



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Análisis y evaluación del sistema de celdas Uber H3 en el
procesamiento de datos AIS*

Grado en Ingeniería Mecánica

ALUMNO: Mario Florido López

DIRECTORES: Miguel Rodelgo Lacruz

CURSO ACADÉMICO: 2024-2025

Universida_{de}Vigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Análisis y evaluación del sistema de celdas Uber H3 en el
procesamiento de datos AIS*

Grado en Ingeniería Mecánica
Intensificación en Tecnología Naval
Cuerpo General

UniversidadeVigo

RESUMEN

El presente trabajo analiza las prestaciones del sistema de indexación geoespacial Uber H3 en el procesamiento de datos geoespaciales generados por el *Automatic Identification System* (AIS), en un contexto donde el tráfico marítimo global exige soluciones capaces de manejar grandes volúmenes de información con precisión y eficiencia. Uber H3 organiza datos mediante particiones jerárquicas en celdas discretas para búsquedas rápidas y escalables, al ofrecer una estructura de geometría regular, se perfila como una alternativa frente a los sistemas tradicionales de indexación, permitiendo reducir distorsiones y mejorar el rendimiento de las operaciones de análisis. A través de simulaciones computacionales y aproximaciones analíticas, se evalúan aspectos clave como el error de posicionamiento inducido por la discretización en celdas, la cobertura espacial en áreas de interés y la representación de trayectorias AIS, todo ello considerando la relación entre granularidad y carga computacional. Los resultados permiten no solo constatar las capacidades del sistema en términos de escalabilidad y precisión, sino también identificar sus limitaciones y el impacto de estas en la interpretación y gestión de los datos. En conjunto, el estudio aporta una valoración crítica sobre la viabilidad operativa de H3 en aplicaciones marítimas, proponiendo nuevas vías para optimizar la organización y el análisis de información geoespacial en entornos dinámicos y de alta demanda.

PALABRAS CLAVE

Uber H3, Indexación espacial, Error, Posición, AIS

AGRADECIMIENTOS

En primer lugar, agradecer a mi tutor, Miguel Rodelgo, por mostrar interés desde el primer momento que le comuniqué que había elegido el trabajo propuesto por él. Su incansable voluntad de corregir todos los pequeños detalles hasta conseguir el objetivo deseado me ha servido para dar lo mejor de mí.

Agradecer a mis compañeros de la promoción 425. Estos 5 años de esfuerzo, con ellos han sido más amenos y llevaderos. Espero disfrutar de esta profesión con ellos muchos más años.

Quisiera expresar mi agradecimiento a la Armada que, desde una perspectiva personal, representa para mí la oportunidad de desempeñar el que considero el mejor trabajo del mundo: ser marino de España.

Por último, a mi familia. Sólo tengo palabras de agradecimiento por los valores que me han inculcado ejemplarmente, sin ellos no estaría aquí. Su paciencia infinita y respaldo inconmensurables han sido admirables en estos años de Escuela. Y, desde luego, a Elena, mi pareja. Su apoyo incondicional y su deseo de superación constante me han servido de piedra angular para esforzarme.

A todos ellos, infinitas gracias.

CONTENIDO

Contenido	1
Índice de Figuras	3
Índice de Tablas.....	5
1 Introducción y objetivos	7
1.1 Contexto y motivación	7
1.2 Objetivos	9
1.3 Estructura de la memoria	10
2 Estado del arte	13
2.1 Las proyecciones cartográficas	13
2.1.1 Antecedentes.....	14
2.1.2 Clasificación	14
2.1.3 Sistemas de proyección cartográfica	15
2.1.3.1 World Geodetic System 1984.....	15
2.1.3.2 European Terrestrial Reference System 1989	15
2.1.3.3 Universal Transverse Mercator	16
2.1.3.4 Proyección de Mercator.....	16
2.1.3.5 Proyección de Lambert Conformal Conic.....	16
2.1.3.6 North American Datum 1983	16
2.1.4 Problemas inherentes	16
2.2 Sistemas de indexado geoespacial	17
2.2.1 Uber H3	17
2.2.2 Google S2	20
2.3 AIS	22
2.3.1 Tipos de transpondedores y mensajes.....	22
2.4 Trabajos relacionados	24
3 Desarrollo	26
3.1 Entorno de trabajo y software	26
3.1.1 Python.....	26
3.1.2 Jupyter Notebook.....	27
3.1.3 Librerías	28
3.1.4 Kepler.gl	29
3.2 Consideraciones previas.....	29
3.2.1 Uso de IA generativa	29

3.2.2 Simulación Montecarlo.....	30
3.2.3 Generación de puntos	30
3.2.4 Cálculo distancias geodésicas.....	32
3.3 Cálculo del error de posición	33
3.4 Análisis de cobertura para un área de interés.....	43
3.5 Análisis de trayectoria.....	50
4 Análisis de resultados	61
4.1 Resumen y validación de resultados.	61
4.2 Discusión de resultados.....	63
4.3 Limitaciones.....	63
5 Conclusiones y líneas futuras	66
5.1 Conclusiones	66
5.2 Aplicaciones en casos reales	67
5.3 Líneas futuras	67
6 Bibliografía.....	69
Anexo I: Diccionario de siglas, acrónimos y abreviaturas	74
Anexo II: Reflexiones Éticas y Sociales	75
Anexo III: Implicaciones ambientales y económicas.....	76
Anexo IV: Código de distribución de distancias en un hexágono	78
Anexo V: Código de distribución de error para una resolución.....	79
Anexo VI: Código de error teórico vs real por resolución	80
Anexo VII: Código para visualizar distribución de error de un área en un mapa en Kepler.gl	81
Anexo VIII: Código de comparación Uber H3 con Google S2	82
Anexo IX: Código para comparación de área real vs área cubierta por celdas Uber H3	84
Anexo X: Código para obtener porcentaje de cobertura respecto al círculo real	85
Anexo XI: Código para el cálculo de desvíos en una trayectoria concreta	88
Anexo XII: Código para cálculo de desvíos de trayectoria para una resolución.....	90
Anexo XIII: Código para la comparación de transpondedores AIS en una trayectoria	93

ÍNDICE DE FIGURAS

Figura 1-1 Visualización de volumen de datos AIS [2]	7
Figura 1-2 Separación espacial Hexagonal de Uber H3 [3]	8
Figura 2-1 Esquema básico de idea de representación [6]	13
Figura 2-2 Mapamundi Al Idrisi, 1160 [10].....	14
Figura 2-3 Proyección conforme cónica [13].....	15
Figura 2-4 Proyección equivalente cilíndrica [14]	15
Figura 2-5 Proyección equidistante [15]	15
Figura 2-6 Icosaedro terrestre [3]	17
Figura 2-7 Distancia a celdas vecinas de lados y vértices [3]	18
Figura 2-8 División jerárquica de S2 [22].....	21
Figura 2-9 Transceptor AIS [23]	22
Figura 3-1 Interfaz del navegador de Anaconda	27
Figura 3-2 Interfaz de celdas de Jupyter Notebook.....	28
Figura 3-3 Ejemplo de visualización de mapa en Kepler.gl [34]	29
Figura 3-4 Distribución de puntos concentrada en latitudes altas.....	31
Figura 3-5 Comparación de distribución corregida por latitud	31
Figura 3-6 Representación en 3D de distribución corregida por latitud	32
Figura 3-7 Generación de puntos aleatorios sólo en agua.....	32
Figura 3-8 Distribución de distancia para hexágono perfecto de lado 1	34
Figura 3-9 Distancia de puntos dentro de un hexágono perfecto	34
Figura 3-10 Distribución de error según resolución 6.....	36
Figura 3-11 Distribución de error según resolución 14.....	37
Figura 3-12 Diferencia de error para dos resoluciones	38
Figura 3-13 Gráfica de error por resolución.....	39
Figura 3-14 Distribución de error de posicionamiento para un área	40
Figura 3-15 Distribución de error de posicionamiento irregular en un área	41
Figura 3-16 Comparativa de error de posición entre H3 y S2.....	42
Figura 3-17 Representación de error mínimo y máximo para un disco de "k" celdas	43
Figura 3-18 Área cubierta con k anillos y después filtrada con celdas cuyo centro está en el interior del círculo.....	45
Figura 3-19 Representación de cobertura para el mismo área para resoluciones 5, 7 y 10.....	45
Figura 3-20 Gráfica área cubierta y número de celdas por resolución.....	46
Figura 3-21 Representación de área cubierta en resoluciones bajas	47
Figura 3-22 Representación de la <i>bounding box</i> para los límites de integración.....	48

Figura 3-23 Gráfica con los resultados de cobertura según resolución.....	50
Figura 3-24 Ejemplo de trayectoria a través de celdas.....	51
Figura 3-25 Representación de la misma trayectoria para las resoluciones 4, 7 y 10.....	52
Figura 3-26 Representación de trayectoria mediante <i>grid</i> y transmisiones	55
Figura 3-27 Ruta con transmisiones en resoluciones medias.....	57
Figura 3-28 Ruta mediante transmisiones en resoluciones muy altas.....	58
Figura 3-29 Gráfica de los resultados de la Tabla 3-7.....	58
Figura 4-1 Resumen de resultados obtenidos.....	62
Figura 4-2 Distorsión del sistema en base al error de posición.....	64

ÍNDICE DE TABLAS

Tabla 2-1 Cantidad de celdas por resolución [3].....	18
Tabla 2-2 Área y longitud media de lado por resolución [3]	19
Tabla 2-3 Valores máximo y mínimo de área por resolución [3].....	20
Tabla 2-4 Área y número de celdas por nivel de S2 [22].....	21
Tabla 2-5 Comparativa de funcionalidad según la clase [25]	23
Tabla 2-6 Intervalos de transmisión de datos [27]	23
Tabla 3-1 Librerías empleadas	28
Tabla 3-2 Resultados de error de posición por resolución	39
Tabla 3-3 Comparación de precisión entre H3 y S2	42
Tabla 3-4 Comparación de geometría entre hexágono y cuadrado	43
Tabla 3-5 Cobertura de áreas por resolución.....	49
Tabla 3-6 Porcentajes de cobertura de área por resolución	49
Tabla 3-7 Análisis de una trayectoria concreta	53
Tabla 3-8 Análisis de 100000 trayectorias aleatorias de 500 km.....	54
Tabla 3-9 Resultados de transmisiones por celdas	56

de grandes volúmenes de datos generados por el AIS requiere metodologías eficientes que permitan almacenar, indexar y recuperar información de manera rápida y precisa. En este sentido los sistemas tradicionales de almacenamiento y análisis de datos espaciales enfrentan limitaciones en términos de escalabilidad y rendimiento, especialmente cuando se trabaja con bases de datos distribuidas o infraestructuras en la nube. Estas limitaciones evidencian la necesidad de adoptar enfoques innovadores que ofrezcan una representación estructurada y eficiente de los datos geoespaciales, optimizando así las operaciones de consulta y análisis.

Los sistemas de indexación geoespacial son esenciales en la gestión y el análisis de datos multidimensionales, especialmente en aplicaciones que requieren una representación eficiente del espacio geográfico.

Un aspecto clave en la indexación geoespacial es la forma en la que se representan los datos espaciales. En lugar de trabajar con coordenadas bidimensionales exactas en un plano cartesiano o en un sistema de referencia geodésico, las técnicas de indexación convierten el espacio en celdas discretas que agrupan puntos cercanos en una estructura jerárquica o en una partición uniforme del espacio. Estas celdas se representan mediante un identificador unidimensional que se puede indexar. De esta manera, estos sistemas permiten organizar, recuperar y procesar grandes volúmenes de información espacial de manera eficiente, facilitando la ejecución de consultas espaciales y espaciotemporales con alto rendimiento.

En este contexto, el sistema Uber H3 [3] es una tecnología de indexación geoespacial basada en una jerarquía de celdas hexagonales (Figura 1-2) que ha surgido como una herramienta prometedora. Este sistema divide la superficie terrestre en celdas uniformes de diferentes resoluciones. En comparación con las cuadrículas tradicionales, el diseño hexagonal de H3 minimiza distorsiones geográficas, ofrece flexibilidad en la resolución espacial y facilita operaciones como la agregación de datos y la búsqueda geoespacial.

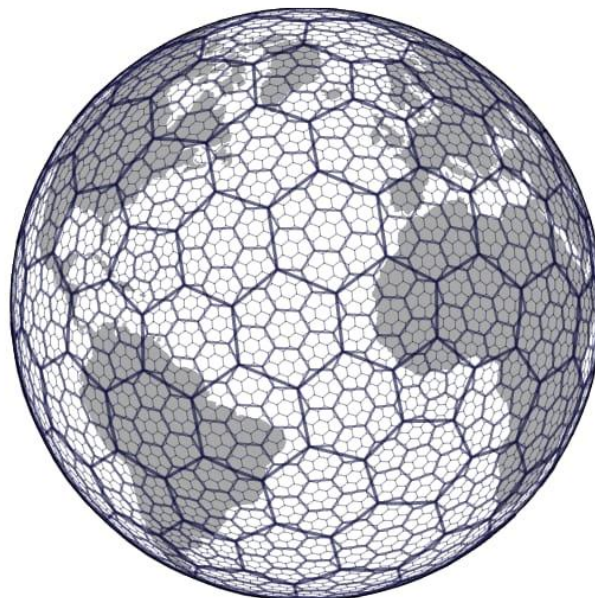


Figura 1-2 Separación espacial Hexagonal de Uber H3 [3]

La indexación basada en celdas presenta diversas ventajas que la hacen especialmente útil. Una de sus principales fortalezas es la optimización del rendimiento ya que, al reducir la precisión de los datos mediante una representación discreta, se simplifican los cálculos y se mejora la eficiencia en la ejecución de consultas espaciales, como la búsqueda de vecinos cercanos o la agregación de información en áreas geográficas específicas. No obstante, su principal ventaja reside en convertir una búsqueda bidimensional (de latitud y longitud) a una unidimensional basada en celdas, que puede ordenarse o indexarse para acelerar las búsquedas. Además, esta técnica favorece la escalabilidad ya que, al dividir

el espacio en celdas predefinidas, permite gestionar grandes cantidades de datos en infraestructuras distribuidas sin que ello afecte de manera significativa en el rendimiento del sistema. Asimismo, su compatibilidad con modelos de análisis geoespacial la convierte en una herramienta valiosa para diversas aplicaciones, incluyendo la visualización de mapas, el análisis de tráfico y el modelado de patrones de movilidad. Por otro lado, pero no menos importante, la inteligencia artificial (IA) desempeña un papel clave en la optimización del procesamiento de datos AIS, facilitando la detección de patrones de tráfico, la predicción de rutas y la identificación de anomalías en la navegación. Sin embargo, para integrar estos datos en modelos de aprendizaje automático, es crucial reducir la cantidad de información procesada sin comprometer la precisión. En este sentido, la conversión de coordenadas geográficas en celdas discretas mediante Uber H3 permite reducir el número de tokens, optimizando la carga computacional y mejorando la eficiencia en el análisis en tiempo real. Esta representación compacta y estructurada facilita la integración con modelos de IA, permitiendo una navegación más precisa y escalable [4].

Además, la aplicación de técnicas avanzadas de indexación geoespacial en el tráfico marítimo no solo tiene implicaciones en la seguridad y eficiencia operativa, sino que también impacta en ámbitos como la sostenibilidad y la optimización de rutas. La capacidad de procesar datos en tiempo real con mayor precisión facilita la identificación de patrones de navegación, la predicción de congestión en rutas marítimas y la detección de posibles incidentes en zonas críticas. Este enfoque puede contribuir significativamente a la reducción del consumo de combustible mediante la optimización de trayectorias, lo que a su vez disminuye la huella de carbono de la industria marítima. En este contexto, explorar el uso de H3 como una solución de indexación geoespacial aplicada al AIS es un paso clave para avanzar en la modernización del sector y fomentar el desarrollo de herramientas de gestión más robustas y escalables.

Por todo ello, la implementación de H3 en el procesado de datos AIS puede transformar significativamente la forma en que se procesan y visualizan los datos marítimos. Al permitir un procesamiento más eficiente y escalable, este sistema podría facilitar tareas críticas como la detección de patrones de tráfico, la identificación de comportamientos anómalos y la evaluación de rutas marítimas, contribuyendo al fortalecimiento de la seguridad y la eficiencia en el sector marítimo.

A pesar de sus numerosas ventajas, la indexación geoespacial basada en celdas introduce ciertas imprecisiones que deben ser consideradas para su uso adecuado. La principal limitación de este enfoque radica en la aproximación de una ubicación específica mediante una celda discreta, lo que conlleva un error de cuantificación, es decir, una pérdida de precisión en la representación del dato original. Esta imprecisión puede tener implicaciones importantes en distintos contextos como, por ejemplo, el rumbo y la velocidad real que lleva un barco comparado con los obtenidos a través de celdas. En aplicaciones de navegación y geolocalización, si la granularidad de las celdas es demasiado grande, la precisión en la determinación de ubicaciones puede resultar insuficiente para casos en los que se requiere alta exactitud, como en la conducción autónoma o la localización de dispositivos móviles. En el análisis de datos espaciales, el nivel de agregación seleccionado puede influir en los resultados de las consultadas espaciales, generando desviaciones en la interpretación de los datos.

Por ello, antes de adoptar plenamente este enfoque, es esencial llevar a cabo un análisis detallado, un estudio de viabilidad y una evaluación de su aplicabilidad a los datos AIS. Por tanto, este trabajo de fin de grado tiene como principal motivación explorar el potencial del sistema Uber H3 para optimizar el procesamiento de los datos AIS, contribuyendo al desarrollo de herramientas más avanzadas para gestión del tráfico marítimo y fortaleciendo la seguridad y eficiencia en dicho ámbito.

1.2 Objetivos

El objetivo principal de este trabajo de fin de grado es llevar a cabo un análisis exhaustivo y una evaluación detallada del sistema de indexación espacial Uber H3, con el fin de determinar su idoneidad y ventajas en el contexto de procesamiento y análisis de datos generados por el AIS. Esta evaluación

implica comparar el rendimiento, la eficacia y las capacidades técnicas de Uber H3 frente a otros enfoques alternativos de indexación y discretización espacial, como Google S2, que se utilizan comúnmente en el tratamiento de información geoespacial. Derivado del objetivo principal del proyecto que es analizar los errores que se cometen al analizar posiciones y trayectorias AIS usando celdas Uber H3, se proponen diversos objetivos específicos.

Si bien se incluye un estado del arte para contextualizar el estudio y proporcionar los fundamentos teóricos necesarios, el enfoque central recae en el análisis práctico del sistema a través de una serie de experimentos. Estos experimentos permiten valorar su eficacia en distintas aplicaciones geoespaciales, especialmente en lo relativo al posicionamiento, la cobertura de áreas y la representación de trayectorias. A continuación, se enumeran los objetivos específicos que guían el desarrollo del estudio:

- Estado del arte: Revisar los conceptos fundamentales relacionados con las proyecciones cartográficas y los sistemas de indexación geoespacial, con especial atención al sistema Uber H3 y, de forma secundaria, al sistema Google S2, así como analizar las características de los datos AIS que justifican la necesidad de herramientas eficientes para su procesamiento.
- Análisis de posición: Evaluar el error que se comete al representar puntos geográficos mediante celdas H3, analizando la distancia entre la posición real y el centroide de la celda asignada, y cuantificando este error en función de la resolución.
- Análisis de área: Examinar la capacidad del sistema H3 para cubrir áreas geográficas específicas mediante celdas, comparando el área teórica con la efectivamente cubierta y valorando la precisión y eficiencia espacial alcanzadas.
- Análisis de trayectoria: Analizar cómo se representan trayectorias mediante el sistema H3, evaluando la distorsión en términos de rumbo y distancia, observando el impacto de la discretización espacial en la fidelidad de la trayectoria real y el análisis de dichas trayectorias en función de la frecuencia de transmisión de los datos AIS.

1.3 Estructura de la memoria

Después de contextualizar el trabajo y definir los objetivos a alcanzar, se expone la organización de la memoria con el propósito de proporcionar una estructura más clara. El documento se compone de cinco capítulos, seguidos de la bibliografía y los anexos.

- Capítulo 1: Introducción y objetivos. En el primer capítulo se introduce el trabajo. En esta introducción se menciona por primera vez los sistemas de indexación geoespacial, así como, los datos AIS y su posible relación.
- Capítulo 2: Estado del arte. En este capítulo se establecen las bases teóricas generales y específicas, pasando a analizar detalladamente lo siguiente:
 - Las proyecciones cartográficas. Se hace una breve reseña histórica de las proyecciones cartográficas. Posteriormente una clasificación para terminar con los sistemas más utilizados en la actualidad.
 - Análisis de sistemas de indexado geoespacial. Se define y analiza detalladamente el sistema desarrollado por Uber. A continuación, una breve descripción del sistema de Google.
 - AIS. Se define el sistema a través de los tipos de transmisiones y mensajes y a través de la calidad y consistencia de los datos.
 - Trabajos previos. Se mencionan y explican brevemente estudios o trabajos relacionados con el ámbito que aborda el presente proyecto.
- Capítulo 3: Desarrollo. En este capítulo se detalla el desarrollo del trabajo. Se describen las herramientas utilizadas como el entorno de trabajo, el lenguaje de programación y una de las metodologías como es la simulación de Monte Carlo. Se profundiza en los análisis de error de posición, área y trayectoria.

- Capítulo 4: Resultados. En este cuarto capítulo se exponen los resultados obtenidos en el capítulo anterior.
- Capítulo 5: Conclusiones y líneas futuras. En estos capítulos se concluye el trabajo y se presentan las posibles líneas futuras en el ámbito para una integración total entre ambos sistemas.
- Bibliografía y anexos.

2 ESTADO DEL ARTE

El presente capítulo comienza con el estudio de las proyecciones cartográficas utilizadas para representar sobre un plano la superficie esférica de la Tierra, una problemática clave en disciplinas geoespaciales. A continuación, se analizan diversos sistemas de indexación geoespacial, centrándose en el sistema H3 desarrollado por Uber, reconocido por su eficiencia en la gestión y representación jerárquica de datos espaciales. Posteriormente, se aborda el sistema AIS, examinando sus diversos campos de información, la calidad de sus datos y las limitaciones que presenta en contextos específicos. Finalmente, se revisan diversos proyectos previos que han tratado objetivos similares, proporcionando un marco de referencia para comprender las contribuciones existentes y los retos actuales en esta área.

