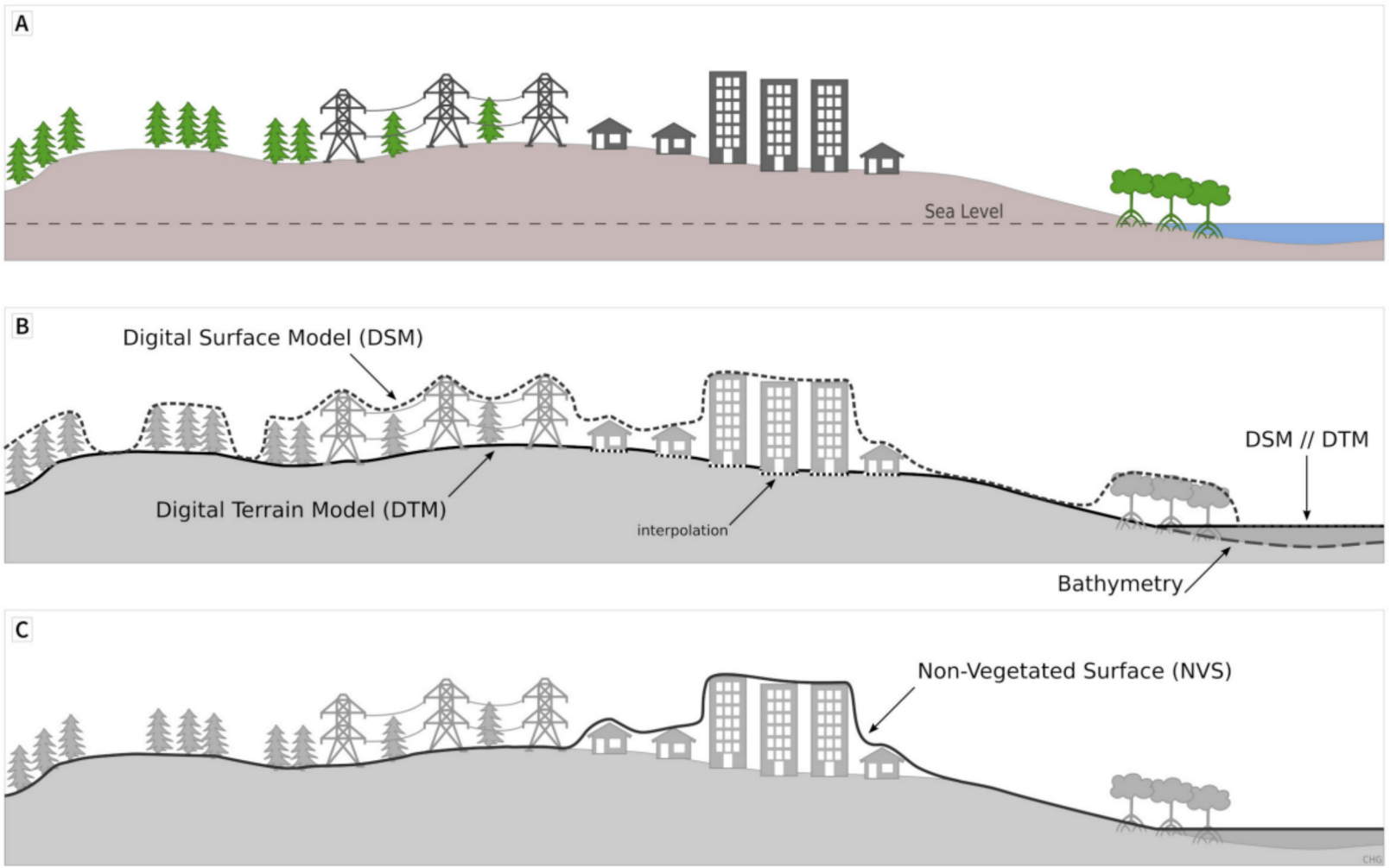


Figura 3: Comparación entre modelos DEM y DSM (DTM).
Fuente: [Guth et al., 2021]

ra una data de elevación mejorada, estimar costos y beneficios, y evaluar los diferentes escenarios de implementación; específicamente en el apéndice H, se habla sobre las especificaciones técnicas que debería tener este sistema. Entre ellas destacan por ejemplo que para poder almacenar un mapa altimétrico completo de Estados Unidos, se necesitaría un plan de adquisición (de acuerdo al nivel de calidad que se requiera) de entre 8 a 25 años. En este reporte se establecen los niveles de calidad (QL) en la Tabla 1, que dependen de la densidad de puntos por metro cuadrado, y que resultan en un diferente tamaño del archivo LAS que requiere ser almacenado. El tamaño de resultado esperado se calculó considerando que un archivo LAS requiere usualmente de 20 a 30 Bytes por punto, por lo que de manera arbitraria necesitaría promedio 25 Bytes por punto (considerar que existen puntos de hasta 63 Bytes [The American Society for Photogrammetry & Remote Sensing (ASPRS), 2019]).

Luego, estos diferentes niveles de calidad de datos son seleccionados de forma exclusiva o diferido dependiendo de su locación geográfica. Para ello, se definen 4 escenarios que utilizan estos niveles de calidad, junto con un plan de adquisición de datos, los que surgen desde los beneficios que se busque tener, así como los costos que se ajusten al presupuesto. Estos son especificados en la Tabla 2. La diferencia entre los escenarios 2 y 4, que poseen los mismos niveles de calidad, radica en las zonas geográficas en las que se tiene cada uno de los QL y la cantidad que se tiene de cada uno de ellos, buscando beneficios para las entidades mencionadas. El escenario 4 presenta la mejor relación costo/beneficio, con una inversión anual de 160.6M USD, y beneficios anuales de 780.2M USD, satisfaciendo 66.1 % de los posibles

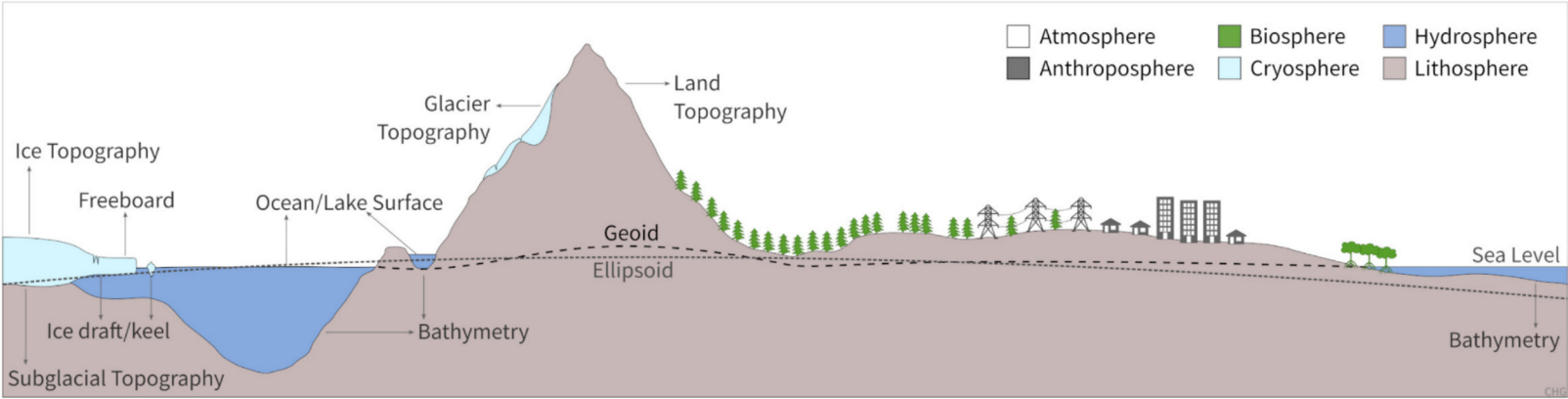


Figura 4: Comparación entre un geoide y un elipsoide.
Fuente: [Guth et al., 2021]

Niveles de calidad	Fuente	Puntos por m^2 horizontal	Tamaño de archivo LAS esperado
QL1	LiDAR	8	2146.6 TB
QL2	LiDAR	2	536.6 TB
QL3	LiDAR	1-0.25	268.3 - 67.1 TB
QL4	Imágenes	0.04	10.7 TB
QL5	IFSAR	0.04	10.7 TB

Tabla 1: Tamaño de archivos estimados para U.S.A. por nivel de calidad.
Fuente: [Dewberry, 2012]

beneficios. Esto refleja el tamaño de los datos que requieren ser almacenados, aumentado por la redundancia que se necesita tener.

En el reporte de Dewberry se detallan etapas por las que pasan los datos, desde la adquisición hasta la provisión. Primeramente en la adquisición de datos, consiste en el estandarizado de todos los datos a formato LAS, edición de las georreferencias y en el sistema de coordenadas, revisión automática de la data, y documentación automática. Luego se disponibiliza la data para ser explorada por el usuario, descargada, modificada (cambios de formato y afines) previo a su descarga, y presentada en servicios web.

Respecto al procesamiento de los datos, en el reporte se estiman mediante un CPU Op-teron ó Xeon del año 2007, por lo que no son proyecciones acordes a la rapidez de procesamiento disponible hoy en día. De acuerdo a lo establecido en el informe, se detalla que tomaría 2779331 horas la etapa de procesamiento para una superficie de 3.79 mi^2 , (9816054 km^2), lo que equivale a aproximadamente 13 veces la superficie de Chile (756950 km^2) [Instituto Nacional de Estadísticas de Chile, 2006]]. Si realizamos un cálculo rápido vemos que nos tomaría alrededor de 213795 horas realizar todos estos cálculos (esto será analizado a profundidad en la Sección 3.2. Como se menciona anteriormente, este número decrece en función de las mejoras de hardware que se realicen.

Actualmente el proyecto Chile 3D tiene relaciones de colaboración con los encargados de la plataforma del Centro de Descargas del Organismo Autónomo Centro Nacional de Informa-

Escenario	Plazo de adquisición de datos	Niveles de calidad	Entidad beneficiada
Escenario 1	25 años	QL3 U.S.A. Continental QL5 Alaska	Ninguna en específico
Escenario 2	8 años	QL1/2/3 U.S.A. Continental QL5 Alaska	Federal
Escenario 3	8 años	QL2 U.S.A. Continental QL5 Alaska	Ninguna en específico
Escenario 4	8 años	QL1/2/3 U.S.A. Continental QL5 Alaska	Federal, Estatal, y No Gubernamental

Tabla 2: Escenarios de planificación por cada QL de la NEEA.

ción Geográfica de España (CNIG²), proyecto que maneja datos de todo el territorio español en diversos formatos, donde destacan datos LiDAR y ortofotos. El proyecto considera a un ente tercero que realiza los sobrevuelos y entrega la información lista para ser almacenada. No obstante, se obtienen productos derivados de estos modelos, los que son procesados de manera previa. De esta forma se disminuyen los tiempos de respuesta, y aumenta el tamaño de almacenamiento necesario. Este sistema trabaja con una interfaz gráfica donde se selecciona el polígono de terreno que quiere descargarse. Posteriormente, el sistema entrega aquellos archivos que contienen dichas coordenadas para ser descargados. No se realiza ningún procesamiento, sino que se envían los archivos tal como son almacenados.

²CNIG, centrodedescargas.cnig.es

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

Inicialmente, se tiene una plataforma donde se destacan 4 grandes entes: La interfaz gráfica (GUI) con la que interactúa el usuario, la API que comunica la GUI y realiza las solicitudes, el sistema de almacenamiento, y el sistema de manejo de los datos. Esta memoria se enfocará en el sistema de almacenamiento y el manejo de datos del proyecto Chile 3D, mientras que otras memorias trabajarán en las otras aristas. Inicialmente realizaremos una descripción general del modelo completo, posteriormente se detallará en lo específico de esta memoria.

El prototipo generado permitirá alojar archivos de nubes de puntos LAS y ráster GeoTIFF (en este informe se utilizarán los conceptos TIFF y GeoTiff de manera indistinta, siempre refiriéndose a GeoTIFF en el marco de imágenes TIFF georreferenciadas), estos archivos podrán ser accedidos a través de una interfaz de usuario gráfica en una página web, la que hará uso de una API para la puesta a disposición de los archivos.

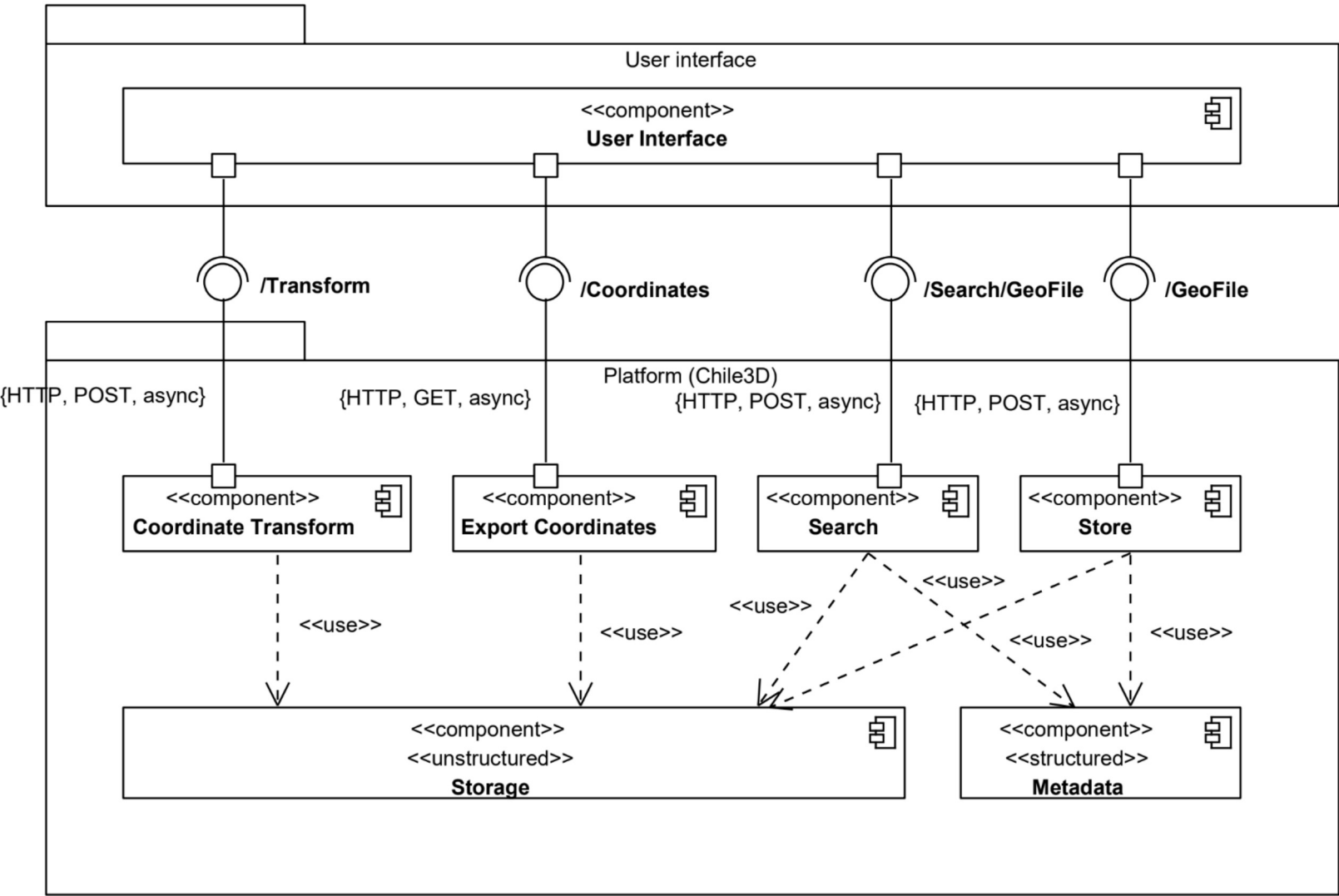


Figura 5: Modelo de dominio: vista funcional inicial Chile 3D.
Fuente: Proyecto Chile 3D.

3.1. Descripción general de la plataforma Chile 3D

En la Figura 5 vemos la vista funcional inicial del proyecto Chile 3D. La plataforma proveerá funcionalidad para el procesamiento de archivos geoespaciales descritos anteriormente. La provisión de la funcionalidad será por medio de un conjunto de APIs. La GUI queda desacoplada de la plataforma de tal manera que clientes externos puedan implementar sus propias interfaces haciendo uso de APIs de uso abierto. En datos geoespaciales, es común encontrar una fuente de datos no estructurada y una fuente de datos estructurada. Lo primero constituye los datos *raw* no estructurados que contienen la información específica del terreno, mientras que en los datos estructurados encontramos la *metadata*, que incluye referencias que nos permiten hacer uso de los datos raw (sistema de coordenadas, georreferencias, entre otros).