2.1 Las proyecciones cartográficas

Las proyecciones cartográficas son métodos matemáticos usados para representación de la superficie curva de la Tierra, considerada tanto esférica como elipsoidal, en un plano bidimensional. Debido a la imposibilidad de desplegar una esfera en un plano sin producir distorsiones, las proyecciones buscan minimizar y controlar estas deformaciones para preservar ciertas propiedades. Esta transformación es esencial para crear mapas, ya que permite la visualización y el análisis de datos geográficos. La elección de la proyección afecta a la representación de distancias, áreas y ángulos, dando lugar a diversas alteraciones que los cartógrafos deben tener en cuenta a la hora de seleccionar la proyección adecuada para aplicaciones específicas.

El proceso de proyección (Figura 2-1) implica transformar coordenadas geográficas (latitud y longitud) en coordenadas cartesianas (X, Y). Para esta transformación existen diferentes tipos de proyecciones basadas en minimización de las imprecisiones en distintas propiedades [5].

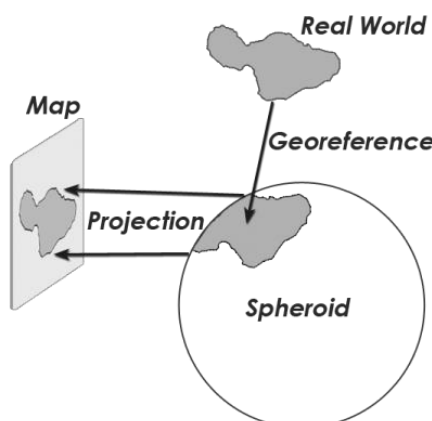


Figura 2-1 Esquema básico de idea de representación [6]

Además, en el contexto de los Sistemas de Información Geográfica (SIG), las proyecciones cartográficas desempeñan un papel crucial en los análisis espaciales y la visualización de datos geográficos. La elección de la proyección adecuada es fundamental para garantizar la correcta superposición de capas de datos provenientes de diferentes fuentes [7].

2.1.1 Antecedentes

Las proyecciones cartográficas tienen una larga historia que se remonta a la antigüedad. Su desarrollo ha sido fundamental para la evolución de la cartografía y la comprensión del mundo. La necesidad de representar la Tierra en superficies planas surgió en las primeras civilizaciones. Los antiguos griegos, con su avanzado conocimiento matemático, fueron pioneros en este ámbito. Claudio Ptolomeo, en el siglo II d. C., sentó las bases de la cartografía científica al definir proyecciones cartográficas y representar elementos geográficos mediante coordenadas en su obra *Geographia* [8].

Durante la Edad Media, el desarrollo de las proyecciones se estancó en Europa, pero continuó en el mundo islámico. Por tanto, se puede afirmar que en el siglo IX la cartografía islámica, al igual que su sociedad, vive una situación de esplendor. Es a través de los mapamundis de la época cómo demuestra el Islam su gran poder e influencia. Algunas de las características de estos llamados mapamundis son: el Sur en la parte superior (Figura 2-2); centrados, principalmente, en Bagdad; el mar Mediterráneo de gran tamaño, superior al real; y, por último, la representación disminuida del territorio ocupado por el Imperio de Roma [9].



Figura 2-2 Mapamundi Al Idrisi, 1160 [10]

Más tarde, en el Renacimiento, debido al interés por la navegación y a descubrimientos geográficos, hubo un interés renovado por la cartografía. Hitos como la circunnavegación de África o el descubrimiento de América hicieron que fuera necesario la incorporación de las recientes tierras en las representaciones de los correspondientes imperios. Es en esta época cuando el matemático y geógrafo, natural de Flandes, Gerardus Mercator desarrolla una de las técnicas más utilizadas hoy en día para la ilustración de la realidad sobre un plano: la proyección Mercator [11].

El siglo XIX vio un aumento en la diversidad de proyecciones, cada una diseñada para minimizar ciertas distorsiones o satisfacer necesidades específicas. Ya en el siglo XX, con el advenimiento de la era digital, las proyecciones experimentaron una revolución. Es a la llegada de los SIG y la cartografía digital cuando se ha transformado radicalmente el campo de las proyecciones. Estas herramientas han permitido una mayor flexibilidad en la creación y uso de diferentes proyecciones, facilitando el análisis espacial y la visualización de datos geográficos.

2.1.2 Clasificación

Las proyecciones cartográficas permiten transformar la superficie curva de la Tierra en un plano, y su clasificación se realiza según diversos criterios. Uno de los más relevantes es el basado en las propiedades métricas que conservan, como las proyecciones conformes (preservan ángulos), equivalentes (mantienen áreas) o equidistantes (reproducen distancias desde ciertos puntos). También se

distinguen por la geometría utilizada, como las proyecciones cilíndricas, cónicas y acimutales, cada una adaptada a diferentes usos y escalas [12].

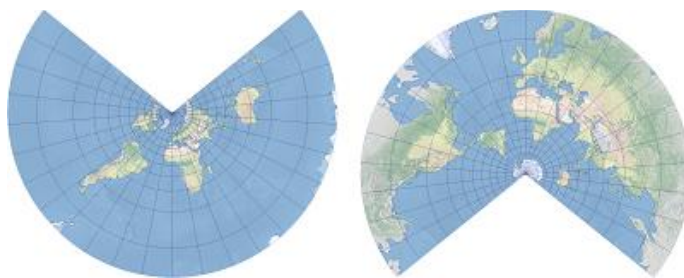


Figura 2-3 Proyección conforme cónica [13]



Figura 2-4 Proyección equivalente cilíndrica [14]

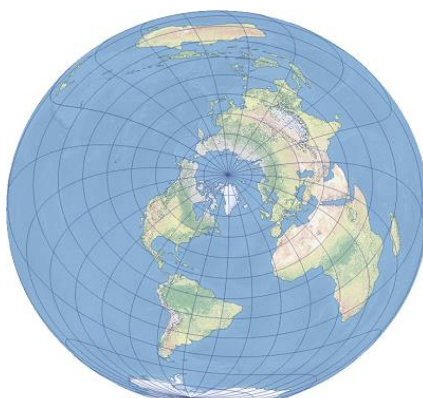


Figura 2-5 Proyección equidistante [15]

2.1.3 Sistemas de proyección cartográfica

Como se ha mencionado en párrafos anteriores de este mismo capítulo, los sistemas de proyección cartográfica son esenciales para transformar la superficie tridimensional de la Tierra en una representación plana. A continuación, se detallan algunos de los sistemas de proyección y datums más relevantes en la actualidad.

2.1.3.1 World Geodetic System 1984

El WGS84 (*World Geodetic System 1984*) es el sistema de referencia geodésico más utilizado a nivel global, especialmente en sistemas de navegación por satélite como el GPS. Este sistema incluye un datum geodésico y un sistema de coordenadas que se basa en un elipsoide con un radio ecuatorial de 6378137 metros y emplea un sistema de coordenadas cartesianas tridimensionales (X, Y, Z). Su precisión lo convierte en la referencia estándar para aplicaciones militares, civiles, de navegación global y en sistemas de información geográfica [16].

2.1.3.2 European Terrestrial Reference System 1989

El ETRS89 (*European Terrestrial Reference System 1989*) es el sistema de referencia geodésico utilizado en Europa, diseñado para estar fijo en la placa tectónica europea. Aunque está alineado con el

WGS84, el ETRS89 evita los desplazamientos de coordenadas causados por los movimientos tectónicos, proporcionando una mayor estabilidad en aplicaciones de alta precisión, como la cartografía europea y los estudios de deformación de la corteza terrestre. Se utiliza comúnmente en combinación con proyecciones como el sistema UTM [17].

2.1.3.3 Universal Transverse Mercator

El UTM (*Universal Transverse Mercator*) divide el mundo en 60 zonas de 6° de longitud cada una, empleando una proyección conforme de Mercator transversa. Este sistema mantiene los ángulos y las formas locales, lo que lo convierte en una opción adecuada para la cartografía a gran escala. Cada zona tiene su propio sistema de coordenadas, generalmente basado en el elipsoide WGS84, aunque puede adaptarse a otros datums. Se utiliza ampliamente en aplicaciones militares, ingeniería y sistemas de información geográfica debido a su precisión y facilidad de uso [18].

2.1.3.4 Proyección de Mercator

La proyección de Mercator es una proyección cilíndrica conforme que mantiene las formas de los objetos geográficos, pero distorsiona las áreas, especialmente cerca de los polos. Es particularmente útil en la navegación marítima, ya que las líneas de rumbo constante (loxodromias) se representan como líneas rectas, lo que facilita la planificación de rutas. Aunque no es adecuada para representar áreas polares debido a su alta distorsión, esta proyección es muy utilizada en aplicaciones de mapas web interactivos, como Google Maps en niveles bajos de zoom [19].

2.1.3.5 Proyección de Lambert Conformal Conic

La proyección de *Lambert Conformal Conic* es una proyección cónica conforme ideal para representar regiones extensas con orientación este-oeste. Mantiene las formas locales y presenta una mínima distorsión en latitudes intermedias. Por esta razón, es ampliamente utilizada en cartografía aeronáutica y en mapas regionales de Europa y América del Norte, donde la precisión en la representación de las formas y la orientación es fundamental.

2.1.3.6 North American Datum 1983

El NAD83 (*North American Datum 1983*) es un datum geodésico utilizado principalmente en América del Norte. Está estrechamente relacionado con el WGS84, pero se encuentra fijado a la placa tectónica de América del Norte para minimizar los desplazamientos causados por el movimiento tectónico. Es compatible con diversos sistemas de proyección, como el UTM, y es ampliamente empleado en la cartografía de Estados Unidos, Canadá y México, así como en estudios de movimientos tectónicos [20].

2.1.4 Problemas inherentes

Uno de los principales problemas de las proyecciones cartográficas es la distorsión de la superficie terrestre, que puede dar lugar a ideas erróneas sobre el tamaño de los accidentes geográficos. Battersby [19] subraya que las personas suelen tener dificultades para comprender las implicaciones de las distintas proyecciones conocidas como la Mercator a otras menos comunes. Esta confusión puede dificultar la lectura e interpretación eficaces de los mapas.

La base matemática de las proyecciones cartográficas también contribuye a sus problemas inherentes. Por ejemplo, la elección de la proyección a menudo implica compromisos entre preservar el área, la forma y la distancia. Ábrahámová y Vájsáblova discuten cómo los métodos matemáticos alternativos pueden optimizar las distorsiones en las proyecciones, aunque estos métodos no pueden eliminar la distorsión por completo [21].

En conclusión, los problemas inherentes a las proyecciones cartográficas se derivan de su capacidad para distorsionar la información geográfica, lo que puede inducir a error a los usuarios y complicar la interpretación de los datos espaciales. La elección de la proyección es fundamental ya que afecta no sólo

a la precisión del mapa, sino que también a la comprensión cognitiva de los usuarios. A medida que la cartografía sigue evolucionando, especialmente con el auge de las tecnologías cartográficas digitales, abordar estas cuestiones sigue siendo un reto importante tanto para los cartógrafos como para los educadores.

2.2 Sistemas de indexado geoespacial

Para entender el desarrollo del presente trabajo es necesario hacer un análisis de algunos de los sistemas de indexación geoespacial. Una proyección cartográfica transforma la superficie curva de la Tierra en un plano, pero es la cuadrícula jerarquizada de un sistema de indexación lo que permite estructurar eficientemente el espacio representado. Mientras la proyección sitúa los datos geoespaciales en un marco bidimensional, la indexación los organiza en unidades discretas, optimizando la consulta, el análisis y la agregación de información a múltiples escalas. Se valúan sus ventajas, limitaciones y aplicaciones, proporcionando una visión comparativa para entender su impacto en la gestión eficiente de datos geoespaciales.

2.2.1 Uber H3

Este sistema desarrollado por Uber requiere, al igual que muchos otros, una cuadrícula superpuesta sobre la superficie terrestre para organizar datos espaciales. La indexación jerárquica hexagonal Uber H3 se basa en una proyección gnómica centrada en las caras de un icosaedro (Figura 2-6), un poliedro compuesto por veinte caras. En lugar de representar la Tierra como una esfera, este modela la conceptualiza como un icosaedro, sobre cuyas caras se establece una cuadrícula geodésica discreta nivel global.

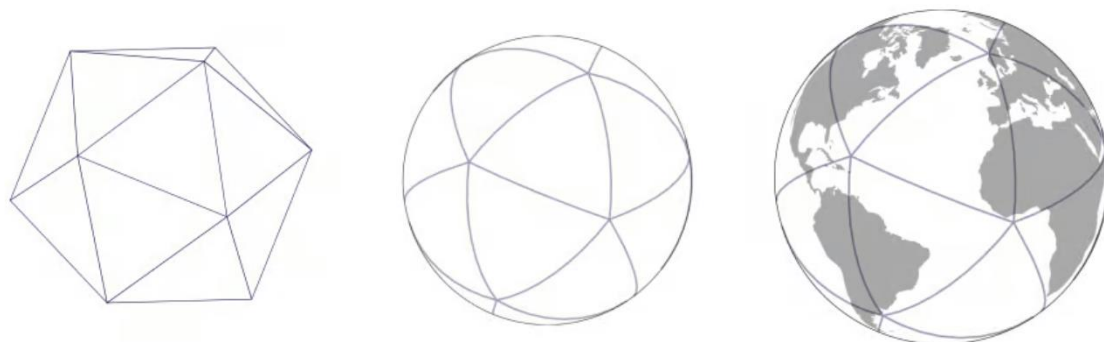


Figura 2-6 Icosaedro terrestre [3]

El sistema Uber H3 ofrece un total de 16 resoluciones de 0 a 15. La Tabla 2-1 representa la cantidad de los hexágonos en cada resolución. En la resolución más baja, la superficie terrestre es segmentada en 110 hexágonos y 12 pentágonos. A medida que la resolución aumenta, cada hexágono se subdivide en siete hexágonos más pequeños, lo que permite una mayor precisión en la representación espacial. Cuanto más alto sea el nivel de resolución, menor será el tamaño de los hexágonos proyectados sobre la superficie del planeta, mejorando así la granularidad del modelo.

Uno de los motivos por los que este sistema está basado en hexágonos y la empresa Uber lo remarca como un aspecto a destacar, es la capacidad equidistante a las celdas vecinas que tiene la geometría hexagonal frente a la triangular o cuadrada como se puede ver en la Figura 2-7.

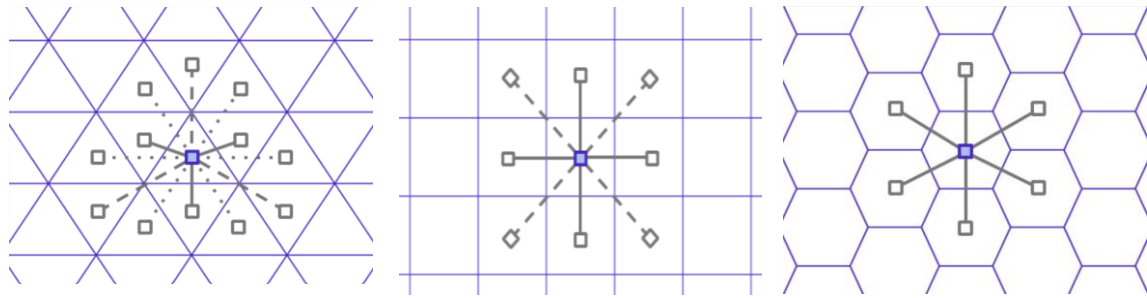


Figura 2-7 Distancia a celdas vecinas de lados y vértices [3]

Como se observa en la Figura 2-7 en la geometría triangular, la celda central tiene doce celdas vecinas con tres distancias distintas. Por otro lado, los cuadrados tienen ocho celdas vecinas con dos distancias distintas. Y, por último, los hexágonos tienen seis celdas vecinas equidistantes. Esta característica facilita y permite el análisis, expandiendo la geometría en anillos de celdas equidistantes al centro.

Resolución	Nº total de celdas	Nº de hexágonos	Nº de pentágonos
0	122	110	12
1	842	830	12
2	5882	5870	12
3	41162	41150	12
4	288122	288110	12
5	2016842	2016830	12
6	14117882	14117870	12
7	98825162	98825150	12
8	691776122	691776110	12
9	4842432842	4842432830	12
10	33897029882	33897029870	12
11	237279209162	237279209150	12
12	1660954464122	1660954464110	12
13	11626681248842	11626681248830	12
14	81386768741882	81386768741870	12
15	569707381193162	569707381193150	12

Tabla 2-1 Cantidad de celdas por resolución [3]

Además de representar la cantidad de celdas por cada resolución, resulta relevante analizar la distribución de las áreas de los hexágonos en comparación con las de los pentágonos dentro del sistema de indexación H3. Dado que la Tierra no puede ser completamente cubierta con hexágonos debido a restricciones geométricas, se requiere la inclusión de pentágonos, los cuales se ubican estratégicamente en los vértices del icosaedro, generalmente coincidiendo en regiones oceánicas. Este aspecto es fundamental, ya que los pentágonos poseen características únicas que los diferencian de los hexágonos, especialmente en términos de su impacto en la indexación y en la agregación de datos geoespaciales. Esta característica del sistema puede ser un inconveniente en el caso de tratar con datos oceánicos, pero como se observa en la Tabla 2-1 la cantidad de pentágonos es muy inferior respecto al total de celdas.

Desde el punto de vista de la distribución del área, los pentágonos de una misma resolución mantienen un tamaño homogéneo, lo que facilita su identificación y procesamiento dentro de los algoritmos de indexación. Sin embargo, el área de los hexágonos presenta variaciones dependiendo de

su posición relativa con respecto a los vértices del icosaedro. Este comportamiento afecta la manera en que se realizan cálculos espaciales, especialmente en aplicaciones que requieren una alta precisión, como la interpolación de datos o la estimación de densidades poblacionales. La capacidad de representar el espacio terrestre con una estructura adaptable, en la que cada celda tiene una relación directa con su entorno, permite que H3 optimice procesos de agrupamiento y análisis geoespacial, facilitando la manipulación eficiente de grandes volúmenes de información. Véase la Tabla 2-2, donde se expone el ratio entre el tamaño del pentágono y el área media para una resolución.

Res	Área media hexágono (km ²)	Longitud media lado (km)	Área pentágono (km ²)	Ratio (P/H)
0	4357449,416078381	1281,256011	2562182,162955496	0,5880
1	609788,441794133	483,0568391	328434,586246469	0,5386
2	86801,780398997	182,5129565	44930,898497879	0,5176
3	12393,434655088	68,97922179	6315,472267516	0,5096
4	1770,347654491	26,07175968	896,582383141	0,5064
5	252,903858182	9,854090990	127,785583023	0,5053
6	36,129062164	3,724532667	18,238749548	0,5048
7	5,161293360	1,406475763	2,604669397	0,5047
8	0,737327598	0,531414010	0,372048038	0,5046
9	0,105332513	0,200786148	0,053147195	0,5046
10	0,015047502	0,075863783	0,007592318	0,5046
11	0,002149643	0,028663897	0,001084609	0,5046
12	0,000307092	0,010830188	0,000154944	0,5046
13	0,000043870	0,004092010	0,000022135	0,5046
14	0,000006267	0,001546100	0,000003162	0,5046
15	0,000000895	0,000584169	0,000000452	0,5046

Tabla 2-2 Área y longitud media de lado por resolución [3]

Otro de los factores importantes a destacar es que, como se ha mencionado anteriormente, el área de las celdas varía en función de su ubicación relativa respecto a los vértices del icosaedro. Esta variabilidad introduce una diferencia significativa en la resolución de los datos geoespaciales, ya que ciertas regiones pueden experimentar una mayor densidad de celdas en comparación con otras. La relación entre el área máxima y mínima de las celdas hexagonales en cada resolución permite comprender el grado de distorsión que puede producirse al emplear este sistema de indexación. En términos prácticos, esta variabilidad afecta a la precisión con la que se pueden modelar fenómenos espaciales, especialmente en aplicaciones que requieren cálculos basados en la homogeneidad de las celdas, como la asignación de recursos o la detección de patrones en datos geoespaciales.

La diferencia entre los valores máximo y mínimo del área de las celdas es un factor determinante en la precisión de los modelos que dependen de la distribución espacial uniforme de los datos. En la Tabla 2-3 se muestran los valores máximos y mínimos de área junto al ratio para cada resolución. Si bien el ratio entre el área máxima y mínima en las primeras resoluciones es considerablemente alto, este valor se estabiliza en niveles más finos, lo que sugiere que las celdas más pequeñas presentan una distribución más homogénea en comparación con aquellas de menor resolución. Esta propiedad es clave para aplicaciones que requieren un nivel de granularidad controlado, como la planificación urbana, el monitoreo ambiental y la optimización de redes de telecomunicaciones.

Res	Mín Hexágono área (km2)	Máx hexágono área (km2)	Ratio (máx/mín)
0	4106166,334463915	4977807,027442012	1,212276
1	447684,201817940	729486,875275344	1,629468
2	56786,622889474	104599,807218925	1,841980
3	7725,505769639	14950,773301379	1,935248
4	1084,005635363	2135,986983965	1,970457
5	153,766244448	305,144308779	1,984469
6	21,910021013	43,592111685	1,989597
7	3,126836030	6,227445905	1,991613
8	0,446526174	0,889635157	1,992347
9	0,063780227	0,127090737	1,992635
10	0,009110981	0,018155820	1,992740
11	0,001301542	0,002593689	1,992782
12	0,000185933	0,000370527	1,992797
13	0,000026562	0,000052932	1,992802
14	0,000003795	0,000007562	1,992805
15	0,000000542	0,000001080	1,992805

Tabla 2-3 Valores máximo y mínimo de área por resolución [3]

Uber ha desarrollado una librería para H3 y la ha puesto a disposición de la comunidad en GitHub con una licencia libre (Apache 2.0). La librería proporciona funciones específicas para el análisis de datos espaciales; desde convertir coordenadas de latitud y longitud a la celda H3 correspondiente, determinar el centro de una celda, obtener la geometría del contorno, identificar celdas vecinas hasta hacer agrupar diferentes niveles de celdas de una resolución para cubrir un área determinada [3]. Estas funciones y otras más concretas se emplearán a la hora de desarrollar los análisis del presente trabajo.