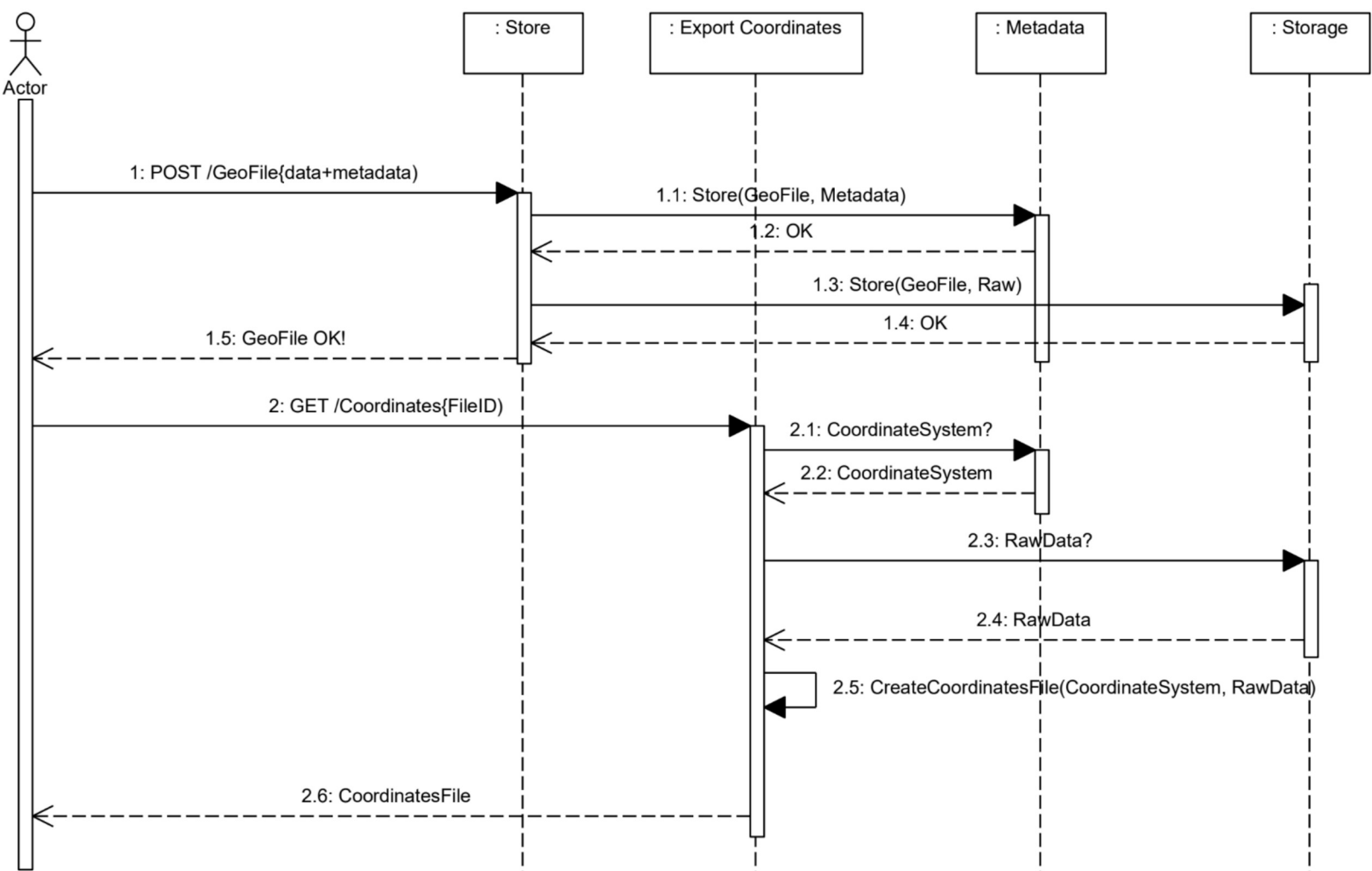


Figura 6: Caso de ejemplo de uso.
Fuente: Proyecto Chile 3D.

En la vista funcional inicial destacan cuatro componentes: (1) Transformación de coordenadas, (2) exportación de coordenadas, (3) búsqueda de modelo/s en particular, (4) almacenamiento (CRUD). Un ejemplo de uso se presenta en a figura 6, donde un usuario mediante envía al sistema un modelo DEM de datos geoespaciales. La plataforma almacenará los metadatos y datos raw de manera separada. Otro analista puede posteriormente solicitar un archivo con las coordenadas y mediante el componente específico *Export Coordinates*, que tomará los metadatos y raw para crear el archivo con las coordenadas y finalmente hacer entrega al analista que las solicitó.

3.2. Adquisición y almacenamiento de datos

El foco de este proyecto está en las nubes de puntos capturados por un sensor LiDAR y en formato LAS. No obstante, también es de utilidad para algunas aplicaciones el uso de ráster. Mientras que una nube de puntos presenta los datos de forma detallada, modelando fielmente las características tanto del terreno como de los objetos y vegetación situados sobre el terreno, un ráster nos permite tener una visión más continua de la superficie, útil para áreas extensas. Un ráster se caracteriza además por su almacenamiento de datos en una cuadrícula uniforme de datos, donde la densidad de puntos se mantiene uniforme, mientras que en una nube de puntos podría no ocurrir de esta forma (por ejemplo quedando esparcidos los puntos de la nube en un espectro de altura).

El objetivo de Chile 3D es el de contar con una base de datos centralizada donde se puedan encontrar datos altimétricos de todo Chile. Esto se realiza mediante la captura de datos a través de sobrevuelos con sensores LiDAR. La adquisición de un gran volumen de datos, análogo al trabajo realizado en Estados Unidos (detallado en la Sección 1.2), conlleva la realización de un plan de adquisición de datos, por las siguientes razones:

- Diferente resolución dependiendo de topografía, es decir, mayor resolución necesaria en zonas costeras, montañosas: las zonas mencionadas tienen mayor interés por ejemplo de parte de entidades de prevención de desastres, por lo que requieren mayor detalle en su modelado.
- Mayor resolución en sectores con mayor densidad poblacional: estos sectores altamente dinámicos, especialmente en sus zonas límite, es decir, zonas de crecimiento urbano. Contar con estos datos de alta calidad permitirá tener información más fidedigna del crecimiento urbano, cambios en la infraestructura, edificaciones ilegales, vegetación, entre otros.
- Zonas de interés de la industria minera, de obras civiles, e agropecuaria: estos sectores productivos requieren del estudio de suelos para realizar sus actividades, por lo que contar con estos datos unificados significaría gran aporte.

Debido a la gran cantidad de factores presentados que afectan en la decisión de las zonas que requieren datos de alta resolución primero, se propone seguir un listado regional, de acuerdo al orden de densidad poblacional. Con los datos adquiridos en el Censo 2017 en Chile [Instituto Nacional de Estadísticas de Chile, 2017], podemos construir la una tabla en orden decreciente de acuerdo a la densidad poblacional (ver Tabla 3).

No obstante la propuesta anterior, esta sectorización debe ser acompañada por una propuesta de quien realizará los vuelos, debido a la superficie de cada celda capturada, cantidad de celdas capturadas por vuelo, entre otros.

Región	km^2	Habitantes	Habitantes/ km^2
Metropolitana	15403.2	7112808	461.8
Valparaíso	16396.1	1815902	110.8
Biobío	23890.2	1556805	65.2
O'Higgins	16387	914555	55.8
Ñuble	13178.5	480609	36.5
Maule	30296.1	1044950	34.5
La Araucanía	31842.3	957224	30.1
Los Ríos	18429.5	384837	20.9
Coquimbo	40879.9	757586	18.5
Los Lagos	48583.6	828708	17.1
Arica y Parinacota	16873.3	226068	13.4
Tarapacá	42225.8	330558	7.8
Antofagasta	126049.1	607534	4.8
Atacama	75176.2	286168	3.8
Magallanes (excluyendo Antártica)	132297.2	166533	1.3
Aysén	108494.4	103158	1.0

Tabla 3: Densidad poblacional por región

3.2.1. Almacenamiento de datos

En el trabajo realizado por la USGS (como por ejemplo en la Figura 7), podemos ver una celda cuadrada que cubre $4\ km^2$ de superficie. Se desconoce si cada vuelo captura celdas de este tamaño o si es el resultado de un post procesado y es solo una parte de la grilla obtenida en cada vuelo único.

Para el almacenado de data altimétrica capturada por sensores LiDAR, se seleccionó LAS por las siguientes razones, las mismas por las que estas son el centro del proyecto Chile 3D:

- Formato diseñado específicamente para nubes de puntos capturados con LiDAR, de ahí el acrónimo del inglés LiDAR Aerial Survey, permitiendo la inclusión de datos de elevación, y soportando metadata para incorporar datos como: parámetros del sensor, detalles de la adquisición, y la referencia geoespacial.
- LAS soporta además de los datos altimétricos, otras capas de datos e.g. valores de intensidad, clasificación, entre otros. Un uso relevante para la clasificación de puntos es útil para diferenciar puntos que representan suelo, vegetación, edificaciones, cableado, ruido, etc.
- Capacidad para ser comprimido con LASzip, preservando todos los datos. Esto facilita su almacenaje y transferencia, disminuyendo en promedio 6 veces el tamaño de archivo, más detalle al respecto a continuación.

Si se fuera a cubrir la superficie de Chile de 756102 km^2 con celdas de 4 km^2 , se tendría que generar una grilla con 189025.6 celdas. Considerando que cada punto de la nube de puntos tiene un peso de 25 Bytes (dependiendo de qué datos extra tendrá) [The American Society for Photogrammetry & Remote Sensing (ASPRS), 2019], y si consideramos los mismos QL propuestos en el reporte de Dewberry, obtenemos columna “Tamaño de archivo LAS” de la tabla 4 con el espacio de almacenamiento necesario para cada nivel de calidad en total (para toda la superficie del país).

Niveles de calidad	Fuente de datos	Puntos por m^2	Tamaño de archivo LAS	Tamaño de archivo TIFF	Tamaño de archivo ECW
QL1	LiDAR	8	165.2 TB	22.0 TB	66.1 TB
QL2	LiDAR	2	82.6 TB	5.5 TB	11.0 TB
QL3	LiDAR	1-0.25	20.7 - 5.16 TB	2.8 - 0.7 TB	5.5-2.75 TB
QL4	Imágenes	0.04	0.83 TB	0.1 TB	0.2 TB
QL5	IFSAR	0.04	0.83 TB	0.1 TB	0.2 TB

Tabla 4: Tamaño de archivo total estimado para Chile por nivel de calidad.

No obstante lo anterior, existe un formato llamado LAZ [Isenburg, 2013] que utiliza el algoritmo zip para comprimir sin pérdida archivos LAS, pudiendo ser utilizados en su forma comprimida por programas GIS. Este formato nos permite diezmar el tamaño de almacenamiento requerido, resultando en una alternativa conveniente para disminuir el uso de almacenamiento, tiempo de procesamiento, memoria, banda ancha, entre otros, el que será analizado a detalle más adelante.

Junto con nubes de puntos, también se almacenarán los datos altimétricos en ráster TIFF, pues si bien el usuario puede procesarlas localmente, idealmente el sistema debe poder entregar formatos de alta demanda de manera eficiente. Los archivos LAS o LAZ en su defecto serán procesados una vez son subidos al sistema. Evaluar los requerimientos espaciales se simplifica al tener que considerar solo dos dimensiones espaciales. Sumado a lo anterior, estos archivos TIFF también podrán ser comprimidos para reducir su tamaño. Cada pixel será una celda de la grilla, donde su peso estará dado por la resolución de los datos de altura, es decir, si está es almacenada en 8, 16, 32, 64 bit u otro. Por lo observado, se suele trabajar con 32 bits, mas se han encontrado archivos a 8 bit y 64 bit (al transformar desde una nube de puntos sin considerar la resolución de entrada por ejemplo).

Realizando el ejercicio de proyectar el espacio requerido por archivos GeoTIFF para representar la superficie de Chile, se utilizó un píxel por cada punto de la nube de puntos. Por lo tanto, 8 puntos por metro cuadrado significarán 8 píxeles por metro cuadrado en la imagen TIFF. Esto resulta en una grilla uniforme, por lo que no representaría fielmente la información de la nube de puntos. Sin embargo sirve para acercarse a un valor cercano al real para esta aplicación.

Se puede visualizar el resultado presentado en la columna “Tamaño TIFF” de la tabla 4. A posterior se indagará en términos de compresión y reducción de resolución, con los diferentes

algoritmos de compresión en la sección 3.5.

También se contempla en un futuro incluir ortofotos en formato Enhanced Compression Wavelet (ECW). ECW es un formato desarrollado por Earth Resource Mapping que permite la compresión en ondículas. Este algoritmo de compresión con pérdida logra ratios de compresión desde 10:1 a 50:1, junto con buen desempeño en términos de procesamiento de compresión y descompresión tanto temporales como espaciales. Mediante la transformada discreta de ondículas (DWT de sus siglas en inglés: Discrete Wavelet Transform), técnica utilizada para analizar señales descomponiéndolas en diferentes componentes de frecuencia, logra a través de una serie de operaciones de filtrado y submuestreo, una representación de la señal original [Triglav, 2000]. No se considera este formato para el presente informe, no obstante, se realizó el ejercicio de agregarlo en los espacios necesarios.

Calculando de forma análoga a las imágenes TIFF el espacio requerido por ortofotos ECW en base a lo observado, una imagen RGB necesita 32 bits por canal, es decir, 12 Bytes por pixel en total. Con estos datos se obtienen los resultados expuestos en la tabla en cuestión. No obstante, en la realidad los valores serán reducidos por la compresión lossy que se expuso anteriormente, con ratios de compresión de 10:1 a 50:1. Esta compresión es liviana de calcular, por lo que podrá ser una buena alternativa para reducir el tamaño de archivos enviados desde el servidor al cliente.

Como complemento, algunos criterios evaluados para seleccionar TIFF como formato de raster de capturas de datos altimétricos son los siguientes: TIFF permite compresión lossless (con métodos que serán analizados a detalle más adelante), preservando valores de elevación. Capacidad de georeferencia de archivos, permitiendo ubicar en el mapa y complementar la información con otros archivos en algún software GIS por ejemplo. Soporte multibanda, lo que posibilita la opción de agregar capas adicionales con información complementaria a la información altimétrica. Integridad de los datos, al contar con validación de checksum, previniendo pérdida de información o corrupción de los archivos durante su almacenaje, transferencia, o procesado.

Los valores presentados son estimados, es decir, dependen de las condiciones reales de terreno, de formato de la entidad adquisidora, entre otros. Además, los siguientes factores prácticos son de alta relevancia y su estimación a priori se dificulta:

Almacenamiento inicial e intermedio: Durante el desarrollo del proyecto, se contará con un servidor de almacenamiento con tamaño necesario para una zona continua, arbitraria, de la que ya existan datos. A modo de prueba de concepto, podrán ser usados datos de otras zonas geográficas disjuntas a Chile.

Plan de adquisición de datos: Si se comienza con las localidades de mayor densidad poblacional, se introduce la variable de tiempo. Es decir, cuántos km^2 o MB de datos son capturados en un ΔT . En base a esta información podremos saber el mecanismo en el que los datos son almacenados en la base de datos tanto en su transferencia, como en su almacenaje a largo plazo. En cuanto al almacenaje, este se hará de forma progresiva, a medida que se incorpo-

ren nuevos datos, se aumentará el escalón de almacenamiento en un servidor escalable. Se deberá seleccionar un servicio de almacenamiento que cuente con redundancia, no obstante, posiblemente se sugiere contar con redundancia de manera local en territorio nacional.

Histórico: Una vez se tengan capturas de la totalidad del territorio nacional, se necesitarán nuevos datos en alguna ventana de tiempo de acuerdo al presupuesto y a las necesidades técnicas. Esto conlleva por lo tanto, además de mantener la base de datos con los datos de la última adquisición, mantener todo el historial de datos. Este crecimiento será lineal de acuerdo al periodo de adquisiciones, las que posiblemente se centrarán en las zonas más densas o aquellas de mayor interés, mientras que las antónimas podrían mantener una frecuencia de actualización menor.

Formato: En un comienzo se considera contar con datos de nubes de puntos LAS obtenidos mediante sensores LiDAR, incluyendo su derivación en formato TIFF. No obstante, este proyecto no se debe cerrar a datos altimétricos en nubes de puntos, sino que debe ser escalable a otros formatos como por ejemplo datos obtenidos mediante interferometría. Esto significa que existirá la necesidad de volver a calcular las necesidades espaciotemporales, así como las de cómputo presentadas en este proyecto, a modo de incluir estas nuevas fuentes de datos. De momento el sistema contará con un indexado que hace uso de los metadatos incluidos en archivos con los formatos indicados, por lo que se necesitará tener estas características en consideración para acomodar diferentes tipos de datos geoespaciales.

Compresión: El sistema cuenta con una API que parte de las funciones que ofrece es la de compresión de archivos LAS y GeoTIFF, además de submuestreo de datos. Esto permitirá reducir tiempos de transferencia de archivos, así como reducir tamaños de descarga.

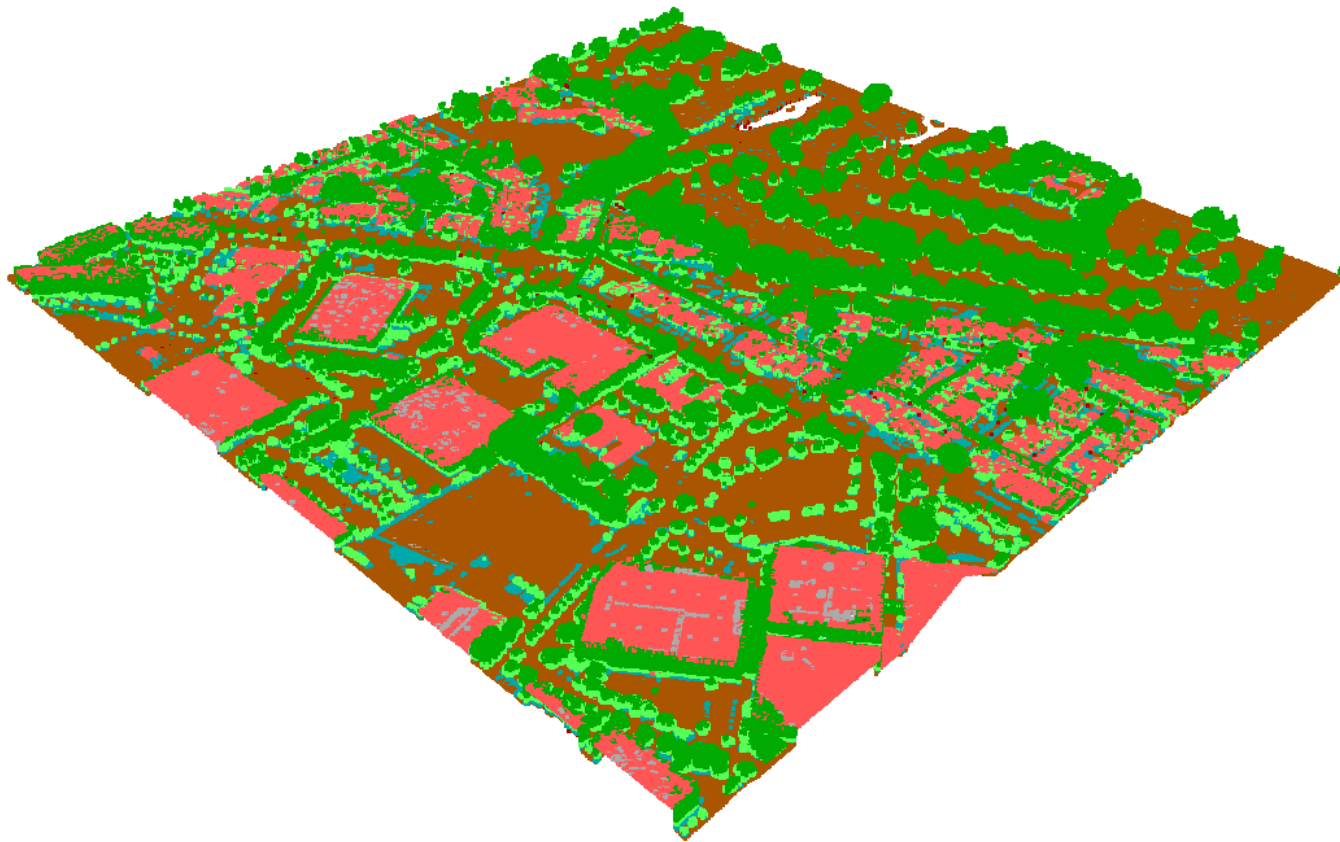


Figura 7: Captura de nube de puntos de [USGS, 2017]

3.3. Submuestreo de datos altimétricos

Por las razones descritas anteriormente, tanto para la academia como para la industria, una plataforma como Chile 3D será útil y ayudará a reducir costos de investigación. Dependiendo del uso, se requieren ciertos tipos de datos, y de acuerdo a lo señalado en entrevistas con expertos de SERNAGEOMIN, la alta densidad de los datos LIDAR dificulta su adquisición y procesamiento en algunas aplicaciones. Es por ello que se implementarán inicialmente algoritmos disponibles en bibliotecas de código abierto para el submuestreo de los datos altimétricos. Se centrará en datos tipo Ráster TIFF y Nubes de Puntos LAS a modo de probar el concepto. Vale mencionar, que la idea es que en base de datos se almacene el dato original, pues esta reducción de resolución conlleva una esperada pérdida de información.

3.3.1. Submuestreo de Nubes de Puntos

Para el submuestreo de datos altimétricos en nubes de puntos LAS, se utilizó la biblioteca Point Data Abstraction Library (PDAL) [PDAL, 2022] para Python. Mediante la misma, se fue

capaz de realizar pipelines independientes para el procesamiento de datos dentro de programas Python o bien directamente a través de una interfaz de líneas de comando en Linux (CLI). En este caso fue inicialmente implementado en scripts Python a modo de realizar pruebas de rendimiento.

Las pruebas de rendimiento se realizaron de manera arbitraria utilizando para LAZ (versión comprimida de LAS), la captura realizada en 2019 por United States Geographical Services (USGS) en 2019 con sensores Lidar [USGS, 2017] de una zona del Condado Cook en Chicago, Illinois en Estados Unidos. Una captura de una vista isométrica de dicho archivo se presenta en la Figura 7.

Los archivos LAS son compuestos por nubes de puntos en 3 dimensiones, razón por la que al realizar algún tipo de submuestreo se debe tener consideración en el algoritmo que se utilizará. Por la naturaleza de los datos, se seleccionaron 3 algoritmos disponibles en la biblioteca PDAL que utilizan la idea del Voxel para este diezmado. En la figura 8 se aprecia un esquema de ubicación de voxels en nubes de puntos para su separación y posterior procesamiento por separado.

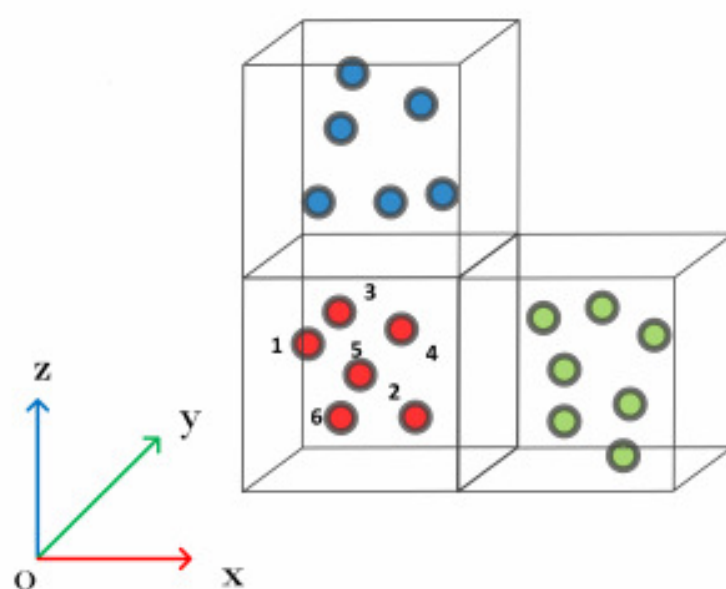


Figura 8: Muestreo de nubes de puntos usando celdas voxel [Wang *et al.*, 2018].

Una breve revisión del funcionamiento de los algoritmos probados [PDAL, 2022]: *sample*, *voxelcenternearestneighbor*, *voxelcentroidnearestneighbor*, y *voxeldownsize*. *Sample* utiliza muestreo de Poisson para la selección de puntos dentro de los voxel definidos anteriormente, de acuerdo al tamaño de celda solicitado, esto se puede visualizar en la Figura 9. El muestreo de Poisson se basa en asignarle a cada punto la probabilidad de mantenerse (en el archivo de salida) de acuerdo a su densidad. La función de densidad se construye basándose en la densidad local de cada punto de la nube de puntos, de acuerdo al tamaño de celda. A mayor densidad, se tendrá mayor probabilidad de ser seleccionados. Finalmente, se realiza una selección aleatoria considerando las probabilidades calculadas previamente. De esta forma, se logran mayores detalles en zonas con densidad de puntos superiores, intentando mantener características importantes del modelo original y reduciendo su resolución [Smothermath, 2017].

Por otro lado, *Voxelcenternearestneighbor* también divide la nube de puntos en voxels, pero luego se selecciona el punto más cercano al centro del voxel. Por otro lado, *voxelcentroidnearestneighbor* trabaja con el centroide en vez del centro. Este último método no conserva la data dimensional, solo las coordenadas XYZ . Por último *voxeldownsize* selecciona por defecto el primer punto que se encuentra en el voxel. Los tamaños de celda son en las dimensiones de XYZ , es decir, si se trabaja en metros, entonces un tamaño de celda 10,0 significa que tendremos voxel de $10 \times 10 \times 10m$ o $1000m^3$.

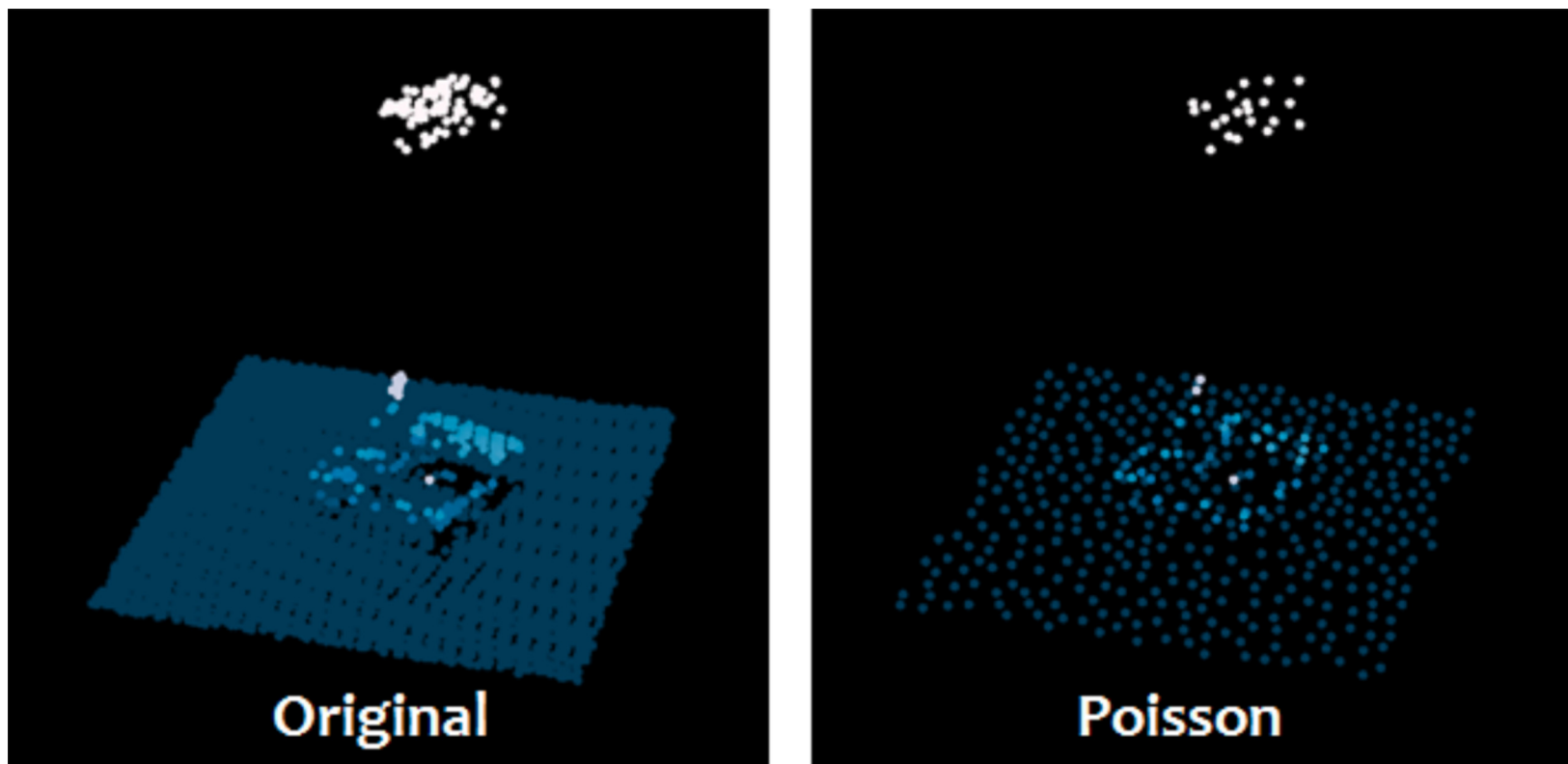


Figura 9: Efecto del método Sample que utiliza el submuestreo de Poisson [Smathermather, 2017].

Cada método de compresión depende de la aplicación específica para la que se requiera la nube de puntos determinada por el experto. Es por ello que se midió el desempeño temporal y espacial de cada uno de los métodos a modo de conocer si existe otro criterio a considerar al momento de su uso. Se realizaron 19 iteraciones, cambiando el tamaño de la celda desde 1 metro hasta 19 metros. Los resultados se visualizan en la Figura 10, donde podemos observar que el tiempo de cómputo no depende en la práctica del método utilizado. Además en la Figura 11, el tamaño del archivo de salida se mantiene constante entre métodos debido a que se utiliza la misma cantidad de vóxels, por lo que la cantidad de datos es similar.

Finalmente, en la Figura 12 podemos apreciar una captura de la misma nube de puntos con un tamaño de celda de 19 pies. A pesar de haber disminuido en gran medida la resolución de la imagen aún son discernibles características visualmente a la distancia a la que se encuentra el observador. Lógicamente si el observador se acerca, comenzaremos a notar falta de datos.

Vale agregar que se realizaron pruebas utilizando CUDA de NVidia, con la idea de paralelizar este proceso. No se tuvo un resultado temporal favorable al menos en la implementación que se hizo, por lo tanto se dejó de lado de momento, pero no se descarta ahondar en esta arista en el futuro.