2.2.2 Google S2

A pesar de que el presente trabajo se centra en la tecnología de Uber, existen alternativas como la empresa Google que, por su parte, también dispone de su sistema de indexación geoespacial denominado S2. Su nombre viene dado por la notación de unidades esféricas (s^2). Este sistema representa los datos en celdas cuadrangulares en una esfera tridimensional, a diferencia de otros sistemas que lo hacen sobre un plano bidimensional. Esta característica hace que las distorsiones en las representaciones de los datos sean muy similares en cualquier parte de la Tierra y poco significativos.

El sistema de indexación geoespacial S2 desarrollado por Google emplea una estructura jerárquica de celdas combinada con la curva de Hilbert para mejorar la eficiencia en la representación y recuperación de datos espaciales. La jerarquía de celdas en S2 se basa en una subdivisión de la esfera terrestre en seis caras de un cubo proyectado, las cuales a su vez se dividen de manera recursiva en celdas más pequeñas, manteniendo una relación de contención jerárquica entre niveles de resolución (Figura 2-8). Este enfoque permite realizar consultas espaciales a diferentes escalas de manera eficiente, ya que es posible agrupar regiones dentro de un mismo nivel jerárquico y acceder a áreas geográficas específicas con una granularidad controlada. Sin embargo, la disposición de estas celdas en una estructura jerárquica conlleva un desafío en términos de coherencia espacial y continuidad, pues las celdas adyacentes pueden terminar teniendo identificadores dispersos en el espacio de direcciones, lo que podría generar ineficiencias en el almacenamiento y en la búsqueda de datos.

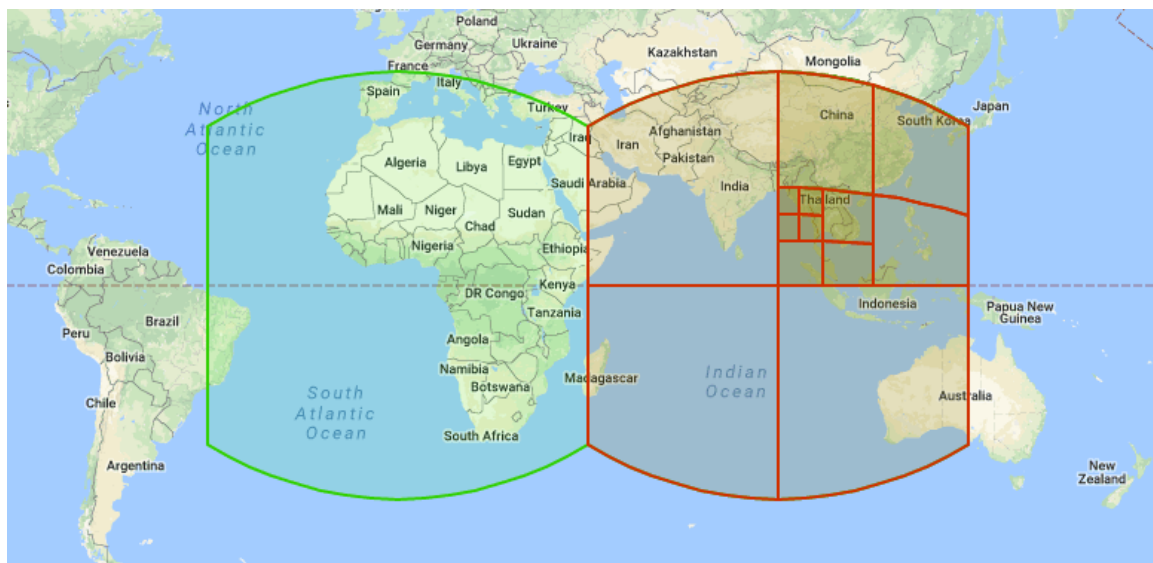


Figura 2-8 División jerárquica de S2 [22]

Para abordar este problema, S2 utiliza la curva de Hilbert, una curva de llenado de espacio que preserva la proximidad espacial en una representación unidimensional. Esta técnica permite asignar un identificador único de sesenta y cuatro bits a cada celda de manera que aquellas geográficamente cercanas conserven una relación de vecindad en el espacio de direcciones. La combinación de la jerarquía de celdas con la curva de Hilbert ofrece un sistema de indexación altamente escalable para operaciones geoespaciales, ya que permite realizar búsquedas de proximidad con menor latencia y organizar datos espaciales en estructuras de almacenamiento compactas. Por esto, S2 es particularmente útil en aplicaciones que requieren acceso rápido a grandes volúmenes de información geográfica, como sistemas de navegación, análisis de cobertura de redes y gestión de infraestructuras geolocalizadas.

La Tabla 2-4 muestra los niveles de resolución de Google S2 junto al área media y número de celdas.

Nivel	Área media	Unidades	Número de celdas	Nivel	Área media	Unidades	Número de celdas
00	85011012,19	km ²	6	15	79172,67	m ²	6B
01	21252753,05	km ²	24	16	19793,17	m ²	25B
02	5313188,26	km ²	96	17	4948,29	m ²	103B
03	1328297,07	km ²	384	18	1237,07	m ²	412B
04	332074,27	km ²	1536	19	309,27	m ²	1649B
05	83018,57	km ²	6K	20	77,32	m ²	7T
06	20754,64	km ²	24K	21	19,33	m ²	26T
07	5188,66	km ²	98K	22	4,83	m ²	105T
08	1297,17	km ²	393K	23	1,21	m ²	422T
09	324,29	km ²	1573K	24	0,30	m ²	1689T
10	81,07	km ²	6M	25	755,05	cm ²	7e15
11	20,27	km ²	25M	26	188,76	cm ²	27e15
12	5,07	km ²	100M	27	47,19	cm ²	108e15
13	1,27	km ²	402M	28	11,80	cm ²	432e15
14	0,32	km ²	1610M	29	2,95	cm ²	1729e15
15	79172,67	m ²	6B	30	0,74	cm ²	7e18

Tabla 2-4 Área y número de celdas por nivel de S2 [22]

2.3 AIS

El *Automatic Identification System* (AIS), Sistema de Identificación Automática en español, es un sistema que proporciona información y datos sobre las características más significativas de los buques en la mar. El propósito originario de dicho sistema era evitar colisiones y para ello los usuarios podían ver información tan relevante como pudiera ser la posición, rumbo o destino entre otros.



Figura 2-9 Transceptor AIS [23]

En la actualidad y desde el 2002, esta tecnología es obligatoria para las embarcaciones de más de 300 toneladas brutas y buques de pasajeros debido al cumplimiento del convenio internacional para la seguridad de la vida humana en la mar (SOLAS) exigido por la *International Maritime Organisation* (IMO) [1].

2.3.1 Tipos de transpondedores y mensajes

La calidad general de los datos AIS está influenciada por el desempeño de los instrumentos de navegación a bordo y por las condiciones de propagación de las ondas de radio durante la transmisión de información. Por otro lado, la calidad de los datos específicos contenidos en los mensajes AIS depende del rendimiento de dispositivos como el giróscopo, el receptor del Sistema Global de Navegación por Satélite (GNSS). Existen 3 tipos de transpondedores AIS [24]:

- AIS Clase A: Es sólo obligatorio para aquellos buques de más de 300 toneladas brutas, buques, buques pesqueros de más de 15 metros de eslora, buques de transporte de pasajeros y para el convenio SOLAS. Los datos de esta clase de transpondedor son más completos en comparación con los de la Clase B. No obstante, son más caros y pesados.
- AIS Clase B: Está diseñado para ser instalados en embarcaciones de recreo y embarcaciones no englobadas en el convenio SOLAS. Este sistema aporta menos información, pero es más económico y ligero.
- AIS Clase B+: También llamados Clase B SOTDMA. Utilizan la misma tecnología que los transpondedores AIS de Clase A. Su ventaja principal es que asegura siempre la transmisión de datos garantizando un time slot, incluso en las zonas donde el tráfico es denso.

En la Tabla 2-5 se resumen las características más esenciales para cada tipo de transpondedor AIS.

FUNCIÓN	CLASE A	CLASE B	CLASE B+
Potencia de transmisión	12,5 W	2 W	5W
Alcance aproximado	20-25 millas náuticas	7-8 millas náuticas	10-12 millas náuticas
Ritmo transmisión	Hasta 2 segundos	Hasta 30 segundos	Hasta 5 segundos
Time slot garantizado	Sí	No	Sí
Datos de ruta	Sí	No	No
Precio aproximado	2500 €	650 €	900 €

Tabla 2-5 Comparativa de funcionalidad según la clase [25]

Los transpondedores AIS emiten datos que se pueden clasificar en 4 grandes grupos [26]:

- Información estática (transmitida cada seis minutos):
 - Número MMSI: *Maritime Mobile Service Identify*. Es una serie de nueve dígitos que identifica a cada estación transmisora.
 - Número IMO: *International Maritime Organization*. Es un número de siete dígitos que identifica al buque, al propietario y a la empresa que lo está empleando.
 - Nombre y distintivo de llamada del buque.
 - Eslora y manga del buque.
 - Tipo de buque: define el tipo de barco mediante un número de dos dígitos desde el 10 hasta el 99.
 - Lugar donde se sitúa la antena transmisora.
- Información dinámica (depende de la velocidad y el rumbo. Véase Tabla 2-6):
 - Posición GPS junto a la precisión.
 - Velocidad sobre el fondo (SOG).
 - Rumbo sobre fondo (COG).
 - Referencia horaria en UTC.
 - Condición de navegación (a vela, fondeado, maniobra restringida...).
- Información detallada del viaje (cada seis minutos o cuando se hayan realizado cambios):
 - Calado.
 - Carga.
 - Plan de ruta mediante *waypoints*.
 - Puerto de destino.
 - Hora de llegada estimada (ETA).
- Mensajes cortos relacionados con la seguridad: Es un mensaje con formato libre que se transmite a las unidades y estaciones para informar de posibles peligros.

CONDICIÓN DEL BARCO	CLASE A	CLASE B	CLASE B+
0-2 nudos y atracado o fondeado	3 minutos	3 minutos	3 minutos
2-14 nudos	10 segundos	30 segundos	30 segundos
2-14 nudos y cambio de rumbo	3,3 segundos	30 segundos	30 segundos
14-23 nudos	6 segundos	30 segundos	15 segundos
14-23 nudos y cambio de rumbo	2 segundos	30 segundos	15 segundos
> 23 nudos	2 segundos	30 segundos	5 segundos

Tabla 2-6 Intervalos de transmisión de datos [27]

2.4 Trabajos relacionados

En el presente capítulo se muestran algunos de los trabajos que se han realizado en el campo de estudio de este mismo trabajo y han servido, en ocasiones, para dar un enfoque más preciso y asentar las bases de ciertos análisis. A continuación, se exponen y explican brevemente en qué consisten.

- *Análisis de los sistemas de indexado geoespacial para el Conocimiento del Entorno Marítimo:* Este trabajo se ve en [28], es un TFG realizado por un antiguo alumno del Centro Universitario de la Defensa de Marín.
Este trabajo examina la aplicación de sistemas de indexación geoespacial como herramienta para mejorar la capacidad analítica del Centro de Vigilancia Marítima (COVAM). Tras analizar distintos métodos, se selecciona Uber H3.
Mediante el uso de MySQL y datos reales, se llevan a cabo consultas espaciales (como la localización de buques o la identificación de zonas de tránsito habitual) que permiten evaluar la eficacia del sistema. Los resultados evidencian una mejora sustancial en la rapidez y eficiencia del análisis geoespacial frente a métodos tradicionales, destacando el valor de H3 como herramienta para el conocimiento operativo del entorno marítimo.
- *Detección de anomalías en el tráfico marítimo basada en inteligencia artificial:* Este trabajo, que se aprecia en [4], analiza cómo detectar anomalías en el tráfico marítimo utilizando inteligencia artificial, con especial atención al papel del Sistema de Identificación Automática (AIS) como herramienta clave para monitorizar la navegación. Ante el crecimiento sostenido del transporte marítimo y su importancia económica y estratégica, se vuelve esencial identificar comportamientos anómalos en tiempo real.
La investigación revisa técnicas tradicionales y modernas aplicadas a la detección de desviaciones en los patrones habituales de navegación. Para ello, se emplea un conjunto de datos históricos del AIS proporcionado por la Guardia Costera de los Estados Unidos sobre el cual se reconstruyen trayectorias de buques. El trabajo propone así un enfoque que combina análisis de datos masivos con técnicas de IA, contribuyendo a mejorar la seguridad y la vigilancia marítima.
- *Revisión sobre la inserción de celdas hexagonales en regiones de interés circulares y hexagonales:* Este artículo [29] ofrece una revisión especializada sobre el uso de celdas hexagonales aplicadas a regiones de interés de forma circular o hexagonal, analizando cómo estas estructuras pueden ser embebidas eficazmente para cubrir dichas áreas sin solapamientos ni huecos. Las celdas hexagonales, frente a las cuadradas o rectangulares, presentan ventajas en numerosos campos por su eficiencia en el uso del espacio, la reducción de recursos, y su mejor comportamiento en algoritmos de proximidad. El estudio recopila distintas fórmulas y expresiones algebraicas empleadas para calcular la cantidad de celdas, vértices o aristas necesarias para cubrir una región de interés, unificando investigaciones previamente aisladas y derivando fórmulas propias cuando los trabajos consultados no ofrecen datos completos.
El artículo guarda una conexión directa con este TFG, ya que ambos trabajos exploran la disposición y eficiencia de celdas hexagonales en un espacio geoespacial concreto. En este trabajo, se aplica el sistema Uber H3 para indexar datos AIS en el entorno marítimo; este estudio complementa dicha aplicación proporcionando fundamentos matemáticos sobre cómo maximizar la cobertura espacial mediante celdas hexagonales, lo cual es clave para optimizar tanto el análisis como la representación de los datos.

3 DESARROLLO

El estudio de los sistemas de indexación geoespacial, junto a las frecuencias de transmisión AIS, permite evaluar la posición, el área y la trayectoria de los objetos marítimos. En estos análisis se procesan grandes volúmenes de datos espaciales. Además, la representación en mapas interactivos facilita una visualización intuitiva y detallada, mejorando la interpretación de los resultados. A lo largo del estudio, se presentan distintos tipos de análisis con explicaciones detalladas de los fragmentos de código más relevantes, garantizando claridad en la metodología empleada. Finalmente, los hallazgos obtenidos se someten a una evaluación rigurosa en el capítulo siguiente, proporcionando una base sólida para su aplicación en contextos operativos.

3.1 Entorno de trabajo y software

La plataforma web elegida para procesar los datos y aplicar técnicas de aprendizaje automático ha sido *Jupyter Notebook* [30], mientras que el lenguaje utilizado para la implementación del código ha sido *Python* [31]. A continuación, se expondrán las razones que justifican esta elección y se describirán los primeros pasos en su uso dentro del desarrollo del proyecto.

3.1.1 *Python*

La elección de *Python* como lenguaje de programación se fundamenta en su versatilidad y accesibilidad lo han convertido en una de las herramientas más poderosas en el ámbito del análisis de datos y la computación científica. Más allá de su popularidad, lo que realmente destaca de *Python* es la claridad de su sintaxis, lo que facilita enormemente la escritura y comprensión del código. Esto no solo optimiza el tiempo de desarrollo, sino que también reduce la probabilidad de errores, permitiendo un flujo de trabajo más eficiente y menos propenso a fallos.

Uno de los mayores atractivos de *Python* es su vasta colección de bibliotecas especializadas, que ofrecen soluciones eficientes para problemas específicos. En el contexto de este trabajo, ha resultado especialmente relevante la biblioteca H3 [32], desarrollada por Uber, diseñada para la manipulación y análisis de datos espaciales a través de una estructura jerárquica de celdas hexagonales. Su uso resulta crucial para evaluar la organización y funcionamiento de las celdas Uber H3 en el tratamiento de datos AIS, permitiendo una representación geoespacial precisa y flexible.

Cabe destacar que, aunque el distribuidor de software Anaconda y proporciona una versión integrada de *Python* y de H3, aunque esta última se encontraba desactualizada en el momento de la implementación. Por este motivo, se optó por instalar la versión 4.1.2 de la biblioteca directamente a través de “pip”, asegurando así la compatibilidad con las funcionalidades más recientes y optimizadas. Esta decisión ha permitido un mejor rendimiento en la manipulación de datos espaciales y una integración más fluida con el resto de las herramientas utilizadas en el proyecto.

```
pip install h3
```

Dentro de este marco, *Python* se convierte en la opción natural debido a su potencia en la manipulación de datos, su extensa comunidad y su ecosistema de bibliotecas optimizadas para el análisis geoespacial. Su capacidad para procesar grandes volúmenes de datos y generar representaciones gráficas detalladas lo hace especialmente adecuado para el estudio de patrones de tráfico marítimo y sistemas de indexación geoespacial.

3.1.2 Jupyter Notebook

En el desarrollo del presente TFG se ha optado por el uso de *Jupyter Notebook* como plataforma principal para la implementación del código y la ejecución de los experimentos. Esta elección se fundamenta en varias razones técnicas y metodológicas que hacen de *Jupyter* una herramienta idónea para este tipo de análisis.

Esta herramienta, instalada a través de Anaconda [33], ofrece un entorno optimizado para el análisis de datos y aprendizaje automático. Más que una simple distribución de *Python*, Anaconda reúne un conjunto de bibliotecas y gestores que facilitan la manipulación eficiente de datos a gran escala. Su ecosistema permite ejecutar código en *Jupyter Notebook* (Figura 3-1), integrar paquetes especializados y garantizar un flujo de trabajo ágil, flexible y reproducible en el ámbito de la ciencia de datos y el modelado geoespacial.

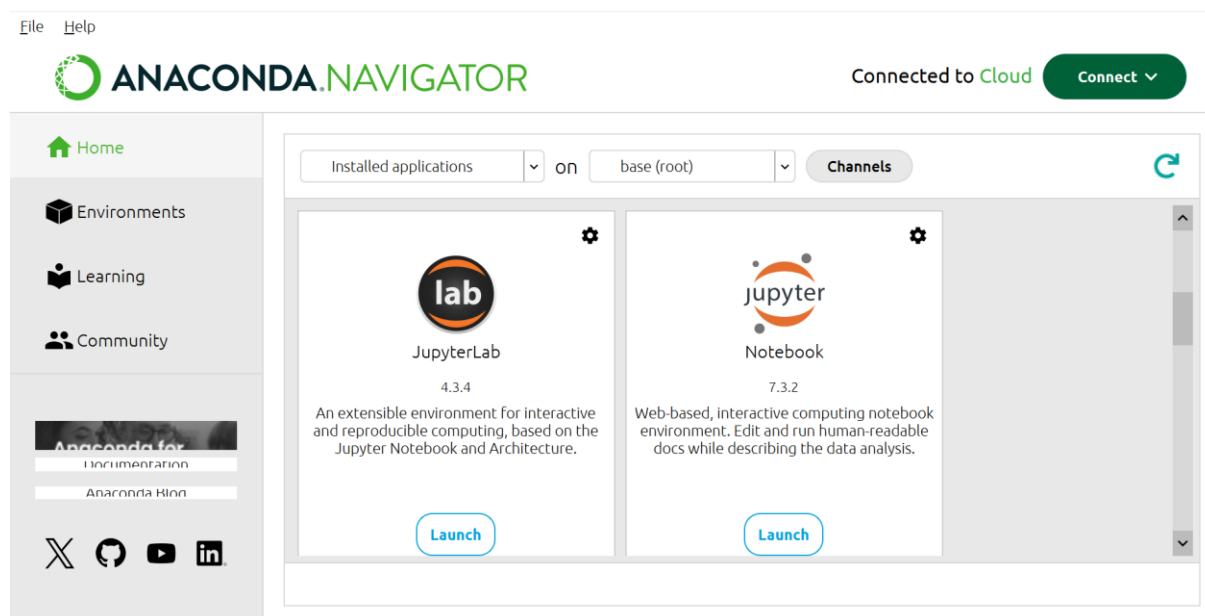


Figura 3-1 Interfaz del navegador de Anaconda

Jupyter Notebook es un entorno interactivo de programación basado en bloques, diseñado para el desarrollo de código en múltiples lenguajes, siendo *Python* el más utilizado. Su flexibilidad y capacidad para integrar código, visualizaciones y documentación explicativa en un mismo documento hacen que sea ampliamente empleado en el ámbito del análisis de datos, la inteligencia artificial y la investigación académica.

Uno de los principales atributos de *Jupyter Notebook* es su arquitectura basada en módulos dentro de un mismo archivo, que permite dividir el código en unidades independientes de ejecución. A diferencia de los entornos de desarrollo tradicionales, donde el código se ejecuta de manera secuencial en un solo archivo, *Jupyter Notebook* facilita un enfoque más dinámico y estructurado, lo cual resulta especialmente útil en el análisis de grandes volúmenes de datos.

La estructura basada en unidades independientes de *Jupyter Notebook* (Figura 3-2) ofrece diversas ventajas para el análisis y evaluación del sistema Uber H3 en el procesamiento de datos AIS:

- Ejecución modular y controlada: la posibilidad de ejecutar celdas de manera individual permite probar funciones y fragmentos de código sin afectar el resto del análisis, facilitando la depuración y la optimización de algoritmos.
- Facilidad de integración con bibliotecas especializadas: gracias a su compatibilidad con Python y múltiples bibliotecas, permite la manipulación eficiente de datos geoespaciales con H3, el análisis de datos con Pandas, entre otros.
- Software gratuito: facilita su adopción sin necesidad de licencias o pagos.
- Ejecución en el navegador: No requiere instalación en el equipo, ya que puede utilizarse directamente desde un navegador web, permitiendo el trabajo en línea sin necesidad de configuraciones adicionales.