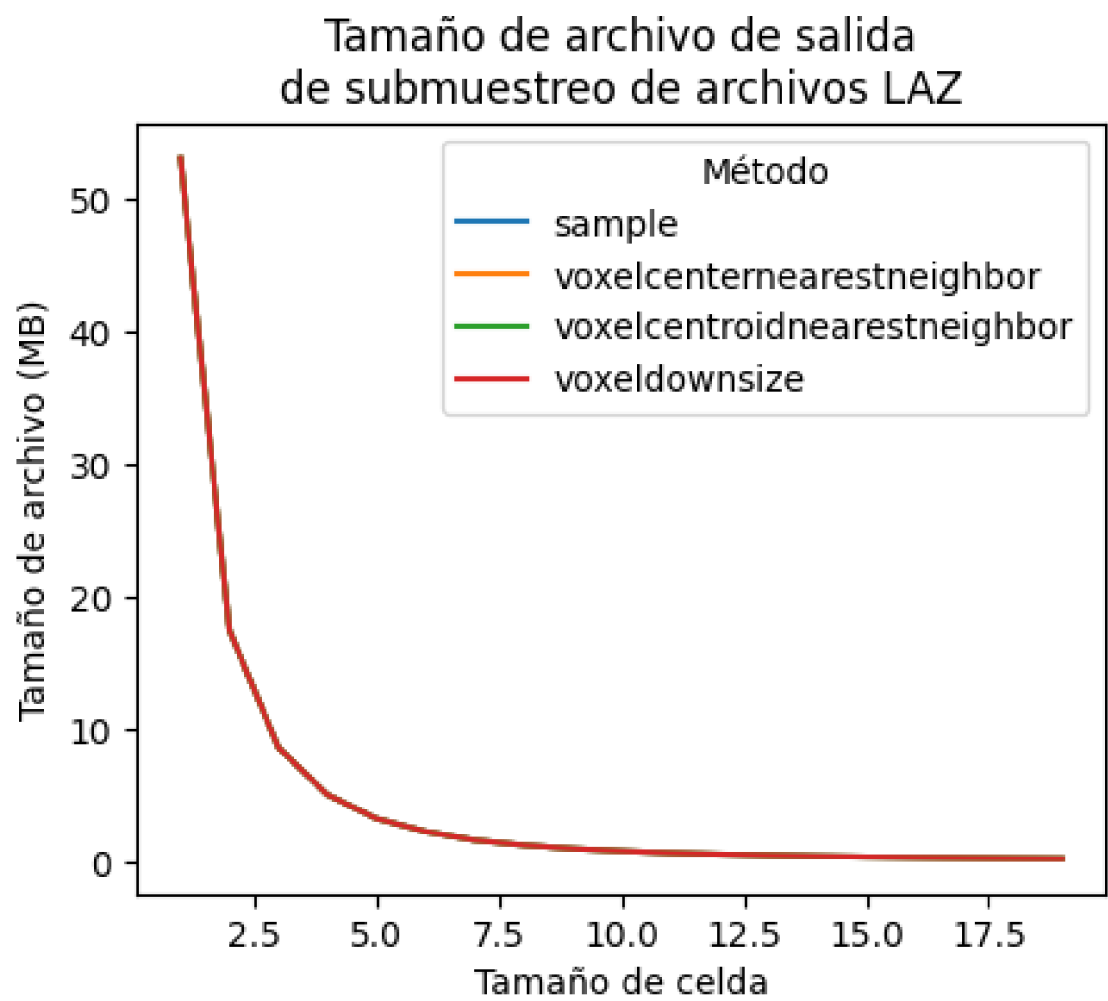


Figura 10: Tiempo de cómputo

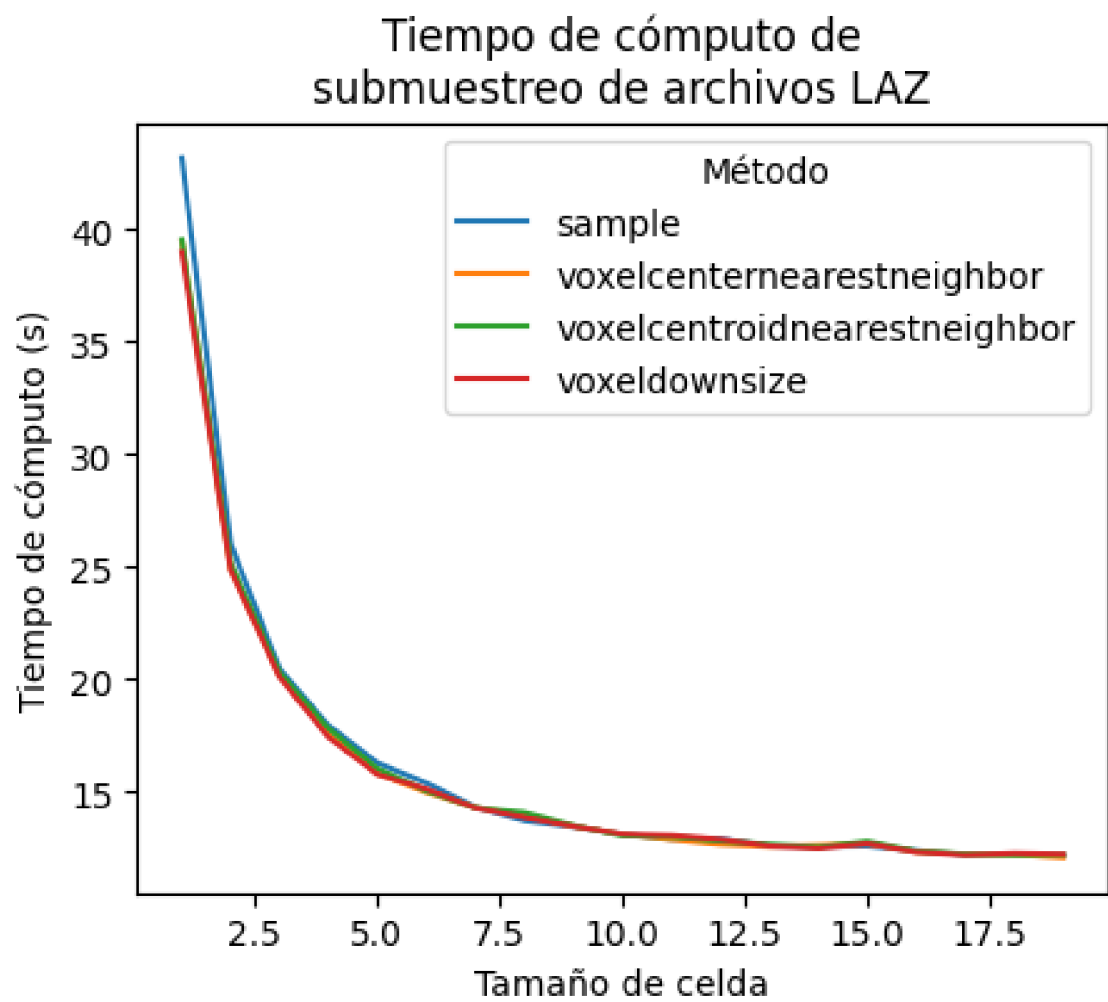


Figura 11: Tamaño de archivo de salida

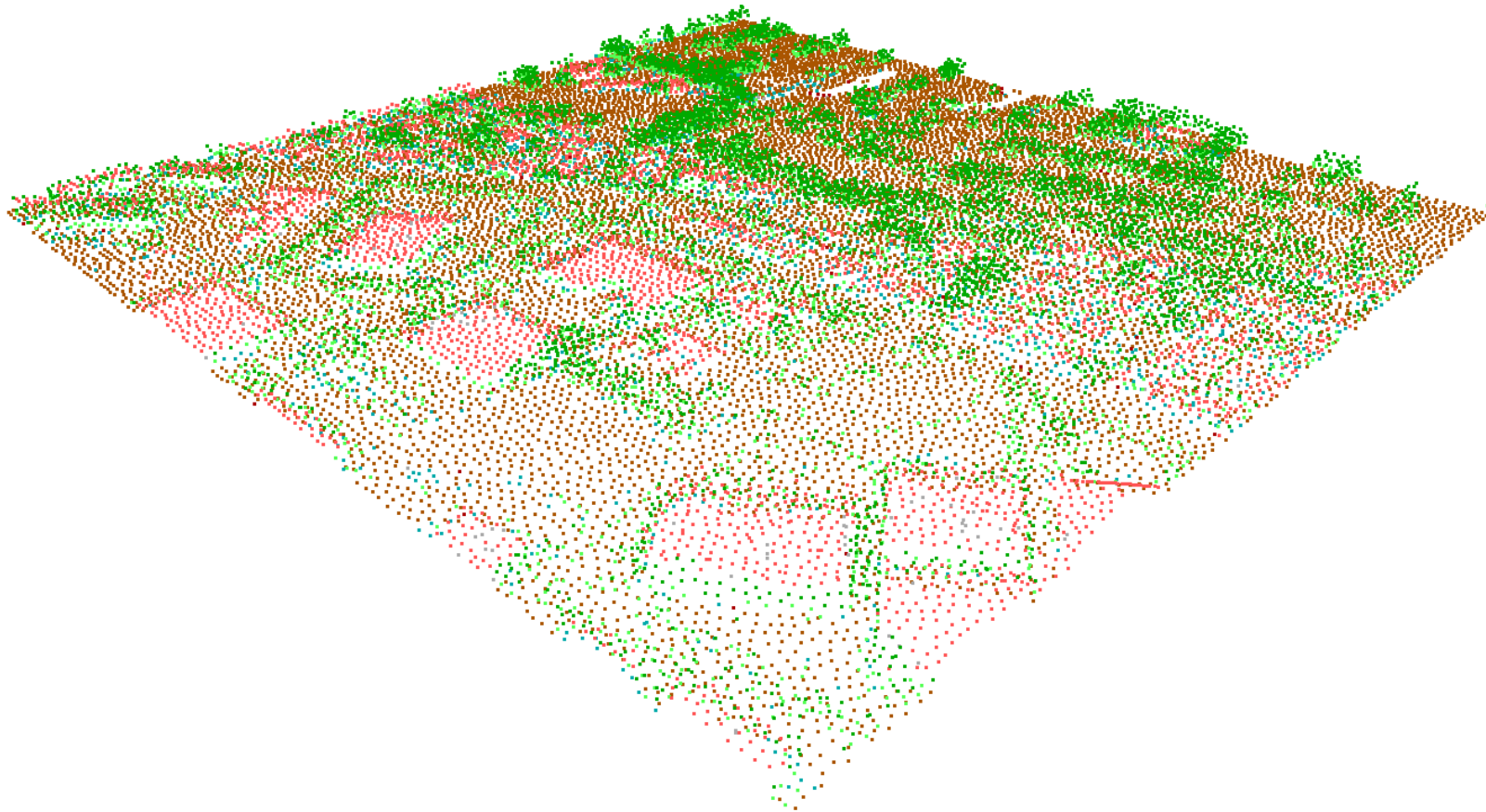


Figura 12: LAZ de Cook County con un tamaño de celda de 19 pies

3.3.2. Submuestreo de Ráster

Similar al caso anterior, se requiere realizar una reducción de resolución de archivos de tipo ráster en formato TIFF. Para efectos demostrativos se utilizará un archivo de la ciudad de Copiapó, Chile [Ministerio de Bienes Nacionales, 2017], imagen capturada por un Dron en la localidad de Paipote. En la Figura 13 se aprecia el ráster posicionado sobre un plano de OpenStreetMap.

En este sentido, al tratarse de una imagen plana, podemos observar comportamientos lineales, como el tiempo de cómputo en función del escalado y el tamaño de archivo de salida que se aproxima a una función raíz (ver Figuras 14 y 15).



Figura 13: Ráster de la localidad de Paipote en Copiapó [Ministerio de Bienes Nacionales, 2017], sobrepuesto a un mapa de Openstreetmap.

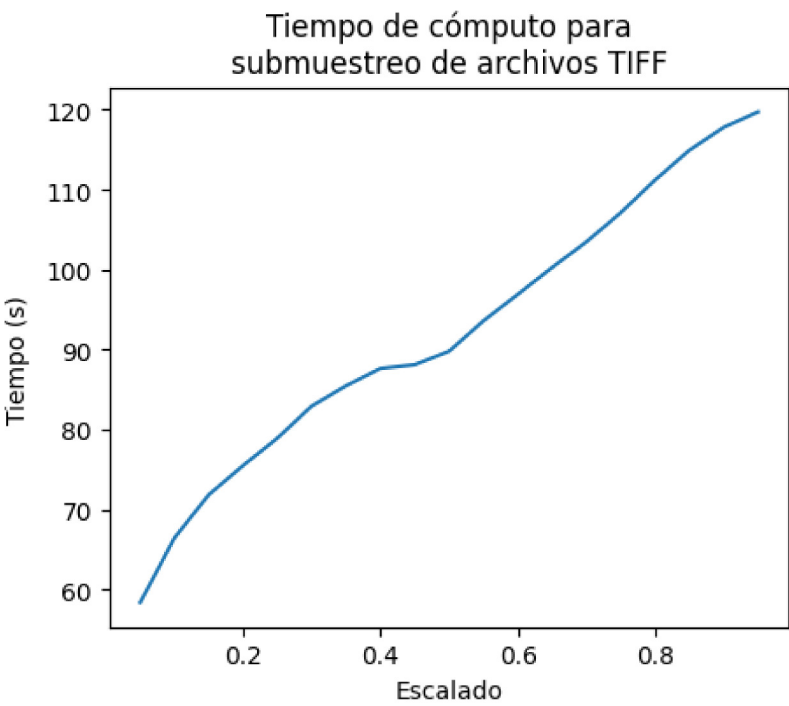


Figura 14: Tiempo de cómputo

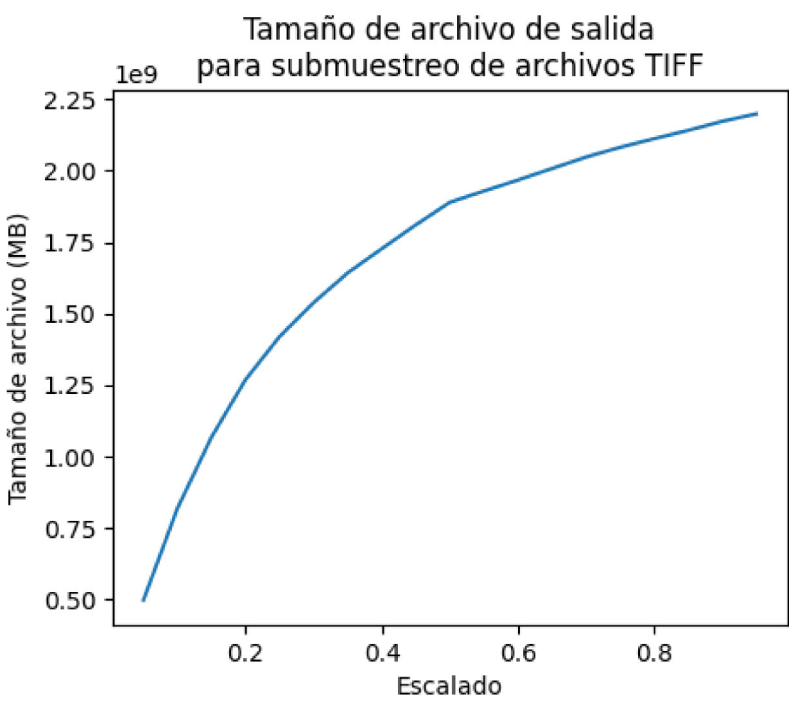


Figura 15: Tamaño de archivo de salida

3.4. Conversión de Nubes de Puntos LAS a Ráster TIFF

La conversión de nubes de puntos LAS a Ráster TIFF también se hará mediante la biblioteca PDAL. En esta transformación de una nube de puntos tridimensional a una imagen bidimensional debemos conservar la mayor cantidad posible de información dentro de las restricciones de resolución que se requieran. Es por ello, que nuevamente se debe evaluar cuales son los parámetros de tamaño de celda y radio de búsqueda.

El tamaño de celda serán las dimensiones laterales de cada uno de los vóxel cúbicos de la imagen, que se transformarán en los píxeles en el raster de salida. Además del tamaño de la celda, se puede definir un radio de búsqueda al rededor de cada punto por sobre el cuál los diferentes algoritmos buscarán datos. Se realizó el mismo ejercicio anterior, pero esta vez variando además el radio de búsqueda dentro de la vecindad de la función descrita.

En el mapa de calor de la Figura 16 podemos identificar que a mayor radio, más se demora el proceso debido a tener que buscar la información de cada punto de la imagen de salida en en más puntos de la preimagen de entrada. Por otro lado a medida que la resolución disminuye (a mayor tamaño de celda, menor resolución), el tiempo disminuye de igual manera.

En las Figuras 17, 19, 18, y 20 podemos apreciar los diferentes resultados al usar diferentes parámetros de tamaño de celda y radio.

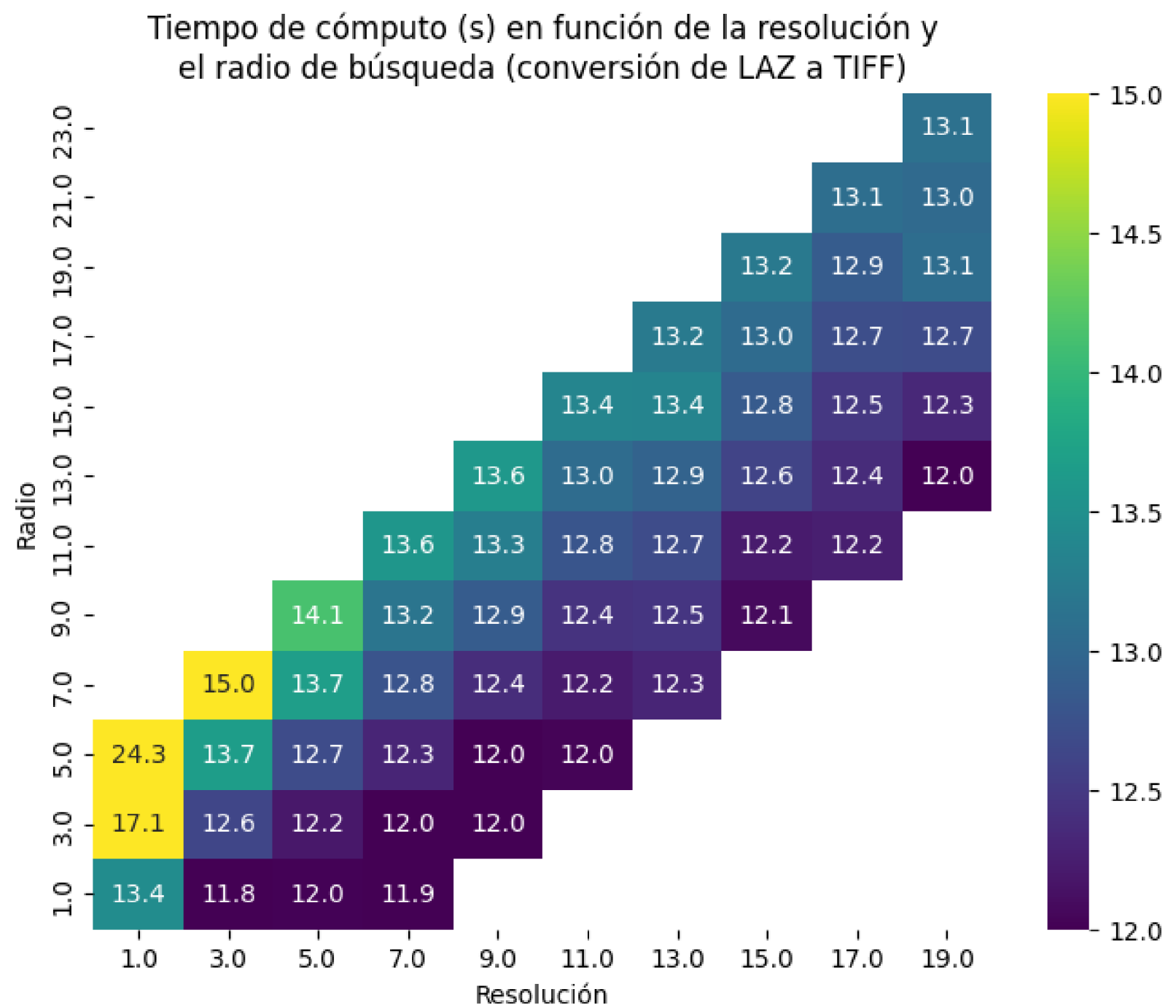


Figura 16: Tiempo de cómputo en función de la resolución (tamaño de la celda) y el radio de búsqueda.