Figura 3-2 Interfaz de celdas de Jupyter Notebook

El uso de Jupyter Notebook como herramienta principal en el desarrollo del TFG responde a la necesidad de un entorno flexible, modular y eficiente para el enfoque de diversos análisis. Su estructura basada en bloques independientes permite un enfoque progresivo y dinámico en la manipulación de datos geoespaciales, facilitando la exploración, visualización y documentación del proceso. Estas ventajas lo convierten en una opción idónea para garantizar un desarrollo metodológico riguroso y reproducible, aspectos fundamentales en un trabajo de carácter académico y técnico.

3.1.3 Librerías

Para llevar a cabo el desarrollo del código implementado en este Trabajo Fin de Grado, ha sido necesario hacer uso de diversas librerías de Python que han facilitado tanto el procesamiento de datos como la representación geoespacial y el análisis de resultados. A continuación, en la Tabla 3-1 se presentan todas las librerías utilizadas, junto con una breve descripción de su funcionalidad principal dentro del proyecto, lo que permitirá comprender el papel que cada una ha desempeñado en la implementación y ejecución de este.

Librería	Funcionalidad
<i>geopy</i>	Cálculo de distancias geográficas
<i>h3</i>	Indexación y análisis espacial Uber H3
<i>keplergl</i>	Visualización geoespacial interactiva
<i>math</i>	Operaciones matemáticas básicas
<i>matplotlib</i>	Generación de gráficos y visualizaciones
<i>numpy</i>	Cálculo numérico y manejo de gran número de datos
<i>pandas</i>	Manipulación y análisis de datos
<i>random</i>	Generación de números aleatorios
<i>s2sphere</i>	Cálculo y modelado sobre esfera terrestre Google S2
<i>scipy</i>	Funciones científicas y estadísticas
<i>shapely</i>	Manipulación de geometrías espaciales
<i>tqdm</i>	Barra de progreso en bucles

Tabla 3-1 Librerías empleadas

3.1.4 Kepler.gl

El uso de Kepler.gl [34] en trabajo, se justifica por su capacidad para visualizar datos geoespaciales de manera interactiva y eficiente. Esta herramienta, desarrollada por Uber y que actualmente es *open source* y está disponible en GitHub, permite representar información marítima en mapas dinámicos, facilitando la identificación de patrones espaciales en los datos AIS. Su compatibilidad con el sistema de celdas Uber H3 resulta especialmente útil en este estudio, ya que permite analizar la segmentación del espacio marítimo y evaluar la eficacia de esta estructura en la organización de grandes volúmenes de información. La Figura 3-3 muestra un ejemplo de mapa interactivo en Kepler.gl

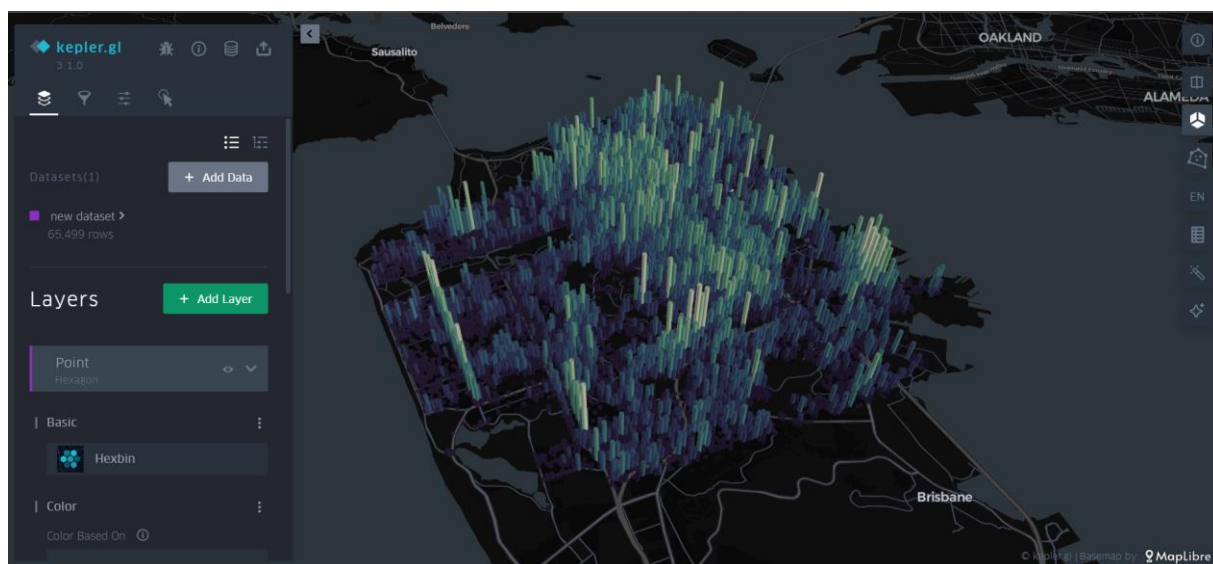


Figura 3-3 Ejemplo de visualización de mapa en Kepler.gl [34]

Además, Kepler.gl se integra fácilmente con Jupyter Notebook mediante la librería *Keplergl*, lo que permite una exploración visual directa de los datos dentro del entorno de desarrollo. Su interfaz intuitiva y capacidad para manejar grandes conjuntos de datos, lo convierten en una herramienta clave para la visualización y análisis de datos de las celdas Uber H3. Gracias a estas características, Kepler.gl facilita la interpretación de resultados y mejora la toma de decisiones en el estudio del sistema Uber H3 aplicado al tráfico marítimo.

La instalación de esta librería para su integración en el entorno de trabajo se lleva a cabo a través de Anaconda, garantizando una gestión eficiente de dependencias y compatibilidades. En este proceso, se instala la versión 0.3.2, asegurando la estabilidad y funcionalidad requerida para el desarrollo de los análisis geoespaciales. Dicha instalación se hace a través del propio distribuidor Anaconda, asegurando así la compatibilidad de la librería.

3.2 Consideraciones previas

Tras haber expuesto las principales herramientas que se emplearán en el desarrollo de los análisis posteriores, en primer lugar, se sientan las bases metodológicas para el análisis y se desarrolla el sistema de generación de muestras (puntos geográficos) necesarios para los análisis posteriores. En esta sección, por tanto, se establecen los fundamentos para los posteriores experimentos.

3.2.1 Uso de IA generativa

Este trabajo combina técnicas tradicionales de investigación con herramientas modernas de inteligencia artificial generativa para mejorar la eficiencia y precisión en su desarrollo. Parte del contenido de la presente memoria ha sido creado utilizando diversas inteligencias artificiales generativas, una herramienta avanzada de generación de texto basada en IA. Asimismo, el código en Python empleado para la recogida de datos, análisis y simulaciones fue desarrollado parcialmente con

el apoyo de ChatGPT (OpenAI), complementado con ajustes y optimizaciones manuales para asegurar la calidad y funcionalidad del programa.

La integración de estas herramientas se realizó de manera ética y responsable, verificando la originalidad y evitando el plagio. Todas las contribuciones generadas por la inteligencia artificial fueron supervisadas, editadas y adaptadas para alinearse con los objetivos específicos del trabajo y las directrices académicas establecidas por la Universidad de Vigo y la ENM.

3.2.2 Simulación Montecarlo

La simulación de Monte Carlo es un algoritmo computacional que se basa en la generación repetida de muestras aleatorias para estimar la probabilidad de ocurrencia de diferentes resultados en un sistema incierto. El nombre "Monte Carlo" se inspira en el famoso casino de Mónaco, reflejando la naturaleza probabilística y aleatoria de estos métodos [35].

El proceso consiste en realizar múltiples iteraciones del modelo, cada vez con un conjunto diferente de números aleatorios que se generan dentro de los valores mínimo y máximo definidos para cada variable. Este procedimiento se repite miles de veces en una simulación típica, generando así un amplio rango de escenarios posibles. Esto permite evaluar modelos complejos donde no se cuenta con una aproximación analítica.

Por otra parte, en este trabajo, algunas métricas pueden calcularse mediante expresiones analíticas directas, como la distancia media entre un punto y el centro de un hexágono. Sin embargo, en otros escenarios, la realidad es más compleja: los hexágonos no son perfectamente homogéneos o no existe una fórmula exacta que describa la relación buscada. Para abordar esta incertidumbre, se emplea la simulación Monte Carlo, generando múltiples puntos o trayectorias al azar y evaluando la métrica en cada uno. A medida que el número de iteraciones crece, los valores estadísticos obtenidos, como promedios y distribuciones, convergen hacia estimaciones fiables de la magnitud real.

3.2.3 Generación de puntos

El primer paso en la realización de los experimentos consiste en la generación automática y aleatoria de localizaciones. Para ello, se generan un gran número de puntos distribuidos aleatoriamente sobre la superficie terrestre. Cuanto mayor sea la cantidad de puntos generados, los resultados serán más precisos.

Para garantizar una distribución uniforme en la superficie terrestre no se puede seleccionar una latitud y una longitud aleatoria dentro de los intervalos posibles. Esto se debe a que la diferencia longitud los paralelos y de las franjas longitudinales encerradas entre ellos. Si la latitud se seleccionara de manera uniforme dentro del intervalo $[-90^\circ, 90^\circ]$, se produciría una concentración excesiva de puntos en las regiones polares y una menor densidad en la zona ecuatorial, lo que afectaría la uniformidad de la distribución. Asemejando la Figura 3-4 a una porción terrestre se aprecia mayor concentración de puntos en los paralelos de latitudes más altas. Para ello, se corrige mediante la fórmula **(1)**, siendo Φ la latitud (véase línea 10 del Anexo V: Código de distribución de error para una resolución)

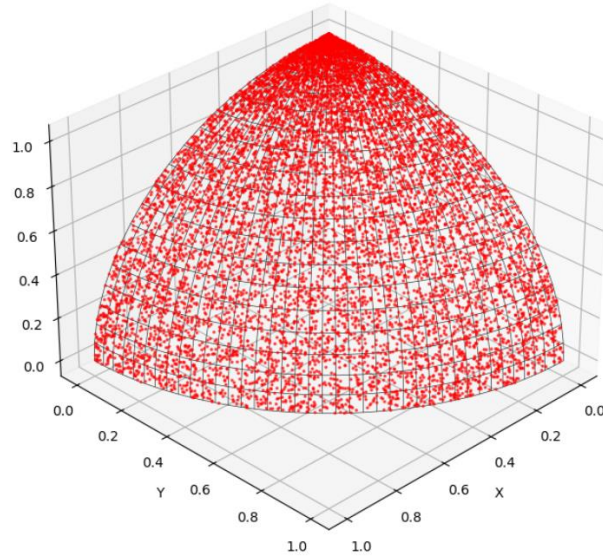


Figura 3-4 Distribución de puntos concentrada en latitudes altas

$$\Phi = \text{asin}(u) \quad (1)$$

$$u \in [-1, 1]$$

Por tanto, al proyectar los puntos generados uniformemente sobre un plano (Figura 3-5 izquierda) sobre la esfera, se genera una concentración en los polos (Figura 3-6 izquierda). Por el contrario, como se aprecia en la Figura 3-5 en la distribución de la derecha, si se plasma sobre una superficie esférica la distribución corregida, la densidad de puntos en los polos es idéntica a la cerca del ecuador.

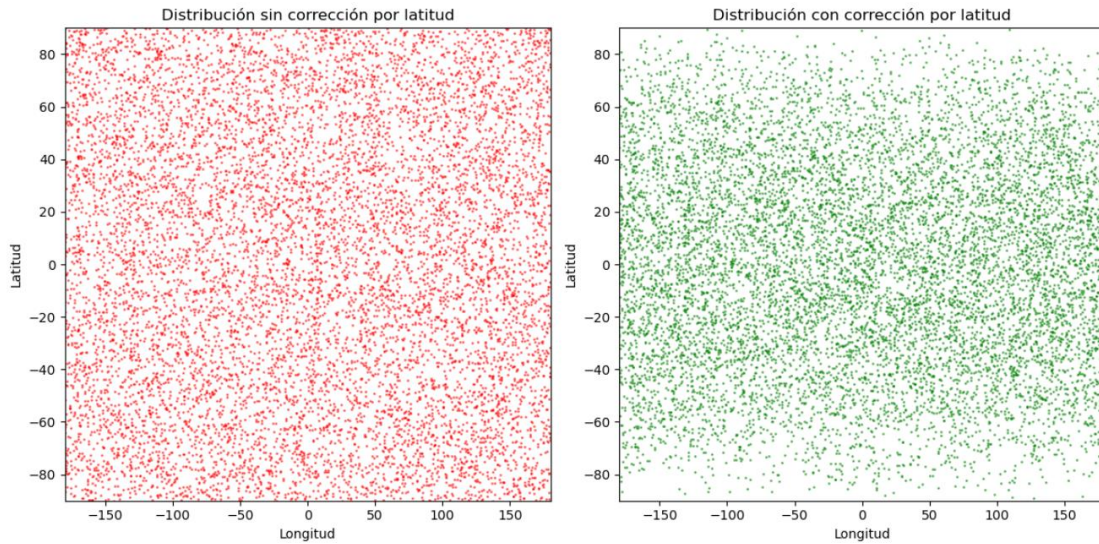


Figura 3-5 Comparación de distribución corregida por latitud

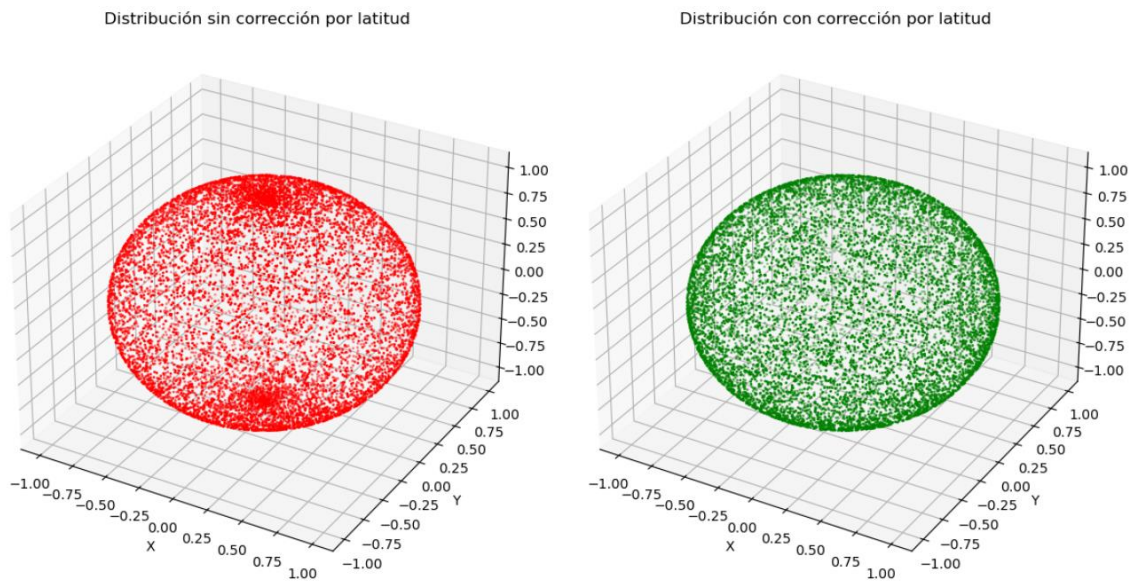


Figura 3-6 Representación en 3D de distribución corregida por latitud

Por otra parte, debido a que el análisis de este trabajo se centra en el entorno marítimo, es necesario hallar la manera para que, de aquellos puntos generados aleatoriamente, el programa deseche los que coinciden con regiones en tierra. Para ello se han utilizado archivos de web *Natural Earth Data* [36]. A su vez, se emplean librerías especializadas en el manejo de datos geoespaciales, tales como *GeoPandas* y *Shapely*. El ejemplo de que este archivo es leído y cargado correctamente se aprecia en una representación de los puntos generados en Kepler.gl, como se puede observar en la Figura 3-7.



Figura 3-7 Generación de puntos aleatorios sólo en agua

3.2.4 Cálculo distancias geodésicas

En el ámbito del análisis geoespacial, no basta con distribuir puntos aleatorios de manera uniforme; es igualmente crucial garantizar que las distancias entre ellos sean calculadas con la mayor precisión posible. Aquí es donde entra en juego el concepto de distancia geodésica, que se refiere a la medida más corta entre dos puntos situados sobre la superficie curva de la Tierra. Este cálculo, lejos de ser trivial, requiere métodos sofisticados que tengan en cuenta la forma real del planeta, evitando simplificaciones que puedan introducir errores significativos.

Afortunadamente, la librería *Geopy* [37] se encarga de esta tarea de forma eficiente, implementando el modelo del elipsoide WGS-84 (*World Geodetic System* 1984), el estándar de referencia adoptado por la mayoría de los sistemas de navegación modernos, incluido el GPS. A diferencia de los enfoques que asumen una Tierra perfectamente esférica, este modelo reconoce su naturaleza ligeramente achatada en los polos y ensanchada en el ecuador, lo que permite cálculos más ajustados a la realidad.

El impacto de esta precisión se hace evidente en numerosas aplicaciones prácticas. Desde la planificación de rutas marítimas y aéreas hasta el análisis de datos de tráfico o el estudio de desplazamientos en grandes superficies, contar con un método fiable para medir distancias geodésicas puede marcar la diferencia entre un modelo analítico riguroso y uno lleno de imprecisiones. En definitiva, herramientas como *geopy.distance* no solo simplifican el trabajo de los desarrolladores, sino que también garantizan que las decisiones basadas en datos geoespaciales se fundamenten en mediciones exactas y realistas.

El cálculo exacto de distancias es esencial para:

- Medir el error de posición entre punto real y el centroide de su celda H3.
- Comparar trayectorias reales con trayectorias discretizadas.
- Evaluar la distancia entre celdas vecinas.
- Determinar la cobertura efectiva entre áreas circulares.

Estas consideraciones previas fueron indispensables para establecer una base sólida que garantizara la fiabilidad de los análisis desarrollados. A partir de las herramientas y métodos descritos, en los siguientes apartados se detallan los experimentos diseñados para evaluar tres aspectos fundamentales del sistema Uber H3: la precisión en la representación de posiciones, la eficiencia en la cobertura de áreas de interés y la fidelidad en la segmentación de trayectorias AIS.

3.3 Cálculo del error de posición

Este primer experimento tiene como objetivo principal evaluar el error de posición cuando se discretizan los puntos mediante el sistema Uber H3, para lo cual se crean puntos aleatorios. El procedimiento consiste en seleccionar puntos al azar, determinar la celda correspondiente según la resolución establecida y, a continuación, identificar el centroide de dicha celda. Posteriormente, se calcula la distancia geodésica entre el punto generado de manera aleatoria y el centro de la celda asignada. Esta distancia se interpreta como el error de posición introducido por el sistema al asumir que cualquier punto contenido en la celda se encuentra efectivamente en su centroide.

A continuación, como primera aproximación, se representa la distribución de las distancias una celda hexagonal perfecta de radio máximo 1, calculada para 10 millones de puntos aleatorios dentro de ese hexágono.

La gráfica de la Figura 3-8 representa un histograma de la cantidad de puntos en función de su distancia al centro del hexágono. El eje de abscisas (“Distancia desde el centro”) representa la distancia de cada punto aleatorio desde el centro del hexágono. Se extiende en 100 intervalos (*bins*) desde 0 hasta 1 debido a que se ha generado un hexágono regular de radio 1. Sin embargo, se representa una línea vertical discontinua de color rojo que representa la apotema del hexágono de radio 1 (2).

$$Apotema = \frac{\sqrt{3}}{2} \times radio \approx 0,866025 \times radio \quad (2)$$

A su vez, el eje de coordenadas (“Frecuencia”) representa la cantidad de puntos que se encuentran a cada distancia dentro del hexágono. La forma de distribución sigue un crecimiento lineal hasta la apotema, lo cual tiene sentido ya que a medida que la distancia aumenta, el número de puntos posibles en el anillo correspondiente aumenta proporcionalmente a su radio. No es así a partir de la apotema, donde empieza a disminuir debido a que el número de puntos posibles disminuye al acercarse al borde

del hexágono; generando una cola que corresponde a los puntos más alejados. Esta cola se debe a que, aunque las esquinas representan una pequeña fracción del área total, los puntos generados en estas regiones presentan distancias mayores al centro.