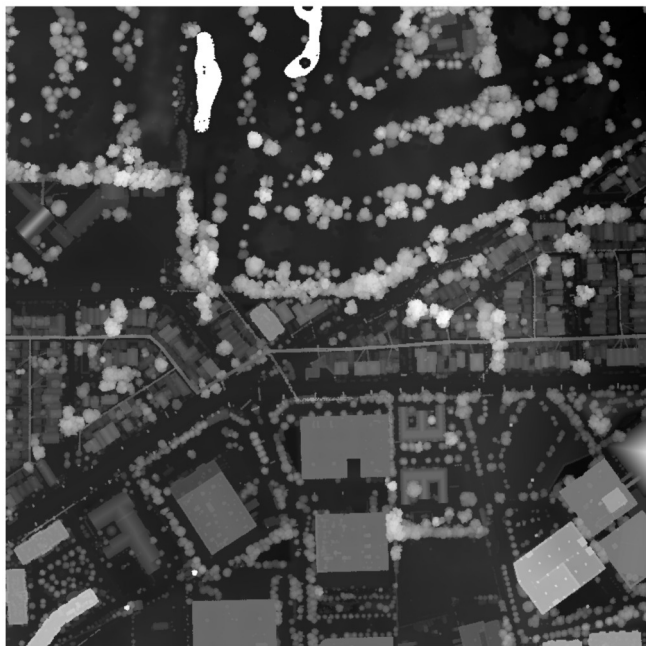


Figura 17: Celda: 1, Radio: 3

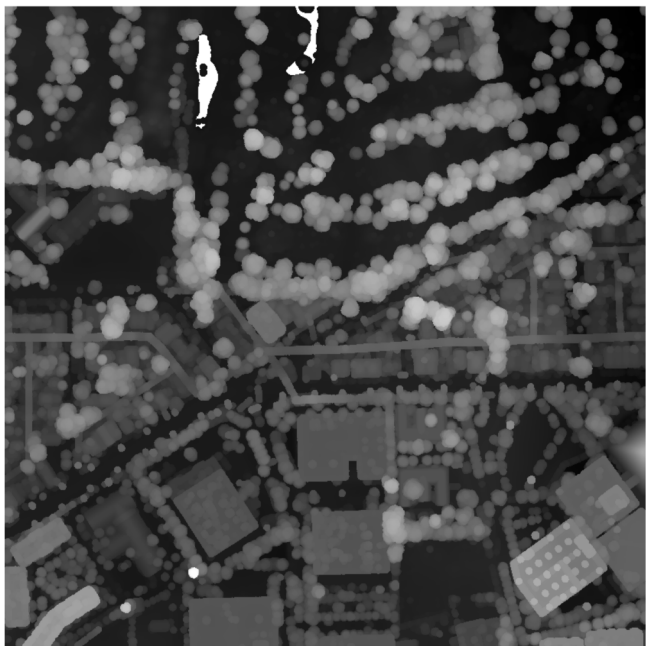


Figura 19: Celda: 1, Radio: 13

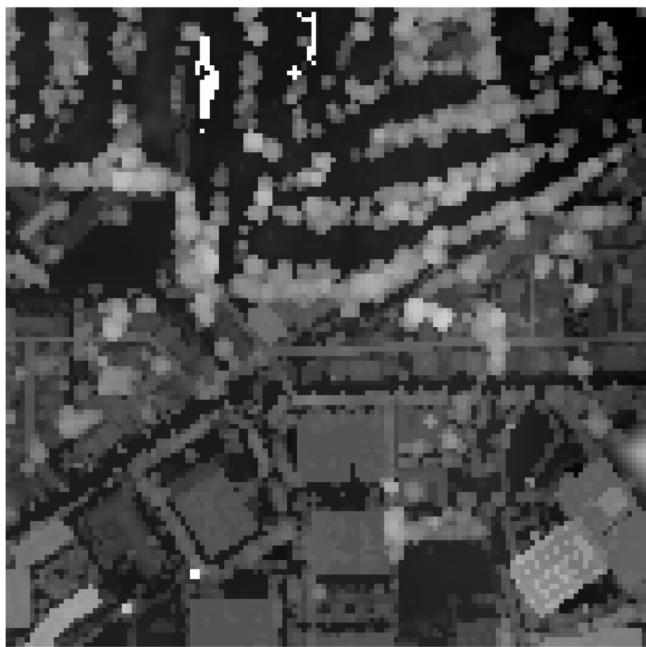


Figura 18: Celda: 19, Radio: 13

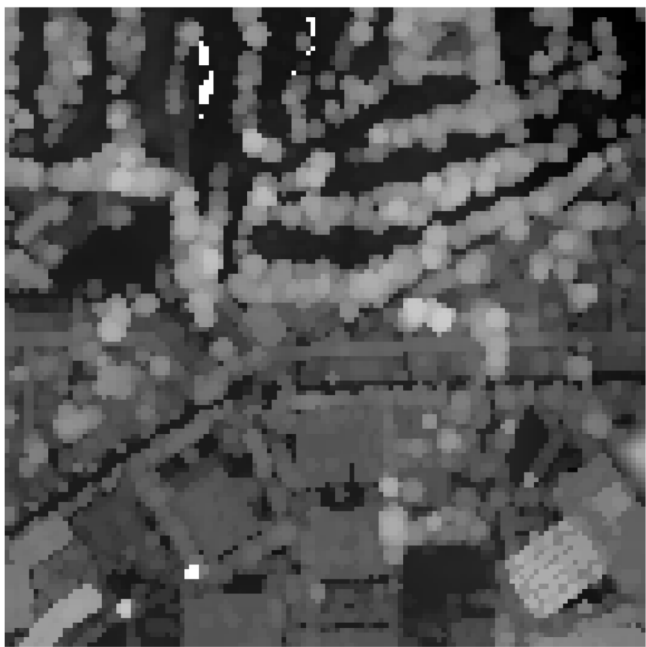


Figura 20: Celda: 19, Radio: 23

3.5. Compresión de datos altimétricos

A modo de reducir el espacio en disco necesario para almacenar datos altimétricos, ya sea en formato ráster o como nube de puntos, se realizó la prueba de diferentes algoritmos de compresión. Se probaron tanto algoritmos con pérdida (en adelante lossy), como algoritmos sin pérdida (en adelante lossless). No obstante, como estos datos son representación de características topográficas y similares, no podemos tener pérdidas de información como por ejemplo en la compresión de imágenes con fines artísticos. Se consideraron estos algoritmos con objetivos comparativos y se deja la opción para ser usados si el usuario lo requiere, pero se invita a no ser implementados debido a las razones mencionadas.

3.5.1. Compresión de Nubes de Puntos LAS

Las nubes de puntos presentan un gran desafío de compresión principalmente por dos razones: el gran volumen de puntos con pequeñas variaciones, la heterogeneidad de datos almacenados en cada punto altimétrico. Cada punto de un archivo LAS puede pesar hasta 67 bytes, dependiendo de la cantidad de información almacenada. Si extrapolamos esto, podemos encontrar que un metro cuadrado en QL1 (8 ppm) podría pesar 536 Bytes o bien 1.34 GB por cada 4 km^2 . Es posible comprimir estos archivos con algoritmos de uso general, como por ejemplo WinZip, WinRar, entre otros. No obstante, ya que su uso no está enfocado en archivos altimétricos, no obtienen resultados favorables en cuanto a su ratio de compresión. A raíz de lo anterior, surge el algoritmo LASzip [Isenburg, 2013] basado en ZIP pero con enfoque específico en archivos LAS.

En su publicación, Isenburg comenta que un problema con algoritmos genéricos es que no permiten el uso de los archivos geospaciales de forma comprimida, por lo que deben ser descomprimidos para ser usados. Por el contrario, un archivo LAZ (extensión de archivo LAS comprimido con algoritmo LASzip) puede ser leído directamente con cualquier aplicación, manejándolo como si se tratara de un LAS.

Isenburg realizó la compresión de 2914 archivos totales de 12 condados del Estado de Minnesota, Estados Unidos, en los se logró reducir desde 776 GB (30 mil millones de puntos, 28.44 Bytes por punto) de origen LAS a 64 GB de archivos LAZ (28.35 Bytes por punto). Esto fue ejecutado en un computador con recursos limitados, no obstante, se compararon a detalle los resultados espaciales, no temporales (lo que se realizará más adelante en este informe). Finalmente, se logró un ratio de compresión para estos archivos específicos de 12,125 : 1.

Posterior a ello, Isenburg realiza una comparación de archivos más heterogéneos, comparando el desempeño de LASzip con otros algoritmos de compresión disponibles en el mercado. Entre 27 archivos diferentes analizados, tanto modelos altimétricos de lugares geográficos, como modelos 3D de objetos sintéticos o no geográficos, se lograron los ratios de compresión promedio resumidos en la Tabla 5.

Algoritmos de Compresión	LAS	ZIP	RAR	LCMP	SID	LAZ
Ratio de Compresión	1.0	2.0	4.5	5.9	4.1	6.4
Tamaño en MB	2.479	1.210	551	425	610	388

Tabla 5: Comparación de desempeño entre algoritmos de compresión aplicados a archivos LAS “de naturaleza típica, experimental, e inusual” [Isenburg, 2013].

Algoritmos de Compresión	LAS	ZIP	RAR	LAZ
Tiempo de Compresión en s	0	116	56	63
Tamaño	1.6 GB	630 MB	334 MB	262 MB
Ratio de Compresión	1.0	2.5	4.8	6.1

Tabla 6: Comparación temporal de desempeño entre algoritmos de compresión aplicados a archivo LAS de Nueva York, Estados Unidos [USGS, 2014].

En esta tabla se evidencia la capacidad de reducir 6.4 veces un archivo LAS con un algoritmo lossless como lo es LASzip, en comparación con el algoritmo con el desempeño más cercano “LasCompression” (LCMP). No obstante lo anterior, la gran ventaja de LAZ es que estos archivos mantienen sus características que lo hacen usable en software GIS, tal como si se tratara del archivo LAS original.

El autor comenta que el compresor LASzip se encuentra optimizado para esos casos donde los puntos se encuentran ordenados en un “orden de escaneo”, es decir, tendrá un mejor desempeño espacial dada esta condición. Otra característica es que si en un caso un avión realiza un sobrevuelo por sobre una región geográfica y posteriormente este sector se transforma en una grilla para ser almacenado en celdas más manejables, el compresor logrará un mejor ratio de compresión si se mantienen las celdas en el orden de adquisición de los datos, tal como lo realiza la herramienta `lastile` de LAStools.

Isenburg realiza solo comparaciones espaciales, por lo que se realizará una comparación temporal entre los algoritmos en la máquina ya descrita. De los algoritmos utilizados por Isenburg, se encuentran con uso gratuito LASzip, ZIP, y RAR. Es por ello que se pudo comparar el desempeño temporal de estos presentado en la Tabla 6

3.5.2. Compresión de Ráster TIFF

Al reducir la resolución de un ráster TIFF o al generar un ráster desde una nube de puntos, se puede comprimir la imagen a modo de reducir el espacio necesario en disco para su almacenamiento. Los algoritmos propuestos no realizan modificaciones sobre los datos, solo sobre la forma en la que estos son almacenados, no obstante, se incluyeron con objetivo de comparación. Cabe destacar que la compresión de un solo archivo solo nos presenta un resultado local y no es una muestra representativa para el comportamiento estos algoritmos. El archivo utilizado fue de las cercanías de Nueva York, Estados Unidos, disponibilizado por la USGS en

Fueron probados los siguientes algoritmos lossless: lzw, deflate, packbits, zstd. Y los siguientes lossy: jpeg. Se midió el desempeño considerando espacio de almacenamiento requerido y tiempo de cómputo en la máquina mencionada anteriormente. Se realizaron reducciones de resolución desde escala 0.05 hasta 0.95, aumentando 0.05 en cada iteración, sobre la captura LiDAR de un sector de la localidad de Paipote, Copiapó presentado en la Figura 13. El archivo original tiene un espacio de almacenamiento de 665.5 MB.

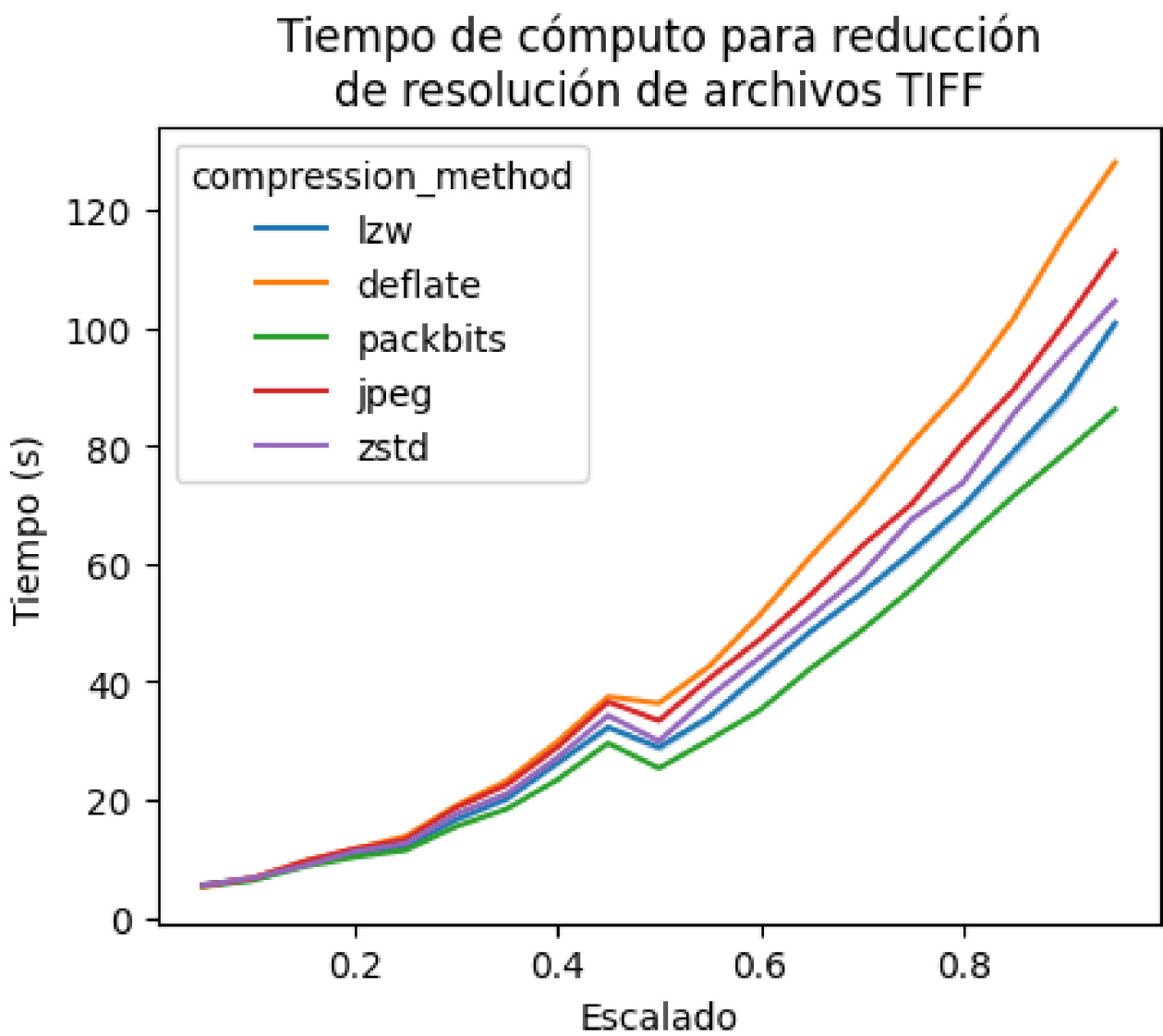


Figura 21: Tiempo de cómputo para reducción de celda de la Figura 13, en función de el porcentaje de escalado.

Una breve revisión de los algoritmos utilizados:

LZW: Lempel-Ziv-Welch es un algoritmo sin pérdida que reemplaza una secuencia repetida de caracteres con un código para reducir el tamaño total de la data. Construye un diccionario de códigos basado en la data de entrada, permitiendo una codificación y decodificación eficiente de los datos [Welch, 1984]. Deflate: Este algoritmo combina las codificaciones LZ77 y Huffman, reemplazando secuencias repetidas con referencias y usando tablas de código de largo variable para representar patrones frecuentes [Feldspar, 2002]. Packbits: Se trabaja reemplazando bytes consecutivos por bytes de control, seguido por la cantidad [Universal Document Converter, 2023].

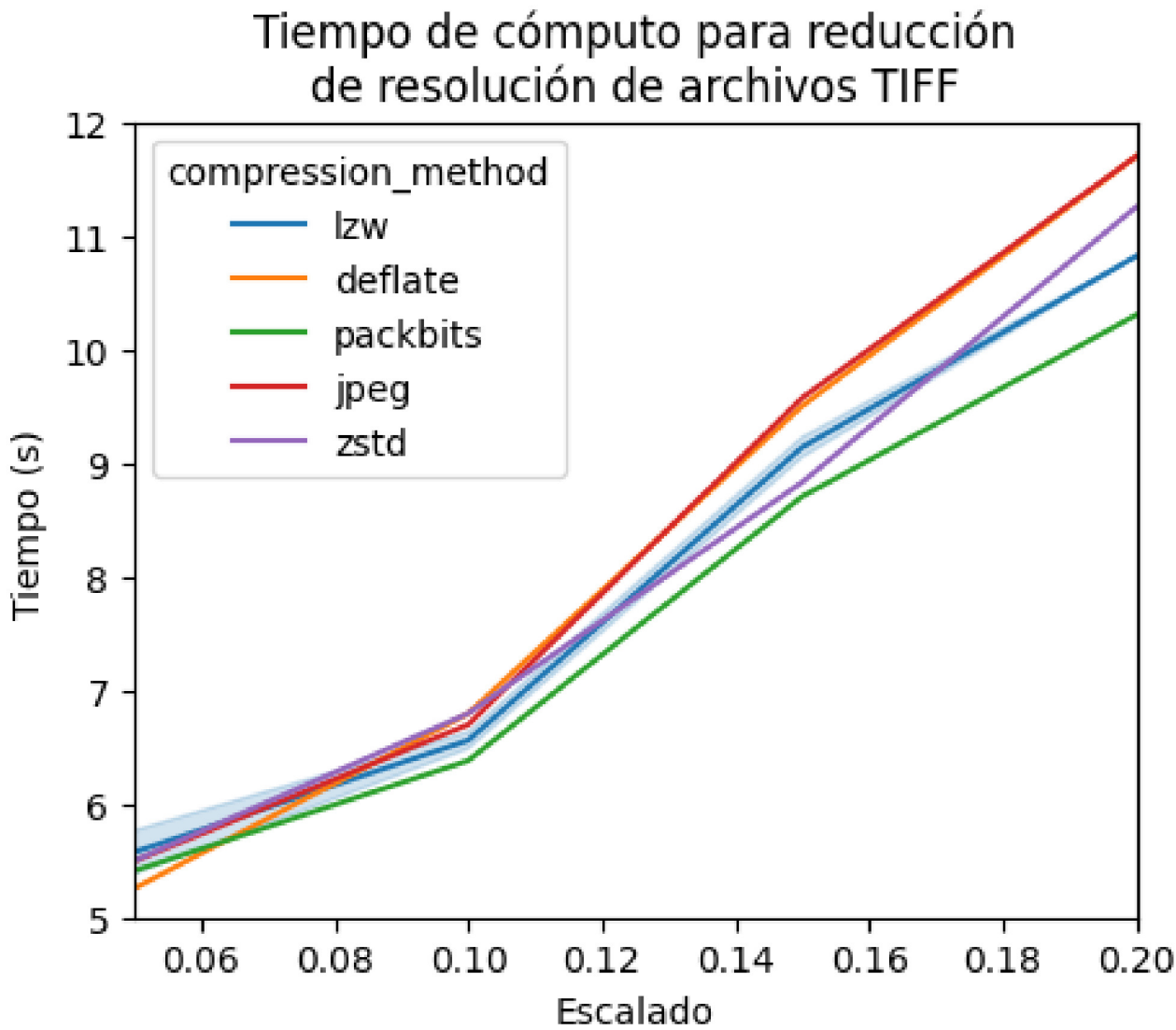


Figura 22: Acercamiento al gráfico de la Figura 21 desde escalados 0,05 a 0,2.