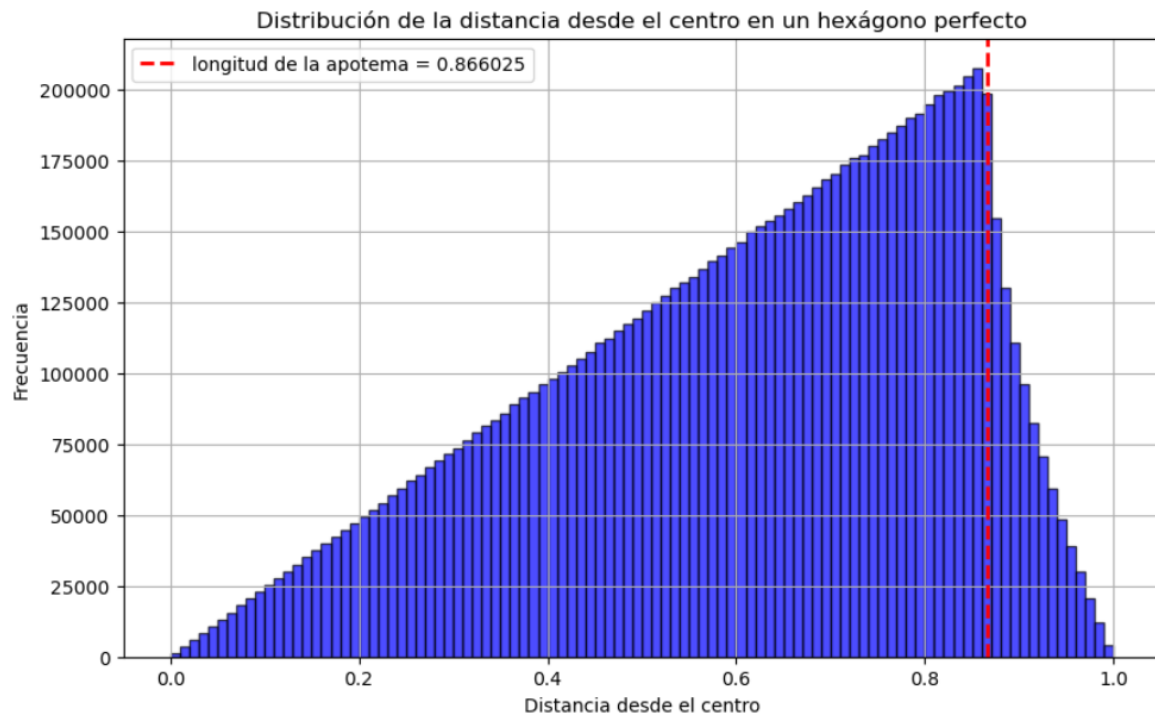


Figura 3-8 Distribución de distancia para hexágono perfecto de lado 1

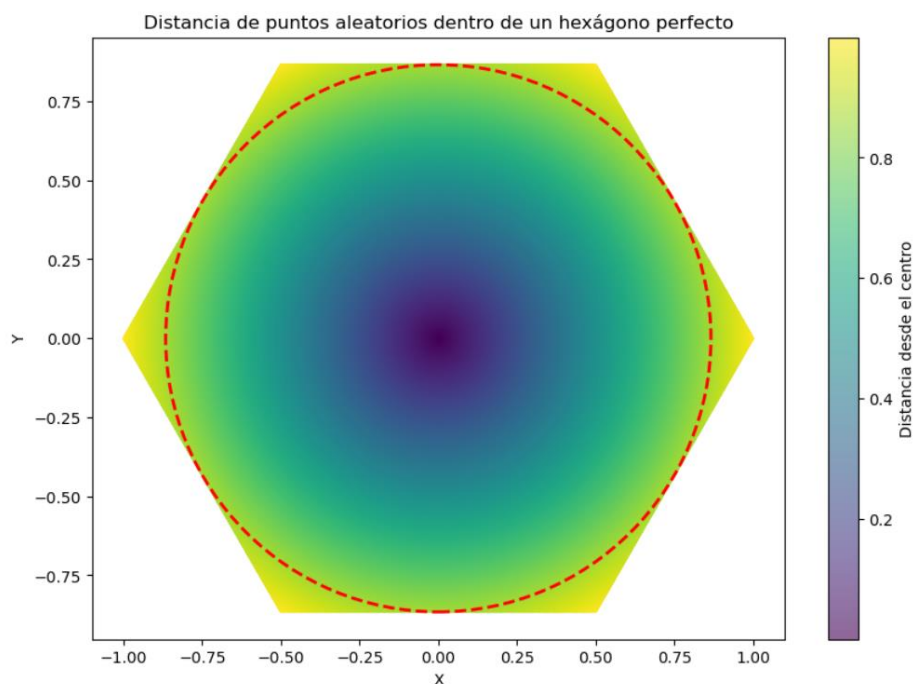


Figura 3-9 Distancia de puntos dentro de un hexágono perfecto

Relacionar la Figura 3-8 y la Figura 3-9 adquiere un significado más completo, permitiendo visualizar no solo la cantidad de puntos según su distancia al centro, sino también su distribución espacial dentro del hexágono. Cada punto generado aleatoriamente se proyecta en un sistema de coordenadas (x, y), delimitado por la geometría de un hexágono regular de radio 1.

La representación mediante gradiente de color en el análisis de distribución espacial dentro de un hexágono proporciona una dimensión adicional de información. Los colores oscuros, como el morado y

el azul, denotan puntos cercanos al centro del hexágono, es decir, a menor distancia del mismo. En contraste, los tonos verdes y amarillos indican puntos próximos al borde, donde la distancia al centro alcanza su valor máximo. Esta codificación cromática permite una comprensión visual inmediata de la distribución de distancias.

Un elemento destacable de esta visualización es el círculo punteado en rojo, que no es un límite arbitrario, sino que corresponde al círculo inscrito en el hexágono, cuyo radio equivale a la apotema. Dicha apotema representa la distancia mínima entre el centro del hexágono y sus lados, delimitando la zona donde se concentra la mayoría de los puntos analizados. Este hecho confirma la observación obtenida en el histograma correspondiente: aunque existen puntos más allá de la apotema, la mayor densidad se encuentra dentro del círculo inscrito, evidenciando un patrón de distribución no uniforme.

La comparación entre esta visualización espacial y el histograma que muestra la distribución de distancias permite complementar los niveles de análisis. Mientras que el histograma ofrece una perspectiva cuantitativa, identificando la frecuencia con la que se presentan diferentes distancias al centro, la representación espacial permite interpretar cómo la geometría del hexágono condiciona dicha distribución en el plano. Así, el análisis integrado de ambas representaciones revela no solo la cantidad de puntos a cada distancia del centro, sino también su disposición relativa dentro de la figura geométrica. (véase Anexo IV: Código de distribución de distancias en un hexágono).

En relación con lo anterior y teniendo en cuenta la irregularidad en el tamaño de las celdas, como se detalla en la Tabla 2-3, es fundamental profundizar en el comportamiento de la distribución del error de posición en este experimento. Por tanto, se realiza el experimento anterior, pero utilizando celdas Uber H3 y, para obtener resultados estadísticamente significativos, se ha generado una muestra de 10000000 de puntos aleatorios distribuidos en el océano, permitiendo así un análisis detallado de los patrones de error que se producen al asociar estos puntos con sus respectivas celdas en el sistema H3.

Antes de analizar los resultados de la figura es necesario destacar 2 de las funciones de la librería de H3 y una de la librería *Geopy*. Con la primera de ellas, introduciendo como variables, la latitud, la longitud y la resolución (en ese orden) se obtiene el índice en formato de cadena hexadecimal de la celda de solicitada resolución a la que corresponden esas coordenadas. La segunda función esencial en este análisis es el paso inverso; es decir, de una cadena hexadecimal obtener las coordenadas (latitud y longitud) del centro de la celda cuyo índice se ha introducido. Y, por último, cabe de destacar la función que a partir de dos puntos se obtiene la distancia geodésica, en este caso, en kilómetros.

```
h3.latlng_to_cell(lat, lng, resolution)
cell_to_latlng(h)
geodesic(punto, centroide).kilometers
```

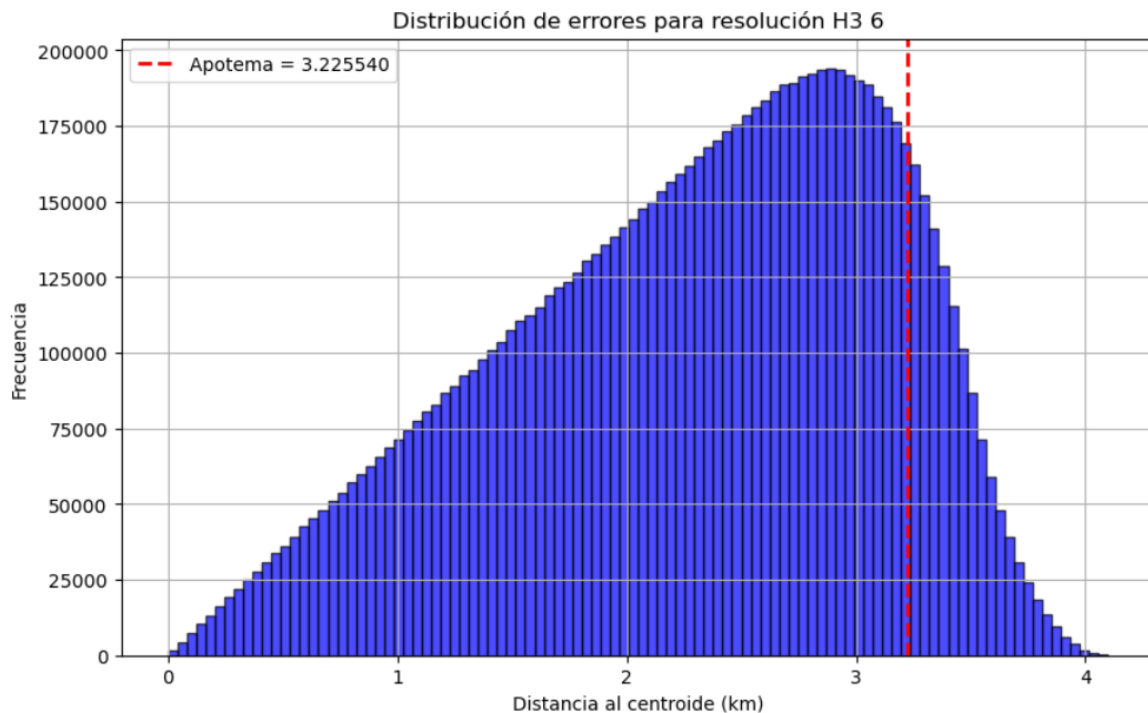


Figura 3-10 Distribución de error según resolución 6

Figura 3-10 representa la distribución de errores en la asignación de puntos dentro de una celda hexagonal en la resolución 6 de H3, expresada en términos de la distancia al centroide de la celda. Se trata de un histograma que permite analizar cómo varían estas distancias y qué patrón siguen dentro del sistema de indexado espacial H3, proporcionando una visión clara de la precisión y los errores introducidos por la discretización en celdas hexagonales.

El eje horizontal muestra la distancia de cada punto al centroide de la celda hexagonal a la que ha sido asignado, con valores que oscilan entre 0 km, cuando un punto se encuentra exactamente en el centroide, y aproximadamente 4 km, lo que marca el máximo error posible dentro de una celda de esta resolución. Por otro lado, el eje vertical representa la frecuencia de puntos dentro de cada intervalo de distancia, revelando cómo se distribuyen en términos de proximidad al centroide. La forma de la curva sugiere que la mayoría de los puntos se concentran en distancias intermedias dentro del hexágono, con una clara acumulación hacia la apotema de la celda.

Un aspecto fundamental de esta gráfica es la línea punteada roja, que marca la apotema del hexágono medio y tiene un valor de 3,225540 km (calculado mediante fórmula (2) y la Tabla 2-2). Esta línea señala un umbral natural en la distribución: la frecuencia comienza a disminuir bruscamente después de este punto, lo que refleja la geometría hexagonal del sistema H3.

El aspecto más llamativo quizás de esta gráfica es que la curva se suaviza cuando está llegando a la apotema. Esto se debe a que, a diferencia del análisis anterior, este análisis se crea para 10000000 puntos, pero cada uno ubicado en una celda distinta y, como no todas las celdas son iguales, no termina tan abruptamente como en el caso ideal, que el crecimiento lineal coincide justamente con el valor de la apotema que varía para cada celda.

La gráfica resultante muestra claramente una cola en el extremo derecho de la distribución como en el caso del hexágono ideal, un fenómeno que no es aleatorio, sino que responde a las características geométricas propias del sistema H3. En particular, esta asimetría en la distribución se explica por diversas causas inherentes a la estructura y proyección del sistema de indexación geoespacial.

El factor más relevante que contribuye a la aparición de esta cola en la distribución se encuentran las irregularidades en la forma de las celdas. Aunque el sistema H3 está diseñado para generar principalmente celdas hexagonales, la proyección de la superficie curva de la Tierra introduce

deformaciones geométricas que pueden alterar la uniformidad del sistema. En ciertas regiones, aparecen celdas pentagonales que provocan distorsiones adicionales en el patrón regular, lo que incrementa la distancia desde ciertos puntos al centroide de la celda. No obstante, estos pentágonos son prácticamente irrelevantes estadísticamente. Como referencia, si se tomase el valor medio del lado de un hexágono de resolución 6 (según la Tabla 2-2), este no debería superar los 3,726 kilómetros. Sin embargo, de la Tabla 2-3 se extrae que sus lados pueden extenderse hasta 4,096 kilómetros, valor que coincide con el máximo registrado en la gráfica.

A continuación, se procede a realizar el mismo análisis, pero par una resolución mucho mayor. Este análisis es de especial interés para demostrar si conforme se aumenta la resolución, la similitud a celdas perfectas es mayor.

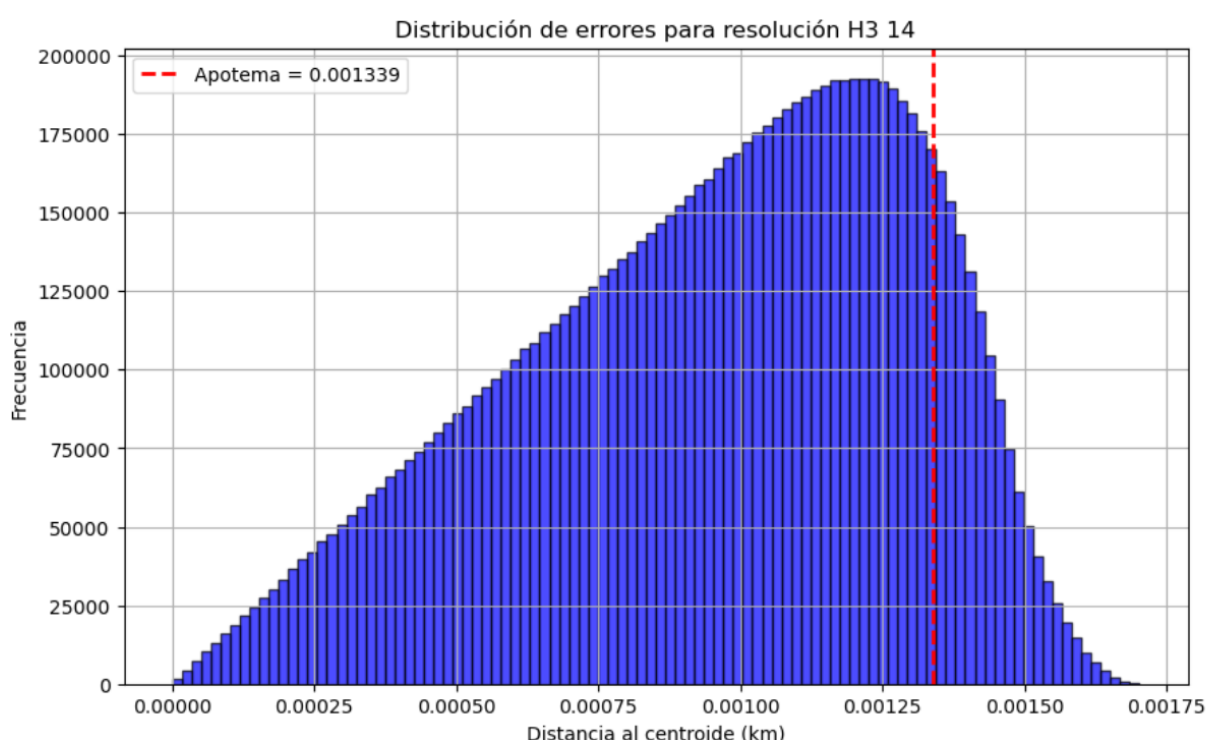


Figura 3-11 Distribución de error según resolución 14

Aunque en el sistema H3 se espera que a medida que aumenta la resolución la distribución de errores mejore o se vuelva más uniforme, la Figura 3-11 revela que la distribución del error se mantiene geoméricamente similar en ambos casos (resolución 6 y 14). Es decir, la forma de la distribución prácticamente no varía, pese a las diferencias en la escala absoluta de los errores: kilómetros frente a milímetros.

Este resultado es especialmente interesante porque, aunque las celdas a resoluciones más altas deberían teóricamente presentar menos irregularidades geométricas y, por tanto, errores más contenidos o uniformes, en la práctica la Figura 3-10 y la Figura 3-11 muestran que la estructura del error se mantiene. La “regularización” que aporta la subdivisión de las celdas no cambia sustancialmente la distribución relativa de los errores, simplemente los reduce en magnitud, pero preserva el patrón geométrico.

Por tanto, como cierre para este análisis, la presencia de la cola en la gráfica refleja los casos en los que el error de posición alcanza sus valores más altos. Estas situaciones se producen cuando los puntos aleatorios se ubican en zonas donde las irregularidades geométricas de las celdas son más pronunciadas. Este comportamiento resalta la importancia de considerar la resolución adecuada al utilizar el sistema H3, ya que optar por resoluciones más altas puede mitigar estos errores, aunque implica un mayor costo computacional (véase Anexo V: Código de distribución de error para una resolución).

Una vez se han tenido en cuenta las consideraciones previas para el desarrollo y cómo se distribuye el error para una misma resolución, el siguiente análisis consiste en cuantificar el error que se comete a la hora de posicionarse dentro de una celda H3. Como es previsible, el error asociado a resoluciones bajas será significativamente mayor que el correspondiente a resoluciones altas. Esto se debe a que, independientemente de la ubicación exacta dentro de la celda, el sistema de indexado espacial de Uber considera que dicho punto se encuentra en el centro de la celda correspondiente y, por tanto, si la celda es más grande se comete más error.

Para ilustrar esta afirmación, se ha generado un ejemplo utilizando las coordenadas GPS del dique exterior de la Escuela Naval Militar ($42,3989964^{\circ}$ N, $8,7049860^{\circ}$ W). A continuación, se ha asociado dicho punto con sus respectivas celdas en resoluciones 5 y 6, lo que permite observar la diferencia de posición entre el punto real y los centros de las celdas asignadas. El resultado de este ejemplo puede verse reflejado en la Figura 3-12:

- La celda de resolución 5 es la 853920c7ffffff y el error de posición es de 5,3729 km.
- La celda de resolución 6 es la 863920c57ffff y el error de posición es de 2,5011 km.

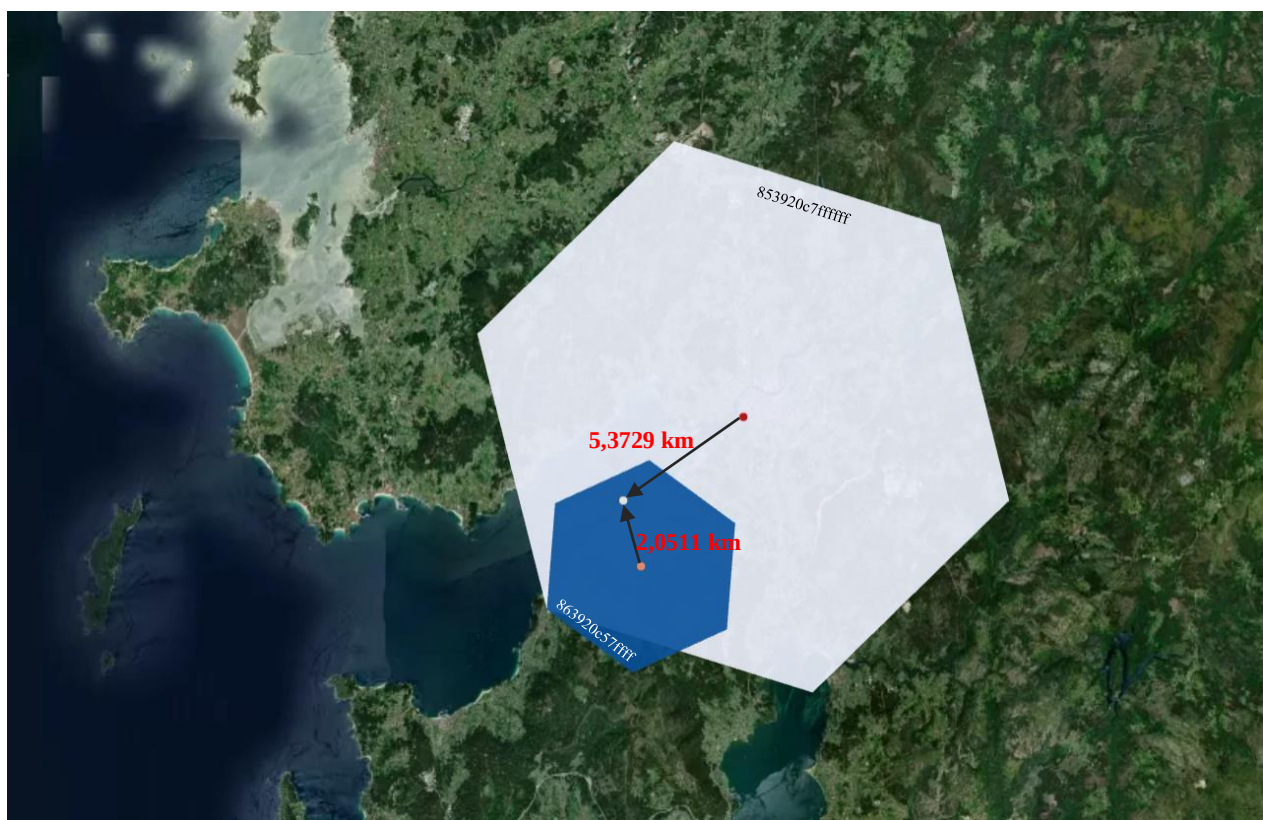


Figura 3-12 Diferencia de error para dos resoluciones

Una vez comprendidos los fundamentos teóricos y los cálculos subyacentes a este experimento, se procede a realizar un análisis estadístico basado en un número suficientemente elevado de puntos para obtener resultados concluyentes. En concreto, este análisis se lleva a cabo utilizando 1000000 puntos aleatorios ubicados exclusivamente en el mar. Tal como se ha mencionado en capítulos anteriores, el sistema Uber H3 en función de la resolución elegida, diseñado originalmente para aplicaciones terrestres, presenta mayores inexactitudes en entornos marítimos debido a la irregularidad en el tamaño de sus celdas.

La Figura 3-13 representa el error medio de posición comparado al error teórico para cada resolución. Este valor viene dado por la distancia media de un punto al centro de un hexágono [38], donde R en este caso se corresponde con la longitud del lado del hexágono para dicha resolución H3 (Tabla 2-2) (3).

$$R \left(\frac{1}{3} + \frac{\ln 3}{4} \right) \approx (0.6079864)R \quad (3)$$

Los resultados mostrados en la Figura 3-13 y Tabla 3-2 Tabla 3-2 Resultados de error de posición por resolución reflejan que la diferencia respecto al caso ideal es inapreciable; a pesar de haber celdas que, prácticamente, son el doble de tamaño respecto a otras para una misma resolución como se aprecia en la Tabla 2-3. Esto se debe al gran volumen de puntos analizados y el impacto ínfimo de la irregularidad del tamaño de las celdas.

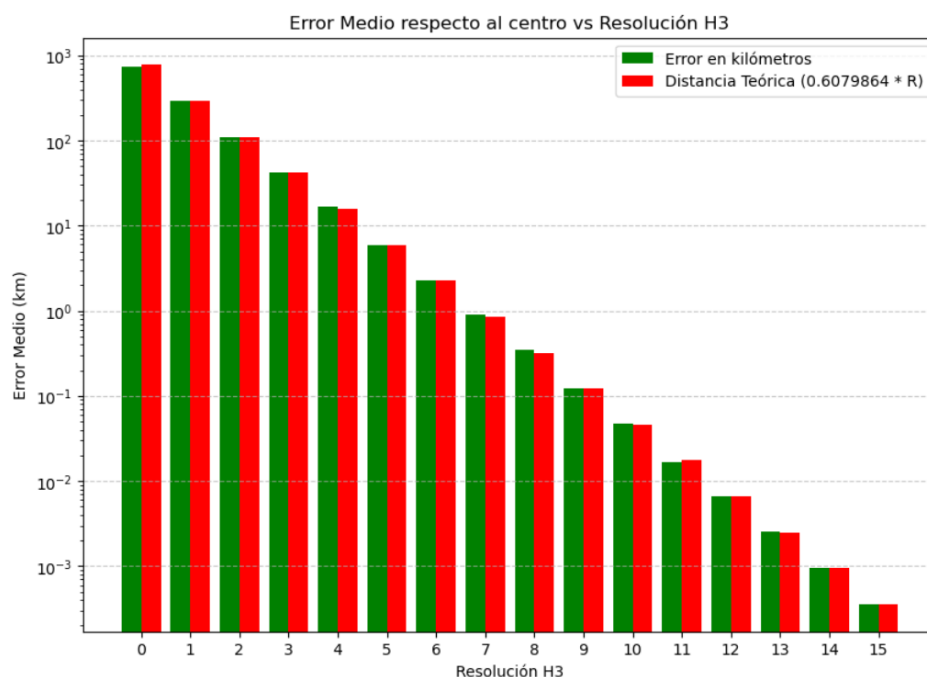


Figura 3-13 Gráfica de error por resolución

Resolución H3	Error Medio (km)	Error Teórico (km)
0	780,349269	778,986230
1	296,087057	293,691989
2	111,962260	110,965395
3	42,347707	41,938429
4	15,992024	15,851275
5	6,044534	5,991153
6	2,285682	2,264465
7	0,863585	0,855118
8	0,326376	0,323092
9	0,123362	0,122075
10	0,046661	0,046124
11	0,017626	0,017427
12	0,006660	0,006585
13	0,002519	0,002488
14	0,000952	0,000940
15	0,000360	0,000355

Tabla 3-2 Resultados de error de posición por resolución

La Figura 3-13 muestra la relación entre la resolución H3 (eje horizontal, niveles 0 a 15) y el error medio respecto al centroide (eje vertical, en escala logarítmica). A medida que la resolución aumenta,

las celdas hexagonales se subdividen, proporcionando mayor precisión en la representación espacial. Como resultado, el error medio disminuye progresivamente, ya que las celdas de menor tamaño abarcan áreas más reducidas y permiten una localización más exacta de los puntos.

Este descenso del error, sin embargo, no es uniforme. En las resoluciones más bajas, la reducción del error es más pronunciada, mientras que en las más altas la mejora se atenúa, indicando una disminución marginal del error frente a un aumento significativo en el coste computacional. Esta tendencia refleja el equilibrio entre precisión y eficiencia característico de los sistemas de indexación espacial.

Desde el punto de vista geométrico, el error medio disminuye en función de la raíz cuadrada de 7 por cada incremento de resolución. Esto se debe a que en H3 cada celda se subdivide en 7 celdas hijas y, dado que el área se reduce en una proporción de un séptimo la longitud característica (y por tanto la distancia media al centro) se reduce aproximadamente por un factor de $1/\sqrt{7}$. Matemáticamente, si E es error medio a una resolución n , se cumple la fórmula (4).

$$E_{n+1} \approx \frac{E_n}{\sqrt{7}} \quad (4)$$

En términos prácticos, la Figura 3-13 proporciona una referencia útil para determinar qué resolución es adecuada en función del nivel de precisión requerido. En aplicaciones donde la ubicación aproximada es suficiente, resoluciones intermedias pueden ofrecer un equilibrio adecuado entre precisión y eficiencia. Por otro lado, en contextos que exigen una localización más detallada, como análisis urbanos o estudios de movilidad, resoluciones más altas pueden ser preferibles, aunque con el costo asociado de mayor consumo de recursos computacionales (véase Anexo VI: Código de error teórico vs real por resolución).

Una vez analizados los errores de posición y su distribución para una misma resolución, otro paso necesario es determinar un área de interés específico y analizar cómo se distribuyen los errores de posición en esa área. La Figura 3-14 generada ilustra la precisión espacial del sistema de indexación H3 aplicado a una zona oceánica frente a la costa de Galicia, España. Los puntos representados en el mapa corresponden a coordenadas oceánicas que han sido analizadas en función de su error de posicionamiento al ser transformadas en celdas hexagonales de resolución 6. La visualización en Kepler.gl resalta los errores de posicionamiento mediante una escala de colores, donde colores oscuros corresponde a errores menores que los colores claros y las variaciones reflejan la diferencia entre la posición real y el centroide de la celda asignada por el índice H3.

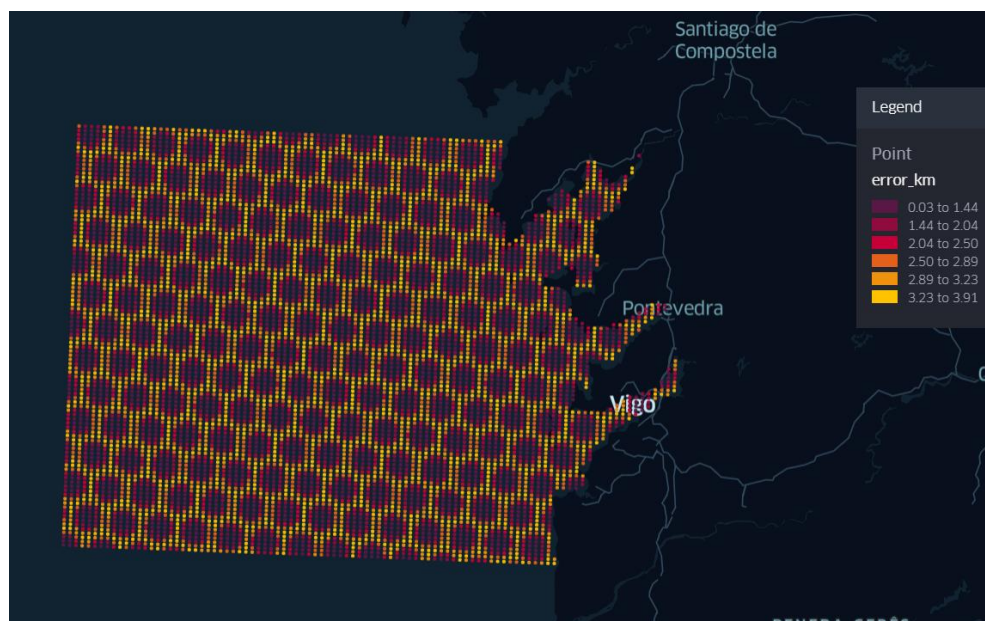


Figura 3-14 Distribución de error de posicionamiento para un área

Lo llamativo del resultado de la Figura 3-14 es el patrón tan regular y uniforme que se observa en la distribución del error para esta área. Esto se debe a que para esta zona geográfica de latitudes medias el sistema no presenta, a penas, deformaciones. No ocurre lo mismo en la Figura 3-15, donde se observa una zona del globo en la que se hace coincidir con uno de los 12 pentágonos del sistema. Por tanto, las celdas hexagonales alrededor del pentágono presentarán más irregularidades y no un patrón tan regular.

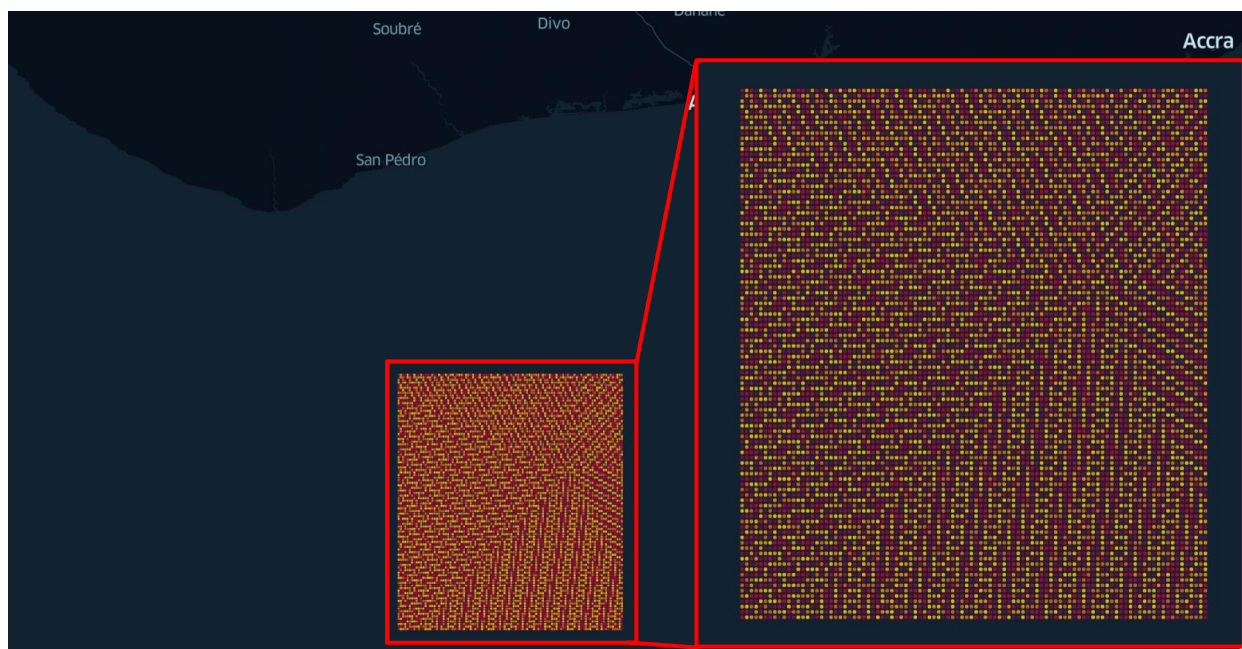


Figura 3-15 Distribución de error de posicionamiento irregular en un área

Para concluir el análisis de posicionamiento, se ha llevado a cabo una pequeña comparativa con el sistema de indexación Google S2, en la que se pretende constatar que la precisión obtenida mediante el sistema de Uber H3 presenta ligeras discrepancias. Es importante señalar que, a diferencia de evaluaciones previas, en esta ocasión se considera, tanto puntos ubicados en tierra como en el océano, lo que justifica las variaciones observadas en los resultados respecto a los obtenidos anteriormente.

En el desarrollo de este estudio, se ha optado por la generación de un millón de puntos distribuidos aleatoriamente en el espacio geográfico. Dado que los niveles de resolución de ambos sistemas no presentan una correspondencia directa, ha sido necesario seleccionar aquellos cuya granularidad de celdas se asemeje en la mayor medida posible, tomando como referencia los valores establecidos en las tablas presentadas en capítulos anteriores. Este criterio ha permitido garantizar que la evaluación se lleve a cabo bajo condiciones comparables, evitando sesgos derivados de diferencias inherentes en la estructura de discretización espacial de cada sistema (véase Anexo VII: Código para visualizar distribución de error de un área en un mapa en Kepler.gl).

La Figura 3-16 representa una comparativa del error de posicionamiento entre los sistemas de indexación geoespacial Uber H3 y Google S2, evaluando cómo varía la precisión conforme aumenta la cantidad de celdas en cada sistema. Se ha utilizado una escala logarítmica en ambos ejes, lo que permite visualizar con mayor claridad la relación entre la granularidad del sistema y la precisión en la geolocalización. En el eje X, se muestra el número total de celdas en cada resolución, que crece exponencialmente a medida que se subdivide el espacio en niveles más altos. Mientras que en el eje Y, se representa el error medio de posicionamiento en kilómetros, que tiende a disminuir conforme la cantidad de celdas aumenta y las regiones geográficas se indexan con mayor precisión.

Los puntos azules, marcados con “x”, representan los errores del sistema H3, mientras que los puntos rojos, identificados con “o”, corresponden a los errores del sistema S2. Además, junto a cada punto se ha etiquetado la resolución a la que pertenece, permitiendo ver cómo cada sistema escala a diferentes niveles de granularidad.

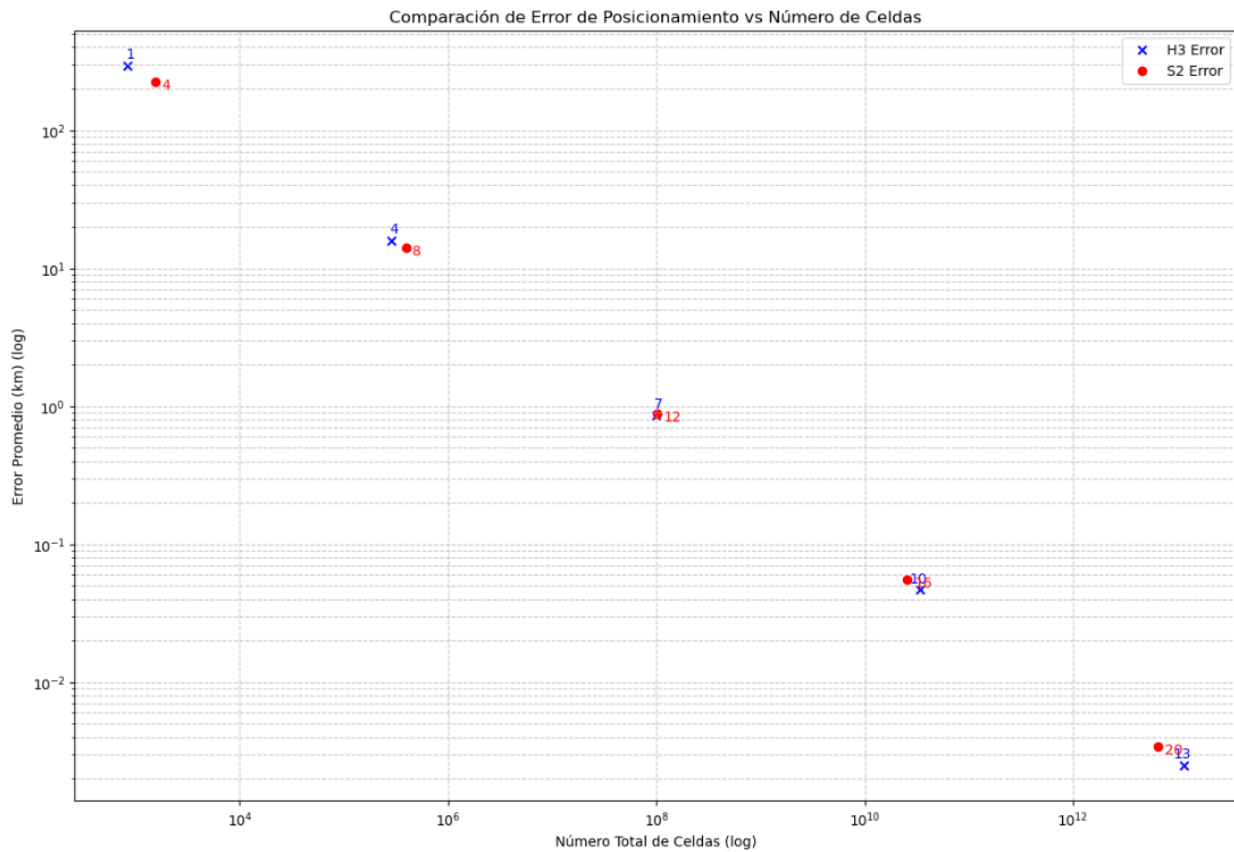


Figura 3-16 Comparativa de error de posición entre H3 y S2

RESOLUCIÓN		Nº DE CELDAS		TAMAÑO MEDIO DE CELDA (km ²)		ERROR DE POSICIÓN (km)	
H3	S2	H3	S2	H3	S2	H3	S2
1	04	842e0	154e1	609,788.44	332,074.27	296.08706	224.51659
4	08	288e3	393e3	1,770.34	1,297.17	15.99202	14.03633
7	12	988e5	100e6	5.16	5.07	0.86359	0.87739
10	16	338e8	257e8	0.01504	0.0197	0.04666	0.054820
13	20	116e11	659e11	0.00004	0.00008	0.00252	0.003426

Tabla 3-3 Comparación de precisión entre H3 y S2

Un aspecto clave que se puede interpretar en la gráfica es la relación entre la eficiencia y la precisión. En resoluciones bajas, S2 parece ser más preciso que H3, pero conforme se avanza en granularidad, las diferencias se reducen y en algunos casos H3 ofrece un error menor con una cantidad menor de celdas. Esto sugiere que H3 podría ser más eficiente en ciertas aplicaciones donde se requiere un equilibrio entre precisión y cantidad de datos manejados, mientras que S2 puede escalar más agresivamente, generando más subdivisiones en niveles altos para alcanzar una precisión extrema.

La disposición de los puntos en la gráfica también evidencia que no hay una alineación directa entre las resoluciones de H3 y S2, lo que confirma que, aunque ambos sistemas intentan aproximar subdivisiones espaciales equivalentes, sus escalas no son idénticas. Esto explica por qué en algunos puntos el error de S2 es mayor o menor que el de H3, dependiendo de la cantidad de celdas utilizadas en cada sistema (véase Anexo VIII: Código de comparación Uber H3 con Google S2).

A pesar de haberse comparado ambos sistemas a través de las resoluciones, esto no es fácil porque las resoluciones no son equivalentes. Por ello, un análisis más relevante es calcular cuál es el cociente entre la distancia media al centro del hexágono entre el área del hexágono frente al mismo cálculo para el cuadrado. Este análisis queda resumido en la Tabla 3-4.

Geometría	Distancia Media $\bar{d}$	Área A	Cociente $\bar{d}/A$
Hexágono	$0,6089864 \cdot R$	$\frac{3\sqrt{3}}{2} \cdot R^2$	$\approx \frac{0,324}{R}$
Cuadrado	$0,7652 \cdot R$	$4 \cdot R^2$	$\frac{0,1913}{R}$

Tabla 3-4 Comparación de geometría entre hexágono y cuadrado

En este análisis, la letra R representa una medida del tamaño de las figuras. En el hexágono es la distancia desde el centro hasta un vértice, es decir, su radio. En el cuadrado, R es la mitad del lado, ya que la figura se extiende de -R a R en cada eje. Utilizar R como referencia permite comparar ambas geometrías de forma proporcional, sin depender del tamaño absoluto, lo que es clave para evaluar su eficiencia espacial.

La interpretación de estos resultados (obtenidos a partir de [38]) se refiere al mero análisis geométrico, pues tiene implicaciones prácticas notables en el ámbito de la indexación espacial, particularmente al evaluar la eficiencia de los sistemas de Uber H3 y Google S2. El cociente representa una medida normalizada de dispersión espacial, que permite comparar la eficiencia con la que cada celda geométrica contiene puntos respecto a su centro, independientemente de su escala. Un cociente mayor implica que, por unidad de área, los puntos están más próximos al centroide, lo que se traduce en una mejor precisión espacial.

En este contexto, Uber H3, que emplea celdas hexagonales para la discretización de la superficie terrestre, presenta una ventaja inherente respecto a Google S2, basado en celdas cuadradas proyectadas sobre una esfera. La mayor eficiencia del hexágono, reflejada en su cociente teórico más alto, sugiere que las celdas H3 minimizan mejor la dispersión espacial y, por tanto, los errores de cuantificación asociados a la representación de datos geoespaciales. Esta propiedad se vuelve crucial al trabajar con grandes volúmenes de información, como los generados por el sistema AIS, donde la precisión en la ubicación de los datos puede afectar significativamente a la calidad del análisis.

3.4 Análisis de cobertura para un área de interés

Otro aspecto relevante en el tratamiento de datos mediante un sistema de indexación geoespacial es el análisis de área desde una posición específica; por ejemplo, el área circundante a un barco. Dado que la geometría principal del sistema de Uber se basa en hexágonos, representar un área circular alrededor de una ubicación fija introduce un error en la estimación (Figura 3-17).

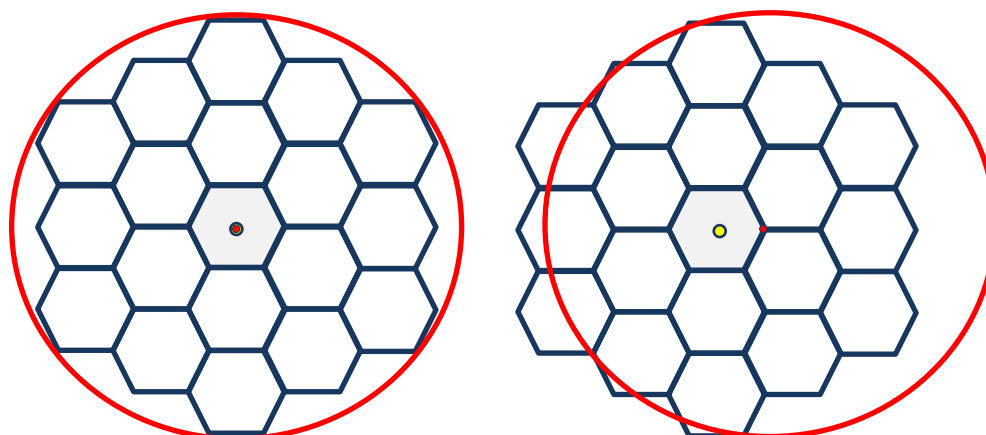


Figura 3-17 Representación de error mínimo y máximo para un disco de "k" celdas

La Figura 3-17 muestra el error que se comete al aproximar un círculo mediante un disco de "k" saltos de celdas H3. En ambos casos, se genera una agrupación de hexágonos a partir de un punto inicial

dentro de la estructura H3, cubriendo aparentemente la misma área total que el círculo. Sin embargo, la ubicación del punto de referencia dentro de la celda inicial afecta significativamente la precisión de la cobertura.