Por otro lado, el algoritmo JPEG se usó solo con motivos comparativos, pues se debe a un algoritmo de alta eficiencia pero con pérdida de información. JPEG es un algoritmo que divide la imagen en bloques, aplicando la transformada de coseno discreta para convertir los datos de la imagen en componentes de frecuencia, cuantificando estos componentes basados en factores de percepción, finalmente se le aplica codificación Huffman. Este algoritmo debe su eficiencia espacial a su capacidad de remover información de alta frecuencia, limitando la información que se mantiene a los límites de la sensibilidad de la visión humana para percibir detalles [Wallace *et al.*, 1991].

En la Figura 21 vemos el desempeño temporal de los diferentes algoritmos. Vemos que a medida que el porcentaje de información que se mantiene, el tiempo de cómputo aumenta y

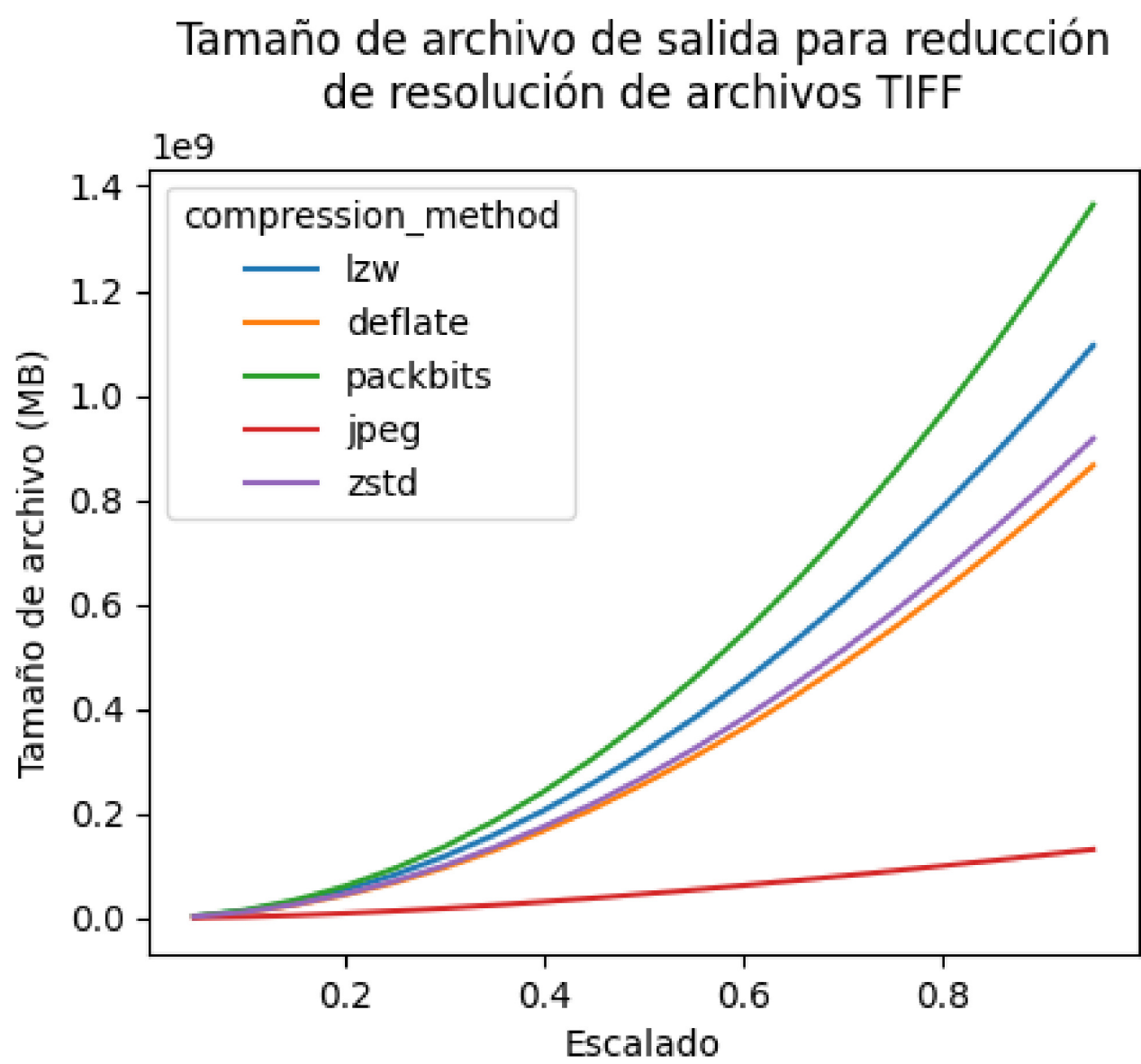


Figura 23: Tamaño de archivo de salida en disco para reducción de celda de la Figura 13, en función de el porcentaje de escalado.

Algoritmos de Compresión	Original	jpeg	deflate	zstd	lzw	packbits
Ratio de Compresión	1.0	14.40	2.59	2.47	2.09	1.76
Tamaño en MB	665.5	46.2	256.5	269.4	318.1	378.2

Tabla 7: Comparación espacial para un escalado de 0,5 del ráster de la Figura 13

la diferencia de tiempo entre estos algoritmos aumenta. Esto es, si tomamos un escalado de 0,8, packbits tiene el mejor desempeño con 60 segundos y deflate el peor con 85 segundos aproximadamente. En la Figura 22 se realizó un acercamiento a los escalados más pequeños, donde podemos identificar que jpeg junto con deflate tienen el peor desempeño, y packbits se encuentra en los menores tiempos de ejecución. Vale mencionar que no se tiene identificada la razón para el decrecimiento uniforme del tiempo de ejecución para los algoritmos alrededor de 0,5 de escalado. Posiblemente se debe a una instanciación de otros servicios del sistema operativo en ese instante.

Posteriormente, en la Figura 23 identificamos el rendimiento de los algoritmos en cuanto al tamaño del archivo de salida en disco. El archivo original tiene un peso de 665.5 MB en disco. Notamos que el mejor desempeño lo tiene jpeg logrando un tamaño de salida de 151.9 MB (ratio de compresión de 4,38 : 1) para un escalado de 1,0. Mientras que los archivos de salida de los algoritmos lossy requieren de un espacio en disco bastante de superior, posicionándose como se presenta en la Tabla 7.

Esta comparación se realizó sobre los archivos escalados, pues al comprimir un archivo con escalado 1,0 aumenta el tamaño del archivo TIFF de salida en los casos probados. De hecho, se dejó como cota superior 0,7 para escalados en el sistema. Posiblemente al agregar metadata o información adicional aumenta el tamaño, o al tratarse de archivos que ya fueron comprimidos anteriormente con el mismo u otro algoritmo.

En vista de los datos, se utilizará **zstd** por defecto. Se seleccionó este algoritmo por su balance entre el tiempo de cómputo y la eficiencia espacial es decir, su ratio de compresión. No obstante lo anterior, se permite el uso de los otros algoritmos si por alguna razón el usuario lo requiriera.

3.6. INTEGRACIÓN CON API

Chile 3D consta de una API ya desarrollada por otra sección del proyecto. Esta API fue desarrollada en el framework web FastAPI ³ para Python, y se encontrará descrita a detalle la memoria “Diseño e implementación de API para la gestión de datos asociados a información altimétrica de alta resolución” de Felipe Otero [Otero, 2023] (en proceso de publicación). El código de la API se encuentra disponible en <https://github.com/foterolu/chile3d> [Aubel *et al.*, 2023], además de una implementación de una API solo con las funciones de este informe en https://github.com/maxaubel/Chile3D_processing

La integración con la API consta de las siguientes fases, la primera ya fue finalizada, y las segundas constituyen desarrollo directo en la API (lo que será detallado en la presente sección):

1. Inicialmente se desarrollaron las funcionalidades en Python en un entorno local. Desde el diseño hasta la implementación de las funciones, incluyendo la iteración hasta tener funciones refinadas. Buscando el desarrollo de un código con bibliotecas con LTS (Long-Time Support: Soporte de largo plazo). Además, se realizaron pruebas de rendimiento en un entorno controlado, permitiendo la medición y comparación de aquellas funciones que son críticas.
2. Luego, se implementarán las funciones seleccionadas de acuerdo al rendimiento y necesidades del usuario en la API en desarrollo por la otra sección del proyecto. El desarrollo se realizó en FastAPI, lo que facilita en gran medida la integración pues comparte el lenguaje de programación.
3. Finalmente se evaluará si la implementación cumple con los requisitos propuestos, asegurándose si las funciones de la API cumplen con las entradas y salidas esperadas. Además, se revisarán los archivos procesados para poder determinar si estos difieren de los archivos procesados de manera local.

La implementación se realizará de manera de dejar acceso a las funciones de la forma:

`ip:port/function`

³FastAPI, <https://fastapi.tiangolo.com>

Siendo `function` el nombre de la función que se quiere ejecutar del siguiente listado de requests tipo POST:

1. `las_to_tiff`: Transformación desde nube de puntos LAS o LAZ a Ráster GeoTIFF. El cuerpo de la request:

```
{
  "file": INPUT_FILE
  "resolution": RESOLUTION, # raster cells side size
  "radius":     RADIUS      # sampling radius
}
```

RESOLUTION hace alusión al tamaño de las celdas del raster GeoTIFF de salida, RADIUS es el radio de muestreo en la nube de puntos original.

2. `las_res_red`: Reducción de resolución de nube de puntos LAS o LAZ. Retorna independiente del formato de entrada, un archivo LAZ. El cuerpo de la request:

```
{
  "file": INPUT_FILE,
  "cell": CELL, # cell or voxel size
  "method": METHOD # sampling method
}
```

METHOD corresponde al método de muestreo del siguiente conjunto:

- `sample`
- `voxelcenternearestneighbor`
- `voxelcentroidnearestneighbor`
- `voxeldownsize`

3. `tiff_res_red`: Reducción de resolución de ráster GeoTIFF, incluida la compresión lossless.

```
{
  "file":    INPUT_FILE,
  "scale":   SCALE,
  "compression_method": COMPRESSION_METHOD # compression method
}
```

`COMPRESSION_METHOD` corresponde al algoritmo de compresión del siguiente conjunto:

- `lzw`
- `deflate`
- `packbits`
- `jpeg`
- `zstd`

Esto permitirá que las funciones sean accedidas tanto desde dentro de la aplicación como desde fuera, contando con los token de autorización correspondientes. Las funciones no retornan el archivo procesado directamente, sino que retornan la dirección del archivo procesado. Lo anterior posibilita a la API a ser usada como el usuario estime, incluso ser accedido desde aplicaciones de terceros teniendo el token correspondiente en la solicitud.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

Por tratarse de un proyecto que se constituye de 3 secciones separadas que convergen en un desarrollo a gran escala, con un número de datos necesarios para su operatividad objetivo no logvable a corto plazo, se dificulta la validación de la propuesta en el entorno para el que fue gestado el proyecto. Es decir, en la presente fase se evaluará si los archivos de salida se alinean con el objetivo de cada función, mas no se podrá evaluar con la cantidad de datos y usuarios a los que apunta el proyecto Chile 3D finalmente.

Fueron ejecutadas las 3 funciones de la API, con archivos desde una máquina local, iterando sobre un conjunto discretizado de todas las combinaciones posibles de parámetros. Debido a que estas funciones trabajan directamente sobre los archivos, son completamente modulares e independientes de otras funciones que tiene la API. Al recibir directamente la dirección de los archivos a procesar y retornar la dirección de los resultados, se logra evitar el manejo de archivos directamente con la función que estaría haciendo uso de estas funciones.

Respecto a la implementación en la API, se desarrollaron inicialmente las funciones de manera atómica, para ser finalmente adaptadas modularmente a la API provista por Felipe Otero. Las características modulares han permitido especialmente evitar problemas de compatibilidad entre ambas secciones del proyecto.

La validación se basará en analizar lo siguiente:

1. ¿Los archivos de salida de las funciones de reducción de resolución entregan archivos con la resolución correcta?
2. ¿Los archivos de salida de las funciones de reducción de resolución entregan archivos con menor tamaño en disco?
3. ¿El archivo de salida de la transformación de una nube de puntos a ráster tiene congruencia visual?
4. ¿Las funciones de la API retornan los formatos correctos y con archivos con el formato correcto en el directorio indicado?

Para las preguntas que aluden al tamaño en disco, se debe considerar que la reducción del mismo podría no existir si estos datos ya han sido comprimidos con ese o con otro algoritmo previamente, lo que no es indicado por el proveedor de la data.

Se trabajará con un archivo que no se había utilizado en pruebas antes, un modelo de la región de Beaver Hills en Alberta, Canadá [Open Data Areas Alberta, 2009] (ver Figura 24 como referencia). En la descarga, es posible encontrar cinco carpetas:

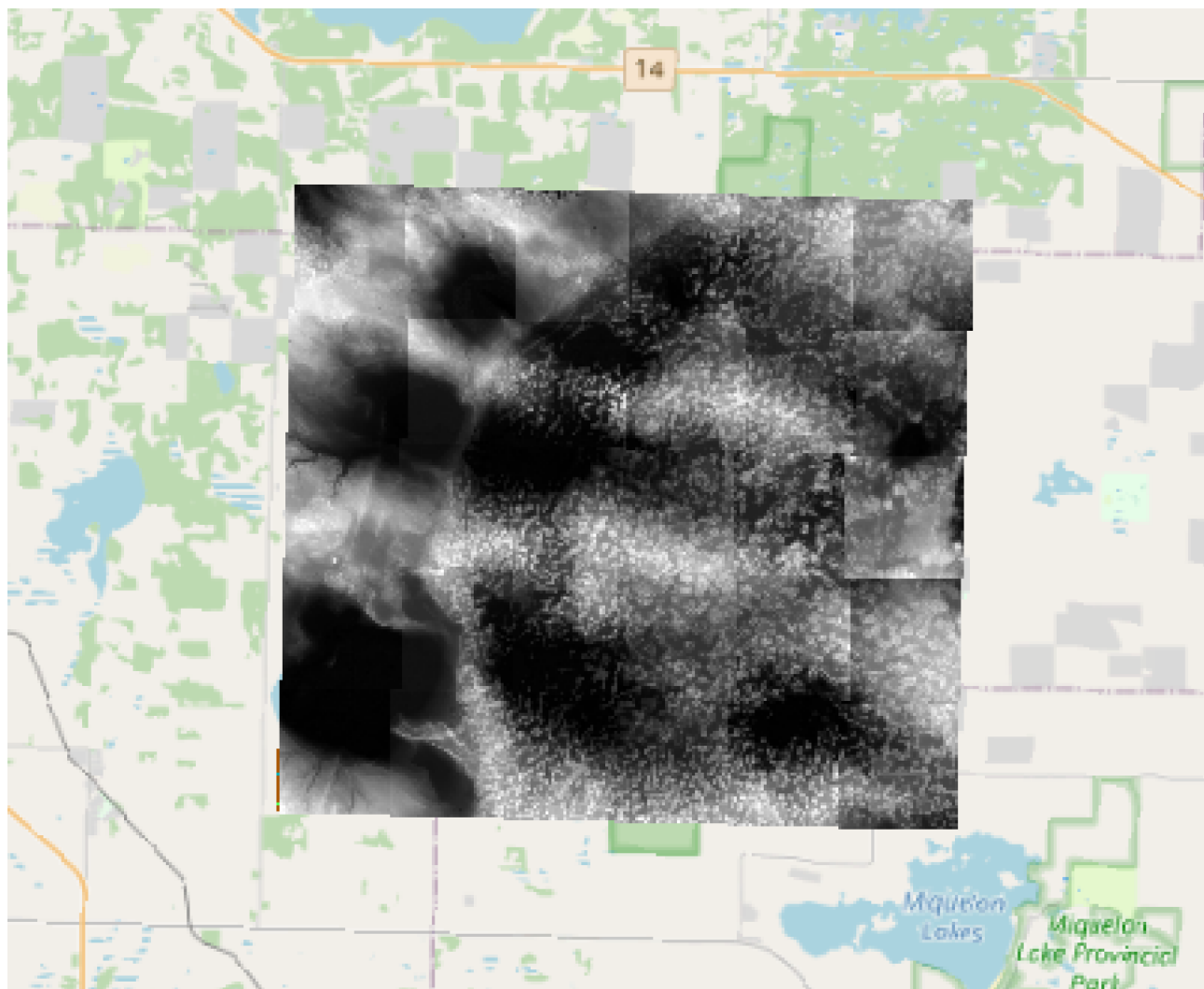


Figura 24: Captura de las 30 celdas GeoTiff que componen el registro de la zona de Alberta.

- **bare_earth_DEM_32bit_Geotiffs**
- bare_earth_hillshade_tif
- full_feature_DEM_32bit_Geotiffs
- full_feature_hillshade_tif
- **point_cloud_las**

Debido al número de variaciones y de procesamientos que ofrece el archivo, este análisis se enfocará en las carpetas resaltadas. Se verá si con la herramienta implementada, se cumplen los objetivos, i.e. se responden las cuatro preguntas anteriores. Se tiene el archivo original en nube de puntos y en raster para poder comparar visualmente las salidas de cada una de las funciones.

En esta sección analizaremos visualmente solo las capturas de la celda E14, que corresponde a la celda al suroeste de la grilla, con un área de 4.847.153,091 metros cuadrados.