En la representación de la izquierda, el disco de celdas se construye tomando como punto de referencia el centro geométrico de la celda inicial. Esto permite que la distribución de los hexágonos se alinee mejor con la región circular de interés, minimizando la discrepancia entre el área real del círculo y la suma de las áreas de las celdas H3.

En la representación de la derecha, el disco se genera tomando como referencia un vértice de la celda inicial en lugar de su centro. A pesar de que la suma de las áreas de las celdas sigue siendo la misma que en la izquierda, la distribución de las celdas está desplazada y no cubre correctamente la región de interés. Esto significa que, a pesar de estar utilizando el mismo número de celdas y la misma área total, una porción significativa de la cobertura mediante celdas se encuentra en una ubicación errónea, generando un error en la representación de la superficie.

Este efecto pone de manifiesto que, en indexación espacial con H3, no solo es importante la cantidad de área cubierta, sino también su correcta ubicación. Un desplazamiento en el punto de referencia puede llevar a que el área cubierta no se alinee con la región que realmente se desea representar, afectando la precisión de los cálculos espaciales.

Para abordar esta discrepancia, el primer paso es desarrollar un método para aproximar un área circular mediante celdas H3. El método utilizado consiste en seleccionar un conjunto de celdas hexagonales H3 que cubran un área circular en torno a una ubicación geográfica dada, asegurándose de que solo se incluyan aquellas cuyo centro se encuentra dentro del radio especificado. Para ello, sigue un proceso iterativo de expansión en anillos concéntricos, refinando la selección de celdas en cada paso hasta que se alcanza el límite de cobertura.

A partir de unas coordenadas concretas, se define un radio de cobertura en kilómetros y una resolución para las celdas H3. La celda central de H3 correspondiente a la ubicación inicial se identifica y se utiliza como punto de partida para la expansión.

```
h3_grid_disk(celda_central, k)
```

A partir de la celda central, el código expande la cobertura en anillos concéntricos, explorando en cada paso un nuevo conjunto de celdas situadas en el perímetro de la zona ya explorada. Sin embargo, en lugar de incluir todas las celdas de un anillo sin distinción y solo aquellas cuyo centro se encuentra dentro del círculo especificado son añadidas a la selección final. Si dentro de un mismo anillo algunas celdas cumplen el criterio y otras no, se eliminan las que no lo cumplen y el proceso continúa expandiendo el área de cobertura. Cuando un anillo completo no contiene ninguna celda cuyo centro esté dentro del radio, se considera que el área ha sido completamente cubierta y se detiene la expansión.

Como se observa en la Figura 3-18, incluso en el peor de los casos, cuando la posición generada coincide con uno de los vértices de la celda central, el área de interés (círculo) queda cubierta por las celdas H3.

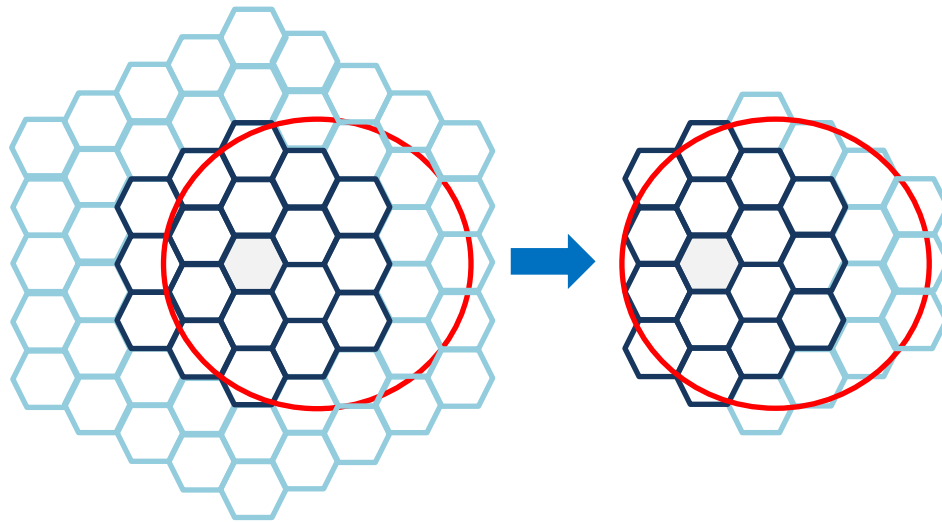


Figura 3-18 Área cubierta con k anillos y después filtrada con celdas cuyo centro está en el interior del círculo

Al finalizar la selección de celdas, se calcula el área total cubierta sumando el área individual de cada celda aceptada. Finalmente, se comparan el área real del círculo y el área cubierta por las celdas seleccionadas, y se imprime el número total de celdas utilizadas junto con los valores obtenidos

Para visualizar el impacto de la resolución en la cobertura de un área determinada, se ha realizado un análisis comparativo utilizando distintas resoluciones del sistema de indexación H3. La ubicación de referencia se ha establecido en 40,00° N 010,00° O, con un área de interés definida por un círculo de 15 kilómetros de radio, lo que equivale a aproximadamente 706,86 km² (Figura 3-19).

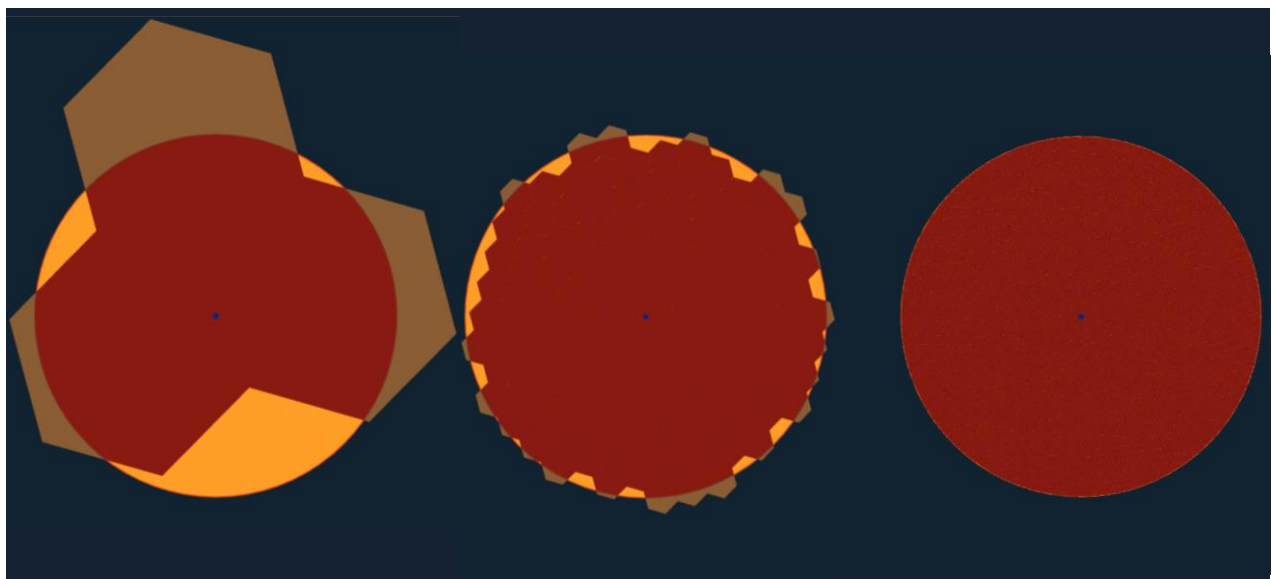


Figura 3-19 Representación de cobertura para el mismo área para resoluciones 5, 7 y 10

En la Figura 3-20, se observa cómo varían el número de celdas utilizadas y el área cubierta en función de la resolución. A medida que la resolución aumenta, se produce un crecimiento en la cantidad de celdas necesarias para representar el mismo área; como cada celda se subdivide en 7, este crecimiento sigue una relación de por 7 respecto a la resolución anterior. Mientras que en resolución 5 solo se requieren 3 celdas, en resolución 10 esta cifra se incrementa drásticamente hasta 42450 celdas, lo que supone un aumento significativo en la carga computacional y la gestión de datos espaciales.

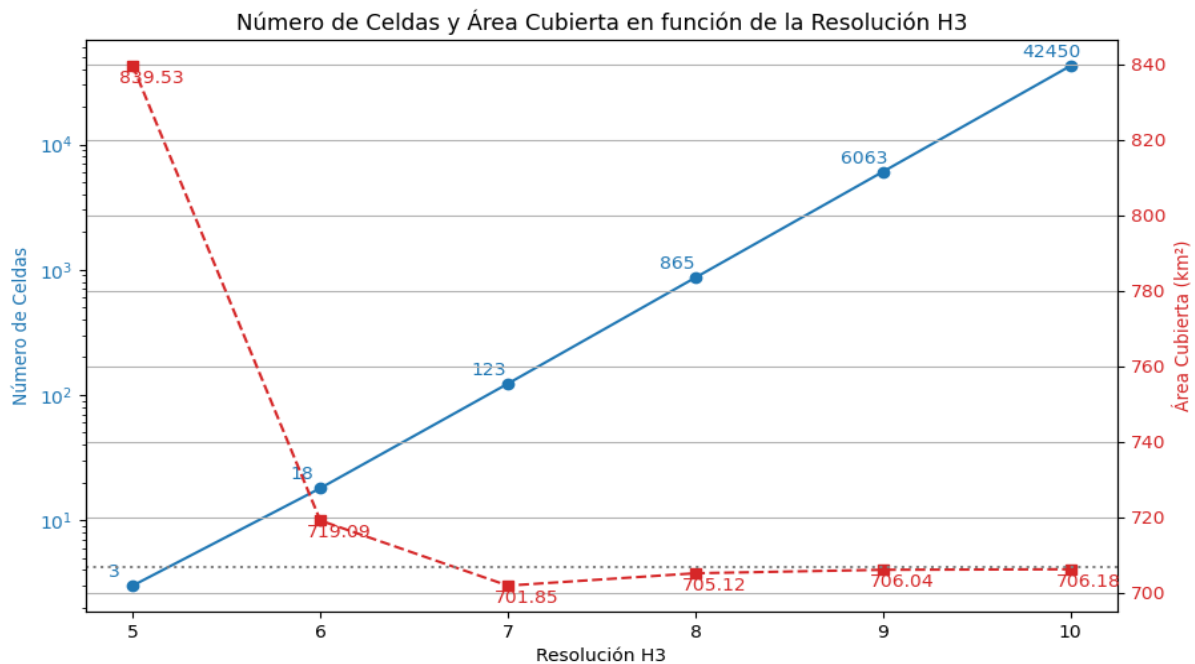


Figura 3-20 Gráfica área cubierta y número de celdas por resolución

Al mismo tiempo, el análisis revela cómo la precisión del área cubierta mejora progresivamente con resoluciones más altas. En resoluciones bajas, la sobreestimación del área es considerable, como se evidencia en el caso de la resolución 5, donde el área cubierta alcanza los 839,53 km², muy por encima del valor real de 706,86 km². Sin embargo, a medida que la resolución aumenta, la diferencia entre el área cubierta y el área real se reduce, alcanzando valores cercanos a la cobertura exacta a partir de la resolución 7. A partir de este punto, la precisión se mantiene estable, logrando minimizar la diferencia con respecto al área real.

Este análisis pone en evidencia el equilibrio que debe alcanzarse entre la precisión y la eficiencia computacional. Mientras que resoluciones bajas requieren un menor procesamiento, pero comprometen la exactitud de la cobertura, resoluciones más altas garantizan una representación más precisa del área de interés, aunque con un alto costo computacional derivado del incremento en la cantidad de celdas a gestionar. Por tanto, la elección de la resolución óptima dependerá de la necesidad específica de cada aplicación, buscando siempre el mejor balance entre exactitud y capacidad de procesamiento (véase Anexo IX: Código para comparación de área real vs área cubierta por celdas Uber H3).

Los resultados obtenidos muestran una tendencia clara en la relación entre la cantidad de celdas y el área cubierta. A partir de cierto punto, aumentar la resolución solo genera divisiones más pequeñas sin mejorar la utilidad del modelo para el análisis geoespacial. Además, las visualizaciones confirman que las resoluciones seleccionadas ya permiten capturar la distribución espacial de manera efectiva. Explorar resoluciones aún mayores no aportaría información nueva, sino que haría el procesamiento más complejo y costoso.

Otro aspecto clave en el análisis del área es evaluar qué proporción de la superficie total ajustada realmente corresponde a la zona de interés. En otras palabras, es fundamental determinar cuánta extensión se encuentra fuera del radio establecido y qué fracción del círculo permanece sin cobertura.

Como se aprecia en la Figura 3-21, la suma de las áreas de los hexágonos utilizados puede aproximarse en cierta manera al área teórica del círculo. Sin embargo, esto no implica necesariamente un ajuste perfecto. Es evidente que, en ciertas resoluciones, una parte significativa de la superficie cubierta por H3 excede el límite del radio de interés, generando una sobreestimación en la cobertura efectiva.

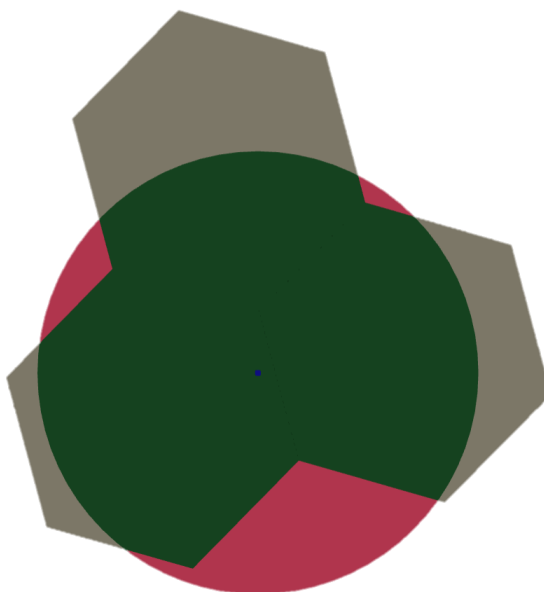


Figura 3-21 Representación de área cubierta en resoluciones bajas

Para comprender mejor esta distribución, en el presente análisis se cuantifica la fracción de los hexágonos que queda fuera del área objetivo y la cantidad de superficie del círculo que, a pesar del ajuste, permanece sin cubrir. Estos valores permiten evaluar con mayor precisión el grado de ajuste de la discretización hexagonal y su eficacia en la representación de un área circular. Para dicho análisis, se generan 100000 localizaciones manteniendo en todas ellas el radio de interés 15 km (como el experimento particularizado anterior) obteniendo en todas esas ubicaciones un área de cobertura de interés de 706,86 km².

Los primeros pasos del código son idénticos al experimento anterior ya que los anillos concéntricos se generan de la misma manera y el criterio de eliminación de celdas es el mismo: que el centro de la celda esté dentro del radio de interés.

Una vez determinadas las celdas que participan en la cobertura, se procede a calcular con mayor detalle qué fracción de cada hexágono realmente queda dentro del área circular. Para ello, se consideran tres posibles escenarios: cuando todos los vértices de la celda están dentro del círculo, cuando todos están fuera (que no debería ocurrir) y cuando la celda se encuentra parcialmente cubierta. En los dos primeros casos, el cálculo se simplifica considerablemente, asignando la totalidad del área del hexágono a la región correspondiente. En el tercer caso, en cambio, se emplea una integración numérica para estimar con mayor precisión qué parte del hexágono pertenece al círculo de interés. Si por algún motivo la integración falla, se utiliza una aproximación basada en la cantidad de vértices dentro del radio.

Por tanto, para el último caso mencionado se utiliza una integración doble (a pesar de poder hacerse mediante una simulación Montecarlo, aunque sería menos preciso) para evaluar el área contenida del hexágono con el círculo. Se evalúa la función característica (5) que devuelve 1 si el punto pertenece al círculo y al hexágono y 0 en caso contrario.

$$f(x, y) = \begin{cases} 1, & \text{si } (x, y) \in \text{hexágono} \cap \text{círculo} \\ 0, & \text{en otro caso} \end{cases} \quad (5)$$

Para llevar a cabo la integración, primero es necesario definir los límites dentro de los cuales se evaluará la función característica. En este caso, esos límites corresponden al rectángulo que envuelve por completo al hexágono. Para determinarlos, se emplea la función *Polygon* de la biblioteca *Shapely*, que permite convertir los vértices del hexágono en una estructura poligonal manejable. Una vez construido el polígono, se utiliza el método *bounds*, el cual devuelve las coordenadas extremas en

términos de latitud y longitud (Figura 3-22). Estas coordenadas representan los valores mínimo y máximo en cada eje, definiendo así el área **(6)** dentro de la cual se realizará la integración numérica.

$$A_{interior} = \int_{lng_mín}^{lng_máx} \int_{lat_mín}^{lat_máx} f(x, y) dy dx \quad (6)$$

Si la integración fallara por cualquier motivo, se usa una aproximación basada en los vértices dentro del círculo; asumiendo que la fracción de vértices dentro del círculo es un buen estimador de la proporción del área cubierta.

Por tanto, para hallar la fracción de área que, efectivamente, está dentro del círculo dividiendo el área obtenida por integración entre el área del rectángulo envolvente.

$$proporción\ dentro = \frac{Área\ integrada}{Área\ Bounding\ Box} \quad (7)$$

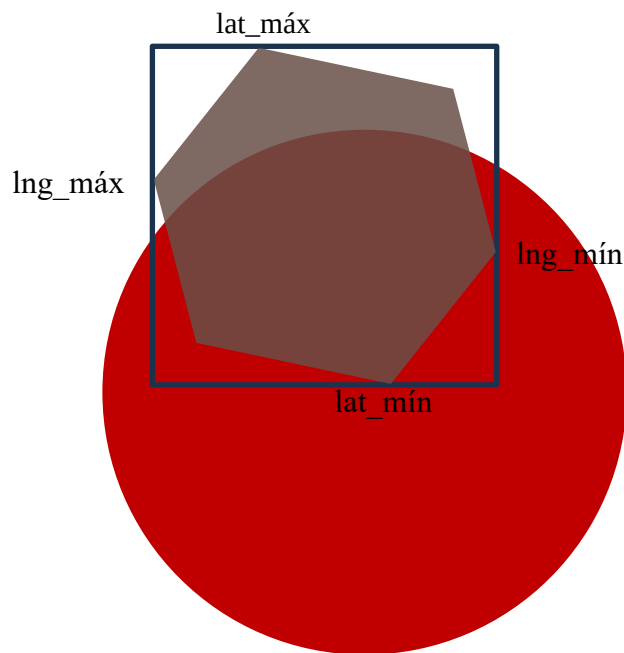


Figura 3-22 Representación de la *bounding box* para los límites de integración

Una vez remarcados los aspectos clave para el cálculo del análisis de la cobertura de un área de interés utilizando celdas Uber H3 se obtienen una serie de resultados reflejados en la Tabla 3-5 y Tabla 3-6, que evidencian una relación directa entre la resolución utilizada y la precisión de la representación. A medida que la resolución aumenta, el sistema logra adaptarse mejor a la forma del área objetivo, reduciendo tanto la falta de cobertura, es decir, las partes del círculo que no quedan cubiertas por ninguna celda, como el exceso de cobertura, que corresponde a las regiones fuera del círculo que quedan incluidas dentro de las celdas seleccionadas. Sin embargo, esta mejora en la precisión tiene un coste, ya que viene acompañada de un incremento significativo en la cantidad de celdas necesarias, lo que repercute en el costo computacional del proceso.

En resoluciones bajas, donde las celdas son de gran tamaño, la representación del área de interés es imprecisa y genera grandes ineficiencias. Debido a la geometría hexagonal de las celdas, muchas de ellas terminan incluyendo porciones sustanciales de territorio fuera del área de interés, al mismo tiempo que dejan sin cubrir zonas dentro del círculo. A medida que se incrementa la resolución, el tamaño de las celdas disminuye, lo que permite un ajuste más fino y una distribución más precisa, reduciendo progresivamente los errores de cobertura. Sin embargo, a partir de cierto punto, el beneficio de aumentar la resolución se vuelve menos significativo en comparación con el costo que implica procesar una cantidad mucho mayor de celdas.

Resolución	N° de celdas	Área cubierta (km ²)	Área dentro del círculo (km ²)	Área no cubierta del círculo (km ²)	Área fuera del círculo (km ²)
5	2.75	714.821	398.329	308.529	316.419
6	19.40	703.809	569.194	137.667	134.615
7	132.82	705.385	652.158	54.701	53.227
8	975.20	706.172	685.303	21.555	20.869
9	6652.62	706.259	698.414	8.444	7.846
10	47147.10	705.839	702.869	4.004	2.970

Tabla 3-5 Cobertura de áreas por resolución

Resolución	Área real	Área cubierta (km ²)	% De área dentro del círculo respecto al área cubierta	% De área no cubierta respecto al círculo	% De área fuera respecto al área cubierta
5	706.858	714.821	55.72 %	43.65 %	44.27 %
6		703.809	80.87 %	19.48 %	19.13 %
7		705.385	92.45 %	7.74 %	7.55 %
8		706.172	97.04 %	3.05 %	2.96 %
9		706.259	98.89 %	1.19 %	1.11 %
10		705.839	99.58 %	0.57 %	0.42 %

Tabla 3-6 Porcentajes de cobertura de área por resolución

El aspecto computacional es un factor crítico en este tipo de análisis. Para resoluciones bajas, el procesamiento es rápido y los resultados se obtienen en cuestión de segundos o minutos. Sin embargo, en resoluciones altas, la cantidad de celdas crece de manera que los cálculos requieren tiempos de ejecución prolongados, pasando de unos minutos a varias horas. Esto se debe a que cada celda debe procesarse individualmente, evaluando su ubicación con respecto al área de interés y, en muchos casos, aplicando métodos de integración para determinar qué fracción de su superficie se encuentra dentro del círculo. Este crecimiento en el número de cálculos hace que el tiempo de ejecución se dispare, convirtiéndose en un factor limitante en el uso de resoluciones muy altas (véase Anexo X: Código para obtener porcentaje de cobertura respecto al círculo real).