	Nro de Puntos	Tamaño archivo	Tipo de Dato	Ancho	Alto
Archivo Original LAS	3.373.669	90.09 MB	N/A	N/A	N/A
Archivo Original TIFF	N/A	19.87 MB	Float32	2170	2398
Archivo Salida TIFF	N/A	38.97 MB	Float64	2147	2378

Tabla 8: Comparación Archivo LAS original y archivo TIFF procesado.

Conversión nube de puntos a ráster

Al ejecutar `las_to_tiff`, generamos una solicitud con el siguiente cuerpo:

```
{
  "file":      "point_cloud_las/pc_083H06SE14.las",
  "resolution": "1",
  "radius":    "1"
}
```

Cuya respuesta es la siguiente:

```
{
  "output_file": "point_cloud_las/pc_083H06SE14.tif",
  "time":        "1.6106441020965576"
}
```

Para esta comparación se necesita utilizar los archivos GeoTiff con todas las características, no los filtrados que se usaron anteriormente (debido a la alta resolución en los datos LAS). Respecto a los valores específicos de resolución y radio al transformar de nube de puntos a ráster, depende de la aplicación. De la misma forma, al realizar este procesamiento, se desconoce cuales son los parámetros que utilizó el proveedor. Lo anterior dificulta comparar precisamente la calidad del ráster obtenido, pues existirán una disparidad incluso en el número de puntos de cada archivo.

No obstante lo anterior, en la Figura 25 se aprecia nube de puntos original del proveedor, junto con el ráster correspondiente del proveedor en la Figura 26. En la Figura 27 se observa el ráster obtenido. En la Tabla 8, es posible dilucidar que se logra una resolución similar a la del proveedor (no de manera intencional por lo dicho anteriormente), además, notamos que el archivo TIFF de salida es tipo Float64, es decir, se está trabajando con datos con el doble de bits de resolución. Por esta razón el tamaño de archivo de salida también es aproximadamente el doble.

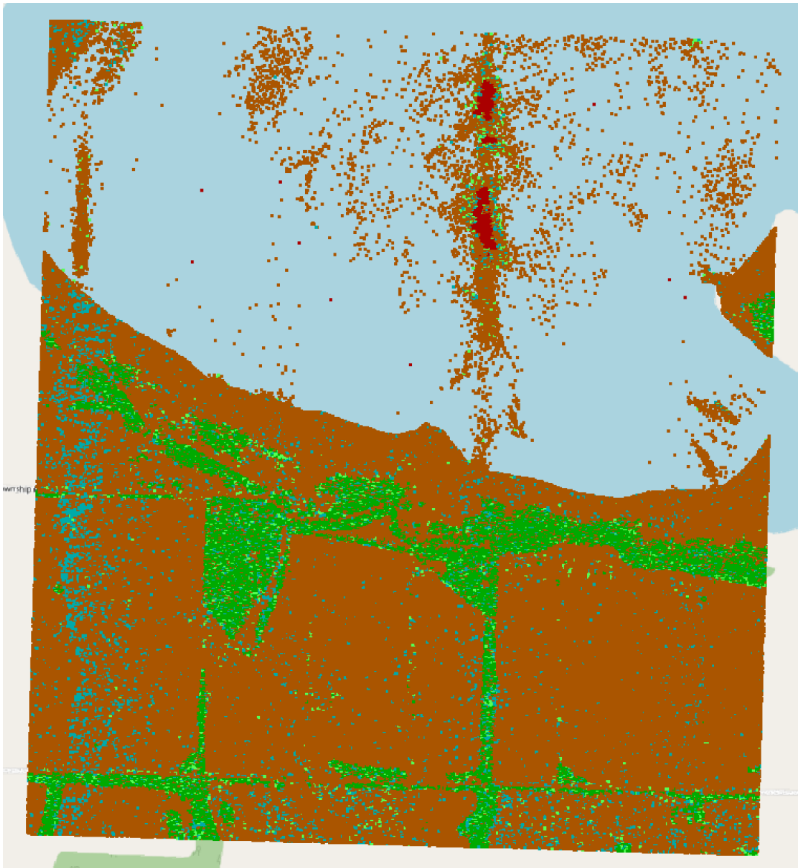


Figura 25: Nube de puntos LAS original.

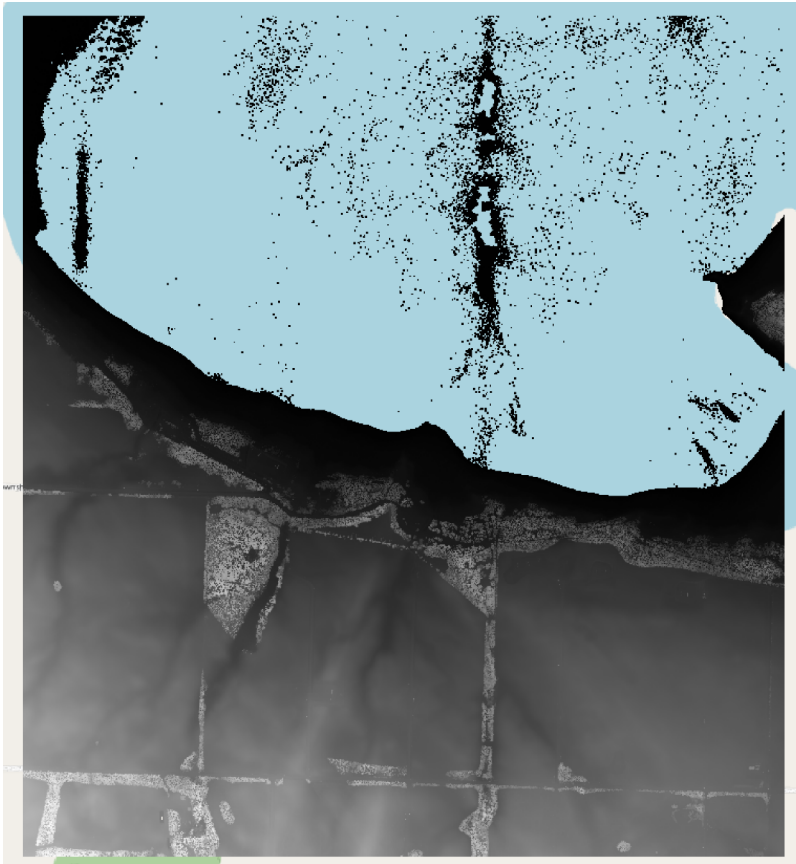


Figura 26: Ráster del proveedor.



Figura 27: Ráster obtenido con implementación.

Reducción de resolución de nubes de puntos

Para las nubes de puntos LAS, se ejecuta la función `las_res_red`, con el siguiente cuerpo:

```
{  
  "file":    "point_cloud_las/pc_083H06SE14.las",  
  "cell":    "2",  
  "method":  "sample"  
}
```

Cuya respuesta es la siguiente:

```
{  
  "output_file": "point_cloud_las/pc_083H06SE14.las_sample_1.0.laz",  
  "time":        "4.499506235122681"  
}
```

En las capturas (ver Figuras 28 y 29) vemos pequeños cambios con un análisis visual. No obstante, al acercarse el observador (ver Figuras 30 y 31), podemos notar diferencias en la densidad de puntos. Además, es posible visualizar el efecto de los voxel al muestrear puntos de manera geométrica, generando un orden sobre los puntos de la nube de puntos en el archivo de salida.

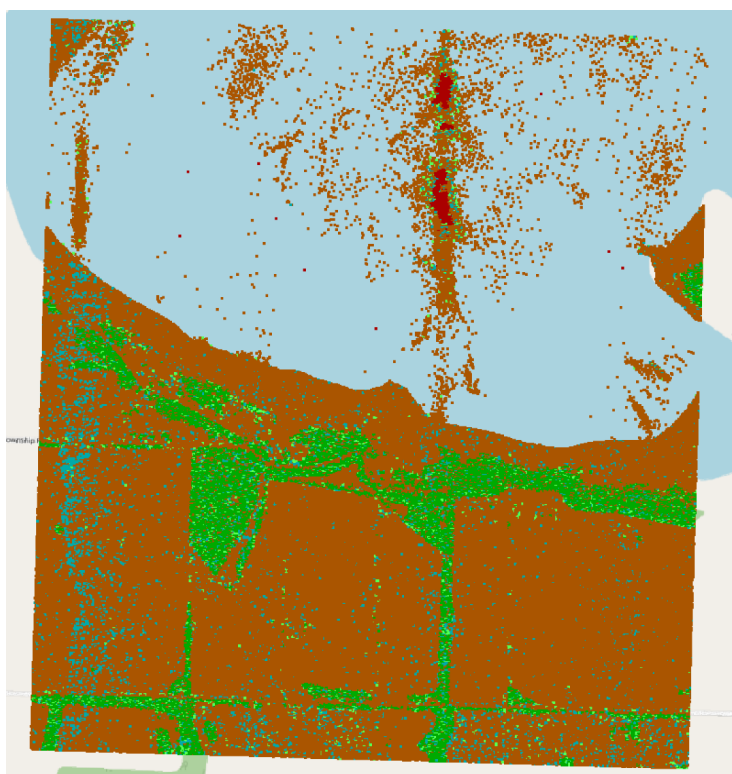


Figura 28: LAS original

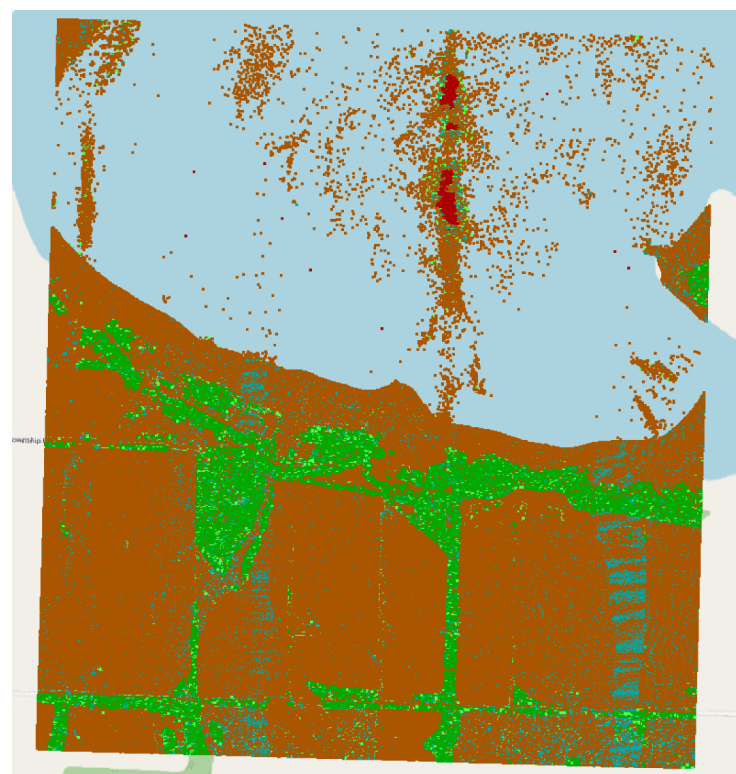


Figura 29: LAZ con celdas de 2 metro

Al comparar ambos archivos (ver Tabla 9), se aprecia que se logra un submuestreo de acuerdo a lo solicitado. Además, el tamaño del archivo de salida no es proporcional con el de entrada, sino que es bastante menor debido a la compresión LAZ.

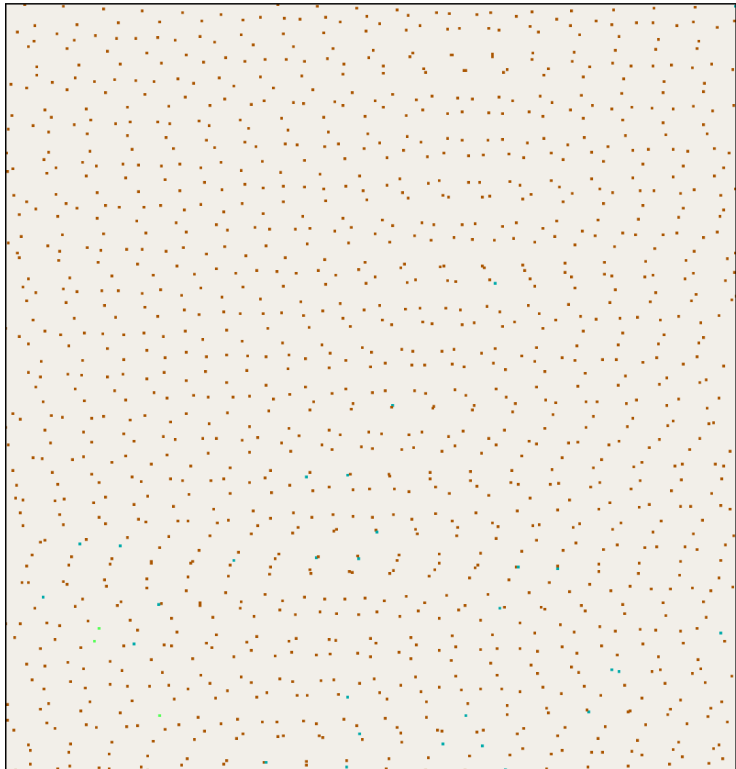


Figura 30: LAS original

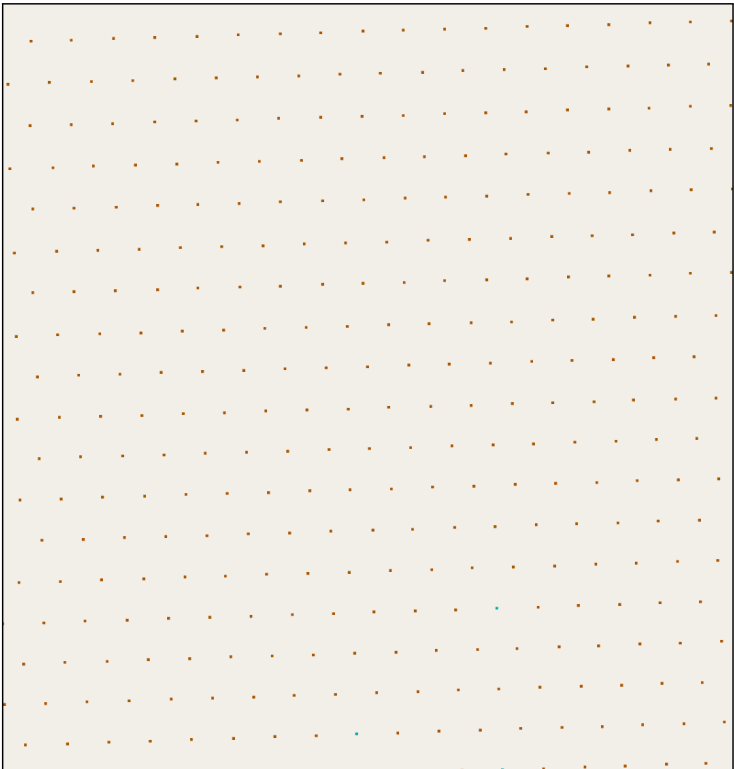


Figura 31: LAZ con celdas de 2 metro

	Celda	Número de Puntos	Puntos/ m^2	Tamaño archivo
Archivo Original	N/A	3.373.669	0.696	90.09 MB
Archivo Salida	2 metros	699.732	0.144	2.8 MB

Tabla 9: Comparación Archivo LAS original y archivo LAZ procesado.

Reducción de resolución de ráster

Se ejecuta la función `tiff_res_red` con el siguiente cuerpo:

```
{
  "file":    "bare_earth_DEM_32bit_Geotiffs/be_dem_083H06SE14.tif",
  "scale":  "0.5"
  "compression_method": "zstd"
}
```

Para la entrada indicada, la API retornó lo siguiente:

```
{
  "output_file":
    "bare_earth_DEM_32bit_Geotiffs/be_dem_083H06SE14_zstd_0.5.tif",
  "time": "0.5570354461669922"
}
```

Al comparar ambos archivos (ver Tabla 10), se aprecia que se logra escalar la resolución correctamente. Además, el tamaño del archivo de salida no es proporcional con el de entrada,

	Escala	Ancho	Alto	Tamaño archivo
Archivo Original	1	2170	2398	19.87 MB
Archivo Salida	0.5	1085	1199	3.0 MB

Tabla 10: Comparación Archivo GeotIFF original y archivo procesado con escala 0.5

sino que es menor debido a la compresión que se le aplica.

En las capturas (ver Figuras 32 y 33) no es posible detectar diferencias al observar la imagen de forma general. Las diferencias son perceptibles al acercarnos al nivel de poder ver los píxeles, los que naturalmente tendrán el doble del tamaño original.

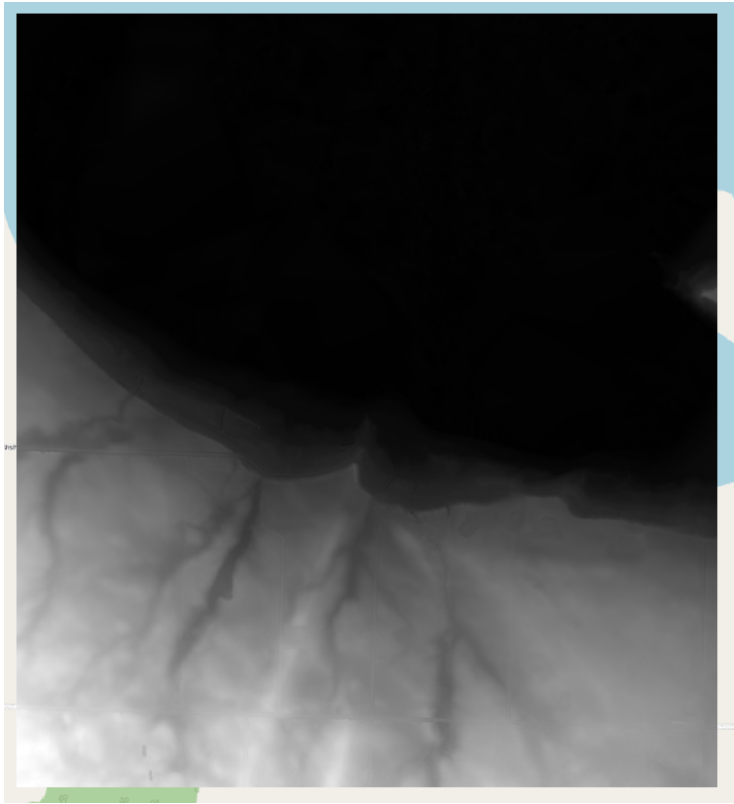


Figura 32: GeoTiff original

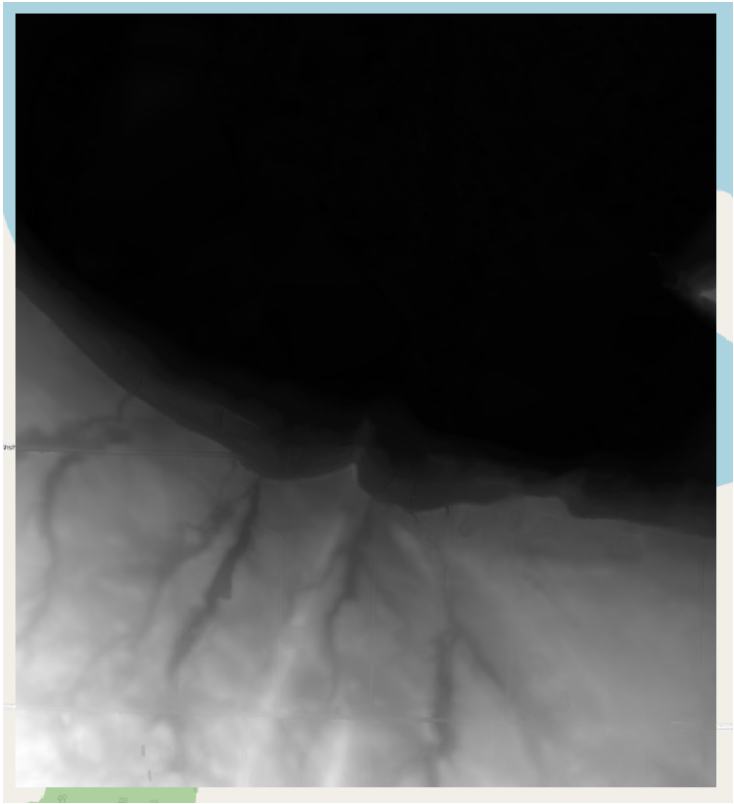


Figura 33: GeoTiff escalado al 0.5

CAPÍTULO 5

CONCLUSIONES

La disponibilidad y precisión de datos altimétricos son fundamentales para una gran número de industrias, desde obras civiles y minería hasta estudios de desastres naturales y agronomía, además de estudios realizados por la academia. En Chile, sin embargo, no existe una plataforma que centralice y proporcione estos datos de manera eficiente y accesible al usuario, tanto empresa, academia, como personas naturales. Con el propósito de abordar esta necesidad, surge el proyecto Chile 3D, una iniciativa destinada a crear una plataforma integral para la recolección, administración, formateo y disponibilización de datos altimétricos de todo el país.

La investigación actual se enfoca en los aspectos clave del formateo de datos, su manejo adecuado y su transformación para adaptarse a las necesidades específicas de los usuarios. Esto implica diseñar una infraestructura sólida y eficiente que permita recolectar y almacenar datos altimétricos, así como implementar procesos de formateo y transformación que garanticen la precisión y coherencia de la información proporcionada.

Al desarrollar una plataforma de este tipo, Chile 3D contribuirá significativamente al avance de diversas áreas de aplicación, brindando a los profesionales y tomadores de decisiones acceso rápido y confiable a datos altimétricos. Esto impulsará un desarrollo más fluido y eficiente en sectores mencionados anteriormente, como la ingeniería civil, la minería, la gestión de desastres naturales y la agronomía, entre otros.

En resumen, el proyecto Chile 3D busca llenar el vacío existente en Chile al proporcionar una plataforma integral y centralizada para la gestión y disponibilidad de datos altimétricos. Su enfoque en el formateo de datos, manejo adecuado y transformación beneficiará a múltiples industrias y contribuirá al desarrollo y avance de diversas disciplinas en el país.

Mediante la API, se puede acceder a las funciones implementadas, las que constituyen solo una sección de todas las funcionalidades ofrecidas por la plataforma Chile 3D. Se requiere solo enviar una request y el sistema retornará la ubicación del nuevo archivo de salida. Se pensó en que las funciones serán utilizadas por otras partes del sistema, no directamente por el usuario, pues son procesamientos que implican una altísima carga de cómputo. Se considera que no es escalable el procesamiento de datos a demanda del usuario por la razón antes expuesta. Por ejemplo, si cada usuario requiriera escalar al 50 % modelos altimétrico con formato original LAS de 2 GB, significaría una alta demanda tanto del tiempo de procesamiento como de la RAM necesaria para realizar estos procesos. Localmente, en la máquina descrita inicialmente, fue posible realizar un solo proceso en un instante. Al intentar ejecutar más de dos, por temas de RAM no fue posible (8 GB de capacidad de RAM).

Objetivos. Revisando los objetivos específicos presentados inicialmente, se logró implementar algoritmos de cambio de formato para modelos georreferenciados de altimetría, centrándose en los formatos más utilizados: GeoTIFF y LAS/LAZ. La inevitable pérdida de información queda en manos del usuario experto, dejando el espacio para configurar los parámetros como el tamaño de los voxels de búsqueda, tamaño de celdas, radio, entre otros. La transformación desde GeoTIFF a nube de puntos se dejó de lado, pues implica convertir un dato que describe usualmente superficies por uno que representa a detalle las características terrestres y de objetos que se encuentren sobre la superficie con nubes de puntos.

El trabajo con datos altimétricos presenta una dificultad para los usuarios por su gran tamaño en memoria. Esto significa tener problemas para abrir los archivos en software GIS por ejemplo, al contar con archivos con niveles de calidad o resolución innecesariamente altos para algunas aplicaciones. Sumado a lo anterior, a mayor tamaño de los archivos, mayor carga sobre el sistema de almacenamiento, mayor carga de la red y mayor tiempo de espera para el usuario.

Esto se buscó solucionar con algoritmos de reducción de resolución y compresión de los archivos. La reducción de resolución se deshace de información innecesaria para la aplicación específica del usuario. Por ejemplo, si un usuario necesita hacer análisis de fenómenos macroscópicos, no será necesario que tenga acceso a detalles en la vegetación que probablemente se pierdan al reducir la resolución.

Por otro lado, el sistema permite comprimir los archivos altimétricos. Las nubes de punto LAS siempre serán transformadas a LAZ, formato que disminuye el espacio de almacenamiento, aprovechando las características específicas de este tipo de datos. La mayor ventaja, además de la ganancia espacial, es que ambos formatos son compatibles. Esto posibilita por ejemplo, abrir archivos LAZ como si fuera un archivo LAS, sin la necesidad de descomprimir como si hubiera sido comprimido con algún algoritmo con fines generales como ZIP o WinRAR. Además, los ráster en formato GeoTIFF también pueden ser comprimidos. Existen dos caminos: algoritmos lossy ó lossless. Los lossy lograrán tener mejor desempeño espacial, pero con pérdida de información. Esto se logra al omitir información similar, o bien, describiéndola de alguna manera que agrupa datos similares. Y los algoritmos lossless, que son algoritmos de compresión sin pérdida, no logran el desempeño espacial que los lossy, pero dependiendo del algoritmo se puede tener un rendimiento aceptable en comparación.

Respecto al objetivo que guarda relación con desarrollar funcionalidades fácilmente integrables con las otras partes del sistema del proyecto Chile 3D, se desarrolló una API con la biblioteca FastAPI. Esto permite tener una API con un desarrollo amigable, que no es dispar al paradigma de programación propio de Python comparado con otras bibliotecas. Esto permitió un desarrollo rápido y modular. Las funciones desarrolladas en el presente área del proyecto fueron creadas siguiendo este paradigma. Sumado a lo anterior, al recibir la dirección de archivos de entrada y entregar la dirección de aquellos de salida, se gana modularidad que facilita su integración en otras partes del sistema, o su uso por una aplicación externa según sea el caso.

Dificultades presentadas. Respecto a las dificultades presentadas, radican principalmente en el estudio exhaustivo de la materia relacionada a los datos georreferenciados.

Para poder interpretar las necesidades de la industria y reflejarlo en soluciones concretas, fue necesario estudiar a cabalidad tanto los usos de cada uno de los formatos, como los formatos en sí. Cada uno de los tipos de datos tiene estructuras, convenciones, y requerimientos específicos, lo que ha exigido un profundo análisis y comprensión, incluso para implementar la función más simple. Además, ha sido necesario familiarizarse visualmente con los datos altimétricos. Esto permitió poder interpretar y evaluar de manera ocular la congruencia de los resultados obtenidos.

El proyecto Chile 3D ha buscado mantenerse siempre dentro del margen del código abierto, es por ello que se han usado, al menos en esta sección del desarrollo, bibliotecas de código abierto. La dificultad relacionada a esto radica en varios factores: primero, la búsqueda de bibliotecas que faciliten la tarea que se busca lograr; segundo, entre las bibliotecas que se encontraron, se deben seleccionar aquellas que funcionen de forma armoniosa con otras bibliotecas más esenciales; tercero: se debe incorporar la biblioteca en la API, lo que además necesita que la biblioteca sea fácil de usar/implementar y mantenga una documentación amigable y actualizada.

El mantenimiento y la compatibilidad a lo largo de la vida útil de Chile 3D puede presentar un desafío, al actualizar otras partes del software por ejemplo, o al migrar a versiones más recientes de las bibliotecas usadas. Esto se solucionó usando Entornos Virtuales lo que permite, en resumen, ejecutar código en un entorno acotado, que incluya las bibliotecas con las versiones específicas. Esto evita tener conflictos con otras instalaciones que se encuentren en la máquina.

No obstante a las dificultades relacionadas al código abierto o el software gratis, el código no requiere de ninguna biblioteca propietaria. La principal ventaja de ello es la disminución de costos. Al escalar el proyecto será necesario evaluar constantemente si el uso de estas bibliotecas afecta al desempeño, y eventualmente migrar a usar bibliotecas propietarias si la demanda de servicios lo requiere.

Trabajo a Futuro. Además de LiDAR, se debería incorporar la interferometría como una fuente de datos. La interferometría es una técnica usada en diferentes campos, incluyendo astronomía, geodesia, e ingeniería; utilizada para medir y analizar los patrones de interferencia creados al combinar una o más ondas. En el contexto de este proyecto, en el campo de la geodesia, la interferometría se usa para medir las deformaciones del terreno con precisión. Esto se logra analizando la interferencia de los patrones creados por rayos de radar o láser, reflejados por la superficie terrestre. De esta forma es posible detectar y monitorear movimientos sutiles e imperceptibles con otros métodos, como por ejemplo movimiento de placas tectónicas, actividad volcánica, hundimiento del suelo, entre otros.

En resumen, al agregar la interferometría se logra incluir al área de la geodesia, que cumple un rol fundamental en un país altamente sísmico como Chile. Las altas resoluciones y precisión permite analizar a fondo los fenómenos nombrados, permitiendo monitorear y estudiar a más detalle el terreno nacional.

La gran mayoría de modelos altimétricos tiene la característica de contar con datos estructurados, atómicamente con poca información, pero simultáneamente destacan por su densidad de datos. Es por ello que surge la idea de utilizar procesamiento paralelo para procesar estas grandes cantidades de datos simples. Se desarrollaron funciones utilizando CUDA, que aprovecha el gran número de núcleos de las unidades de procesamiento gráfico, también usadas para algoritmos con datos similares. La implementación construida sin mayor análisis no logró superar a aquellas que hacían uso de la CPU, no obstante, existe una posibilidad que una implementación que aproveche concurrencia y paralelismo logre superar con creces al desempeño de la ejecución secuencial.

REFERENCIAS BIBLIOGRÁFICAS

- [Abdel-Fattah *et al.*, 2017] Abdel-Fattah, M., Saber, M., Kantoush, S., Khalil, M., Sumi, T., y Sefelnasr, A. (2017). A hydrological and geomorphometric approach to understanding the generation of wadi flash floods. *Water*, 9:553.
- [Aubel *et al.*, 2023] Aubel, M., Otero, F., y Lanás, J. P. (2023). Chile 3D API. <https://github.com/foterolu/chile3d>.
- [Council, 2007] Council, N. R. (2007). *Elevation Data for Floodplain Mapping*. The National Academies Press, Washington, DC.
- [Council, 2009] Council, N. R. (2009). *Mapping the Zone: Improving Flood Map Accuracy*. The National Academies Press, Washington, DC.
- [Dewberry, 2012] Dewberry (2012). *National Enhanced Elevation Assessment Final Report Fairfax*.
- [Feldspar, 2002] Feldspar, A. (2002). An Explanation of the Deflate Algorithm. <https://zlib.net/feldspar.html>.
- [Guth *et al.*, 2021] Guth, P. L., Van Niekerk, A., Grohmann, C. H., Muller, J.-P., Hawker, L., Florinsky, I. V., Gesch, D., Reuter, H. I., Herrera-Cruz, V., Riazanoff, S., López-Vázquez, C., Carabajal, C. C., Albinet, C., y Strobl, P. (2021). Digital elevation models: Terminology and definitions. *Remote Sensing*, 13(18).
- [Instituto Nacional de Estadísticas de Chile, 2006] Instituto Nacional de Estadísticas de Chile (2006). Compendio: Estadístico 2006. https://www.ine.gob.cl/docs/default-source/nacimientos-matrimonios-y-defunciones/publicaciones-y-anuarios/anuarios-de-estad%C3%ADsticas-vitales/ine_anuario-de-estad%C3%ADsticas-vitales_2006.pdf.
- [Instituto Nacional de Estadísticas de Chile, 2017] Instituto Nacional de Estadísticas de Chile (2017). *sisntesis-de-resultados-censo2017.pdf*. <https://www.ine.gob.cl/docs/default-source/censo-de-poblacion-y-vivienda/publicaciones-y-anuarios/2017/publicaci%C3%B3n-de-resultados/sintesis-de-resultados-censo2017.pdf>.
- [Isenburg, 2013] Isenburg, M. (2013). Laszip: lossless compression of lidar data. *Photogrammetric engineering and remote sensing*, 79(2):209–217.
- [Ministerio de Bienes Nacionales, 2017] Ministerio de Bienes Nacionales (2017). Imagen copiápó: Paipote 08-09-2017. <https://www.ide.cl/index.php/imagenes-y-mapas-base/item/1582-imagen-copiapo-ciudad-parte-1-11-09-2017>.

- [Ministerio del Interior y Seguridad Pública, 2021] Ministerio del Interior y Seguridad Pública (07/08/2021). Ley 21.364 establece el sistema denacional de prevención y respuesta ante desastres, sustituye la oficina nacional de emergencia por el servicio nacional de prevención y respuesta ante desastres y adecúa normas que indica. <http://bcn.cl/2qhx8>.
- [Open Data Areas Alberta, 2009] Open Data Areas Alberta (2009). Beaver Hills: Lidar point cloud and derivative products. <https://www.opendataareas.ca/products/beaver-hills/lidar/>.
- [Otero, 2023] Otero, F. (2023). Diseño e implementación de API para la gestión de datos asociados a información altimétrica de alta resolución. Manuscrito en preparación para su publicación.
- [PDAL, 2022] PDAL (2022). Pdal. <https://github.com/PDAL/PDAL>.
- [Rafferty, 2009] Rafferty, J. P. (23 Jul. 2009). Hypsometry. *Encyclopedia Britannica*.
- [Smathermather, 2017] Smathermather (2017). Taking Slices from LiDAR data: Part VII. <https://smathermather.com/2017/02/15/taking-slices-from-lidar-data-part-vi-2>.
- [The American Society for Photogrammetry & Remote Sensing (ASPRS), 2019] The American Society for Photogrammetry & Remote Sensing (ASPRS) (2019). Las specification 1.4 - r15. https://www.asprs.org/wp-content/uploads/2019/07/LAS_1_4_r15.pdf.
- [Triglav, 2000] Triglav, J. (2000). Ecw: Wavelet compression beyond limits. *Geoinformatics*, pp. 34-41.
- [Universal Document Converter, 2023] Universal Document Converter (2023). PACKBITS Compression or Why We Support Lossless TIFF Compression Method? <https://www.universal-document-converter.com/news/packbits-tiff-i2498.html>.
- [USGS, 2014] USGS (2014). New york clinton essax lake champlain ql2 lidar. https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/LPC/Projects/2014_New_York_Clinton_Essex_Lake_Champlain_QL2_LiDAR/NY_ClintonEssexLakeChamplain_2014/LAZ/?C=S;0=A.
- [USGS, 2017] USGS (2017). Project il_4county_cook_2017. https://rockyweb.usgs.gov/vdelivery/Datasets/Staged/Elevation/LPC/Projects/USGS_LPC_IL_4County_Cook_2017_LAS_2019/laz/USGS_LPC_IL_4County_Cook_2017_LAS_13009275_LAS_2019.laz.
- [Wallace et al., 1991] Wallace, G. K., Friel, T., y Pratt, W. K. (1991). The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30-44.
- [Wang et al., 2018] Wang, R., Peethambaran, J., y Dong, C. (2018). Lidar point clouds to 3d urban models: A review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, PP.

[Welch, 1984] Welch, T. A. (1984). A technique for high-performance data compression. *Computer*, 17(6):8–19.