Para complementar los resultados mostrados anteriormente, se realiza una grafica con los distintos porcentajes de cobertura. Esto gráfica se ve en la Figura 3-23.

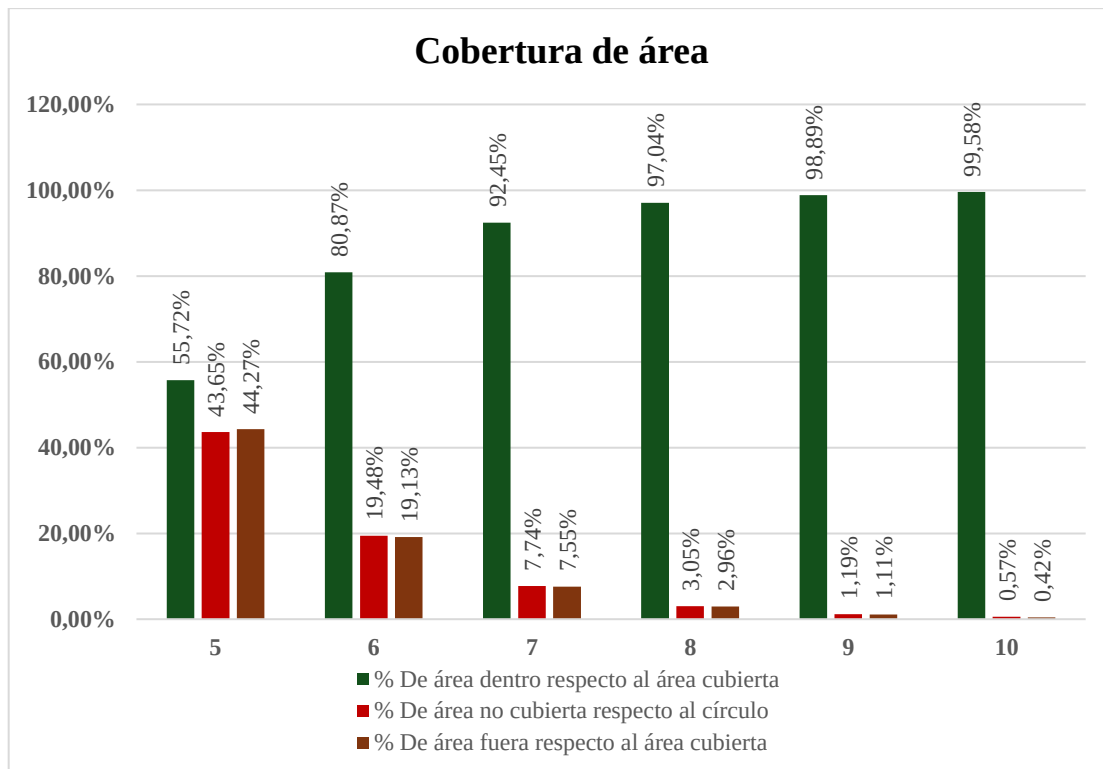


Figura 3-23 Gráfica con los resultados de cobertura según resolución

El análisis sugiere que existe un punto de equilibrio entre la precisión y el costo computacional. En resoluciones intermedias, la representación del área de interés es lo suficientemente precisa sin necesidad de manejar cantidades masivas de celdas. En cambio, optar por resoluciones más altas solo es justificable cuando se requiere una precisión extrema y se dispone de recursos computacionales adecuados para soportar el procesamiento. En definitiva, la elección de la resolución debe basarse en la necesidad específica del análisis, buscando un balance entre la calidad de la cobertura y la viabilidad computacional del proceso.

3.5 Análisis de trayectoria

El uso de celdas Uber H3 para representar trayectorias introduce una serie de problemas que afectan tanto la dirección como la velocidad de los objetos en movimiento. A diferencia de un modelo continuo basado en coordenadas geográficas precisas, la discretización espacial en hexágonos provoca que la trayectoria real quede distorsionada, generando desviaciones en la estimación de los parámetros de navegación.

Uno de los problemas más evidentes es la alteración del rumbo. En una trayectoria normal, el desplazamiento de un objeto sigue una línea recta entre dos puntos, determinando así su dirección real sobre el terreno. Sin embargo, al usar celdas hexagonales para modelar este movimiento, la ruta resultante no es una línea continua, sino una secuencia de transiciones entre centros de celdas adyacentes. Como consecuencia, la trayectoria se ve fragmentada en segmentos con cambios bruscos de dirección, lo que genera una distorsión en la estimación del rumbo. En la Figura 3-24, esta diferencia es clara: la línea azul representa el recorrido real, mientras que la secuencia en rojo muestra cómo el sistema de celdas introduce una ruta en zigzag que no refleja con precisión el movimiento del objeto.

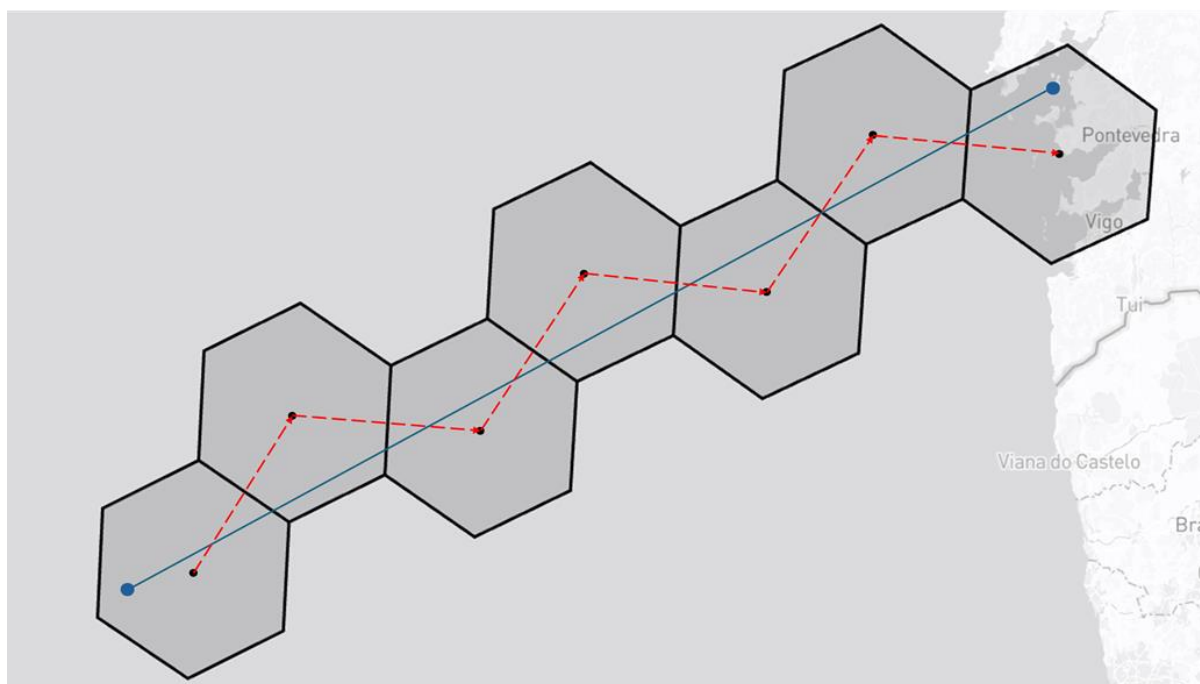


Figura 3-24 Ejemplo de trayectoria a través de celdas

Este experimento comienza analizando únicamente una única una trayectoria de ejemplo, para posteriormente generalizar el comportamiento posterior con un número más elevado. Primeramente, por tanto, se genera una trayectoria a partir de dos puntos. Estos puntos se asocian a su correspondiente celda H3 y para recorrer todas las celdas de dicha trayectoria se usa una función específica de la API de H3. En esta se ingresa la celda inicial y final y como resultado se obtiene un camino más corto de celdas adjuntas.

```
h3.grid_path_cells(celda_inicial, celda_final)
```

Por otra parte, es necesario ir calculando el rumbo para comprobar cuál es la desviación que se comete respecto al rumbo real, que será aquel reportado en el campo COG de los mensajes AIS. Para ello, el cálculo de rumbo (independientemente si es el real o los intermedios) se calcula a partir de las coordenadas GPS (latitud y longitud) iniciales y finales mediante las siguientes fórmulas, donde λ son las longitudes y φ las latitudes:

$$\Delta\lambda = \lambda_f - \lambda_i \quad (8)$$

$$\Delta\varphi = \varphi_f - \varphi_i \quad (9)$$

$$\theta(^{\circ}) = \arctan\left(\frac{\Delta\lambda}{\Delta\varphi}\right) \cdot \frac{180}{\pi} \quad (10)$$

$$Rumbo = (\theta + 360) \% 360 \quad (11)$$

Tras definir tanto los cálculos necesarios para la determinación de rumbos como la segmentación espacial con celdas H3, se genera trayectoria de referencia (Figura 3-25). Sin generalizar pero para graficar y asentar las bases del análisis, se considera, inicialmente, una trayectoria entre las coordenadas 41,33° N, 11.95° O y 42.33° N, 9.00° O, con un rumbo teórico de 71,27° y una distancia real de 269,03 km.

Para evaluar el impacto de la resolución H3 en la precisión de la trayectoria, se repiten los cálculos para distintas resoluciones, obteniendo los resultados que se presentan en la Tabla 3-7. Esta muestra que no existe una relación lineal entre la resolución de las celdas H3 y la mejora en la precisión del rumbo.

A medida que la resolución aumenta, la trayectoria se segmenta en un mayor número de celdas, lo que introduce múltiples cambios de dirección. Estos cambios, aunque pequeños, generan una acumulación de desviaciones que incrementan la variabilidad del rumbo final. Como resultado, en resoluciones más altas, la trayectoria presenta más ajustes direccionales que pueden afectar la estabilidad del cálculo del rumbo.

En cuanto a la distancia recorrida, se observa que con resoluciones más altas el número de celdas en la trayectoria crece significativamente. Esto hace que la distancia medida a través de los centros de las celdas sea generalmente mayor que la distancia real entre los puntos de inicio y fin. Sin embargo, la diferencia porcentual en la distancia no sigue un patrón constante con el aumento de la resolución, ya que la manera en la que las celdas se ajustan a la trayectoria real introduce distorsiones particulares en cada nivel de granularidad.

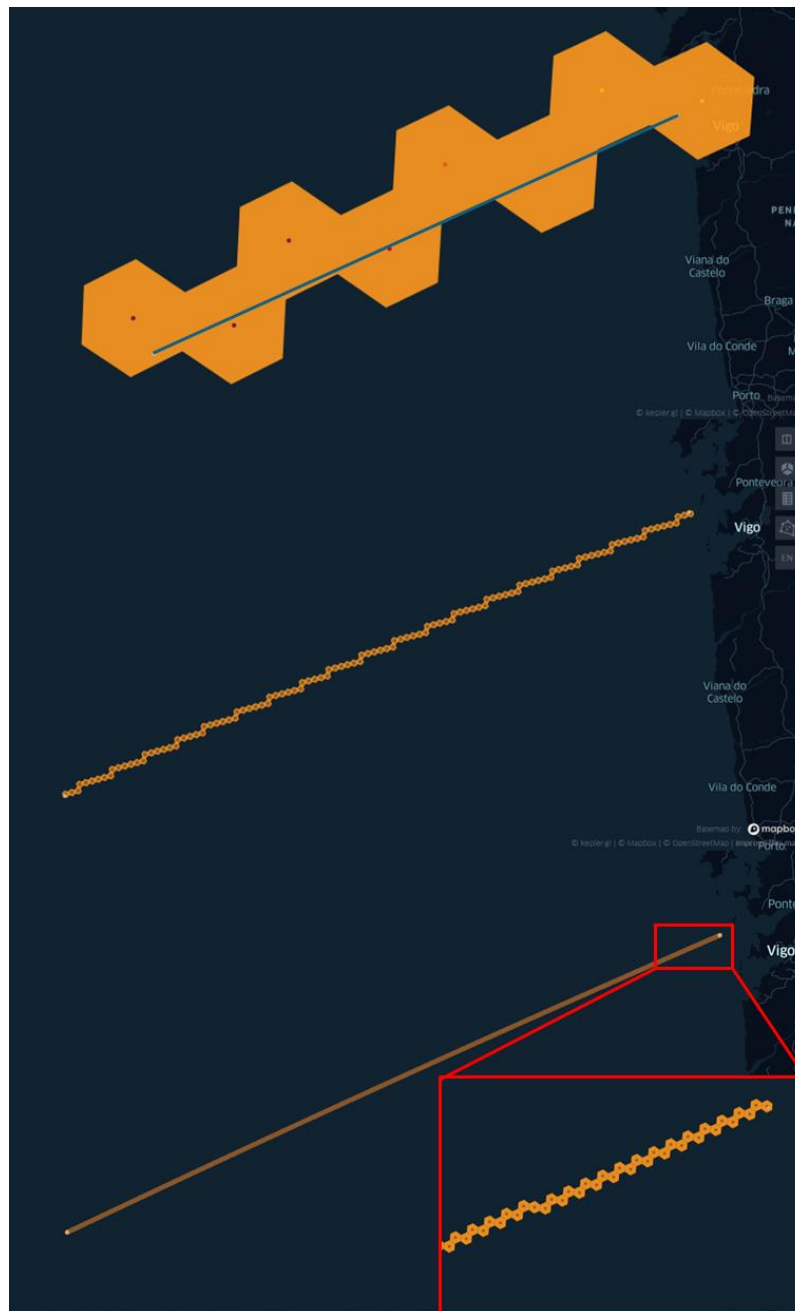


Figura 3-25 Representación de la misma trayectoria para las resoluciones 4, 7 y 10

Respecto al número de celdas utilizadas, se aprecia que este aumenta considerablemente con la resolución. A pesar de que una mayor cantidad de celdas podría sugerir una representación más detallada

de la trayectoria, en muchos casos esto introduce un exceso de cambios direccionales que incrementan la desviación del rumbo sin necesariamente mejorar la fidelidad de la ruta con respecto al trayecto original.

Por otra parte, dos de los resultados más relevantes obtenidos en el ejemplo de la trayectoria es la diferencia de distancia y la desviación de rumbo. El primero de ellos se obtiene sumando la distancia geodésica entre los centros de las celdas contiguas de la trayectoria, restándole la distancia geodésica real entre el punto inicial y final y dividiéndole por ella misma. Para la desviación de rumbo se obtienen los rumbos intermedios aplicando iterativamente la fórmula (11) y comprando cada rumbo intermedio entre el teórico. Por tanto, las fórmulas resultantes para el cálculo de disco porcentajes son respectivamente (12) y (13):

$$Desviación\ rumbo = \frac{1}{n} \sum_{i=1}^n \frac{\Delta\theta_i}{|\theta_{teórico}|} \times 100 \quad (12)$$

$$Diferencia\ distancia = \frac{\sum d_{H3} - d_{real}}{d_{real}} \times 100 \quad (13)$$

Resolución	Nº de celdas	Distancia recorrida por celdas	Diferencia distancia %	Desviación de rumbo %
4	8	329,79	22,58 %	35,84 %
5	17	284,08	5,59 %	22,67 %
6	48	316,38	17,60 %	36,49 %
7	116	291,63	8,40 %	21,82 %
8	327	313,52	16,54 %	36,41 %
9	808	292,38	8,68 %	22,11 %
10	2278	312,82	16,28 %	36,40 %

Tabla 3-7 Análisis de una trayectoria concreta

Este análisis pone de manifiesto la necesidad de encontrar un equilibrio entre la resolución óptima y la estabilidad direccional de la trayectoria. En aplicaciones geoespaciales y de navegación, seleccionar una resolución demasiado alta puede generar mayor ruido direccional sin necesariamente mejorar la representación del recorrido real (véase Anexo XI: Código para el cálculo de desvíos en una trayectoria concreta).

Para obtener con resultados más generales, se repiten los mismos cálculos, pero para trayectorias generadas aleatoriamente. Para ello, se generan 100000 trayectorias aleatorias de 500 km. Otro requisito a mayores, para mayor realismo, las trayectorias en latitudes polares se han reducido.

La Tabla 3-8 muestra cómo varían la cantidad de celdas utilizadas, la diferencia porcentual de distancia y la desviación de rumbo a medida que se incrementa la resolución en el sistema de celdas H3. Sin embargo, a pesar de la notable diferencia en el número de celdas entre resoluciones, los valores de desviación de rumbo y diferencia en distancia se mantienen prácticamente constantes.

Resolución	Nº de celdas	Diferencia distancia %	Desviación de rumbo %
4	13,15	10,27 %	11,59 %
5	33,10	10,32 %	11,54 %
6	85,92	10,32 %	11,52 %
7	225,87	10,41 %	11,61 %
8	595,59	10,12 %	11,31 %
9	1575,22	10,49 %	11,67 %
10	4175,53	10,52 %	11,37 %

Tabla 3-8 Análisis de 100000 trayectorias aleatorias de 500 km

Este comportamiento se debe a la naturaleza del sistema de indexación H3. Aunque el aumento de resolución provoca que la trayectoria se represente con más celdas, el sistema sigue aproximando la trayectoria mediante la conexión de los centros de las celdas. En resoluciones bajas, las celdas son más grandes, por lo que los saltos entre centros de celdas pueden generar desviaciones más perceptibles. Sin embargo, a medida que la resolución aumenta, los centros de las celdas están más próximos, y aunque hay más saltos en la trayectoria, el efecto acumulativo de la desviación no cambia significativamente.

Otro factor importante es que, en resoluciones altas, el mayor número de celdas no significa necesariamente una mejor representación de la trayectoria real, ya que las conexiones entre los centros de las celdas pueden generar múltiples pequeños cambios de dirección. Estos cambios pueden cancelar o equilibrar las desviaciones acumuladas, lo que explica la estabilidad de la desviación de rumbo en torno al 11%.

Respecto a la distancia, la diferencia porcentual se mantiene estable en torno al 10%, lo que indica que el método de conexión entre centros de celdas introduce un error sistemático constante en la distancia recorrida. Este error proviene de la estructura hexagonal de H3, que en ciertos ángulos de trayectoria puede generar rutas ligeramente más largas que la distancia real (véase Anexo XII: Código para cálculo de desvíos de trayectoria para una resolución).

Para entender mejor este fenómeno, es útil recordar que en la aproximación digital de una trayectoria se genera una “longitud digital” al sumar los desplazamientos entre centros de celdas. Por ejemplo, en una cuadrícula cuadrada, utilizando la métrica Manhattan [39], la distancia digital para una trayectoria de longitud L y ángulo θ se expresa como la expresión (14).

$$D_{Manhattan} = L \cdot (|\cos \theta| + |\sin \theta|) \quad (14)$$

Si se asume que el ángulo es uniformemente distribuido, el valor esperado $|\cos \theta| + |\sin \theta|$ resulta ser $\pi/4$, lo que implica que, en promedio, la distancia digital es un 27.3% mayor que la distancia euclidiana real. Sin embargo, el sistema H3, debido a su estructura hexagonal, permite moverse en seis direcciones y obtener una representación más suave de la dirección de la línea. En este caso se coge un ángulo entre 0° y 60° dando como resultado dos vectores $(1, 0)$ y $(1/2, \sqrt{3}/2)$. Por tanto, la dirección unitaria de $(\cos \theta, \sin \theta)$ se expresa como combinación lineal de estos dos vectores (15).

$$a(1, 0) + b\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right) = (\cos \theta, \sin \theta) \quad (15)$$

$$a = \cos \theta - \frac{\sin \theta}{\sqrt{3}}, \quad b = \frac{2 \cdot \sin \theta}{\sqrt{3}}$$

$$a + b = \cos \theta + \frac{\sin \theta}{\sqrt{3}} \quad (16)$$

Por tanto, al promediar la expresión (16) entre 0° y 60° se obtiene un valor promedio de 1,103, lo que indica que la media digital sobrestima la distancia euclidiana en aproximadamente un 10,3%. Este valor corresponde con la diferencia porcentual observada experimentalmente en el sistema H3 [40].

Por último, para hacer un análisis más realista se procede a, en base a la Tabla 2-6, hacer un experimento en el que, dependiendo de la velocidad del buque simulado, se seleccione una frecuencia de transmisión de datos. Puesto que el análisis anterior se asume que se reciben mensajes AIS de posicionamiento en todas las celdas de la trayectoria, en un caso real esto no sucede así y puede que la trayectoria formada por las celdas correspondientes a los mensajes recibidos sea inconexa. Lo primero, como en caso anteriores, se procede a crear a una trayectoria a partir coordenadas y, seguidamente, se introduce el tipo de transpondedor de datos AIS (A, B o B+), la velocidad, la resolución y se calcula la distancia geodésica real de la trayectoria. Según qué tipo de transpondedor y la velocidad introducida, se elige de manera automática la frecuencia de transmisión de datos. Con estos datos, dividiendo la distancia total, entre la velocidad y a su vez por el intervalo de transmisión, se obtendrá el número total de transmisiones que se harán en la trayectoria. A continuación, se cuentan las transmisiones que se hacen en cada celda correspondiente, en caso de hacerse. El siguiente paso, servirá, simplemente, para una comprobación posterior de si las celdas por las que realmente pasa la trayectoria pertenecen al camino más corto mediante celdas usando una función mencionada con anterioridad como es *grid_path_cells*.

Para este análisis se ha creado la misma trayectoria del ejemplo anterior (Figura 3-25) cuya distancia total es de 269,02 km. Este análisis no requiere aplicarse a múltiples trayectorias porque el objetivo no es evaluar variaciones en rutas específicas, sino demostrar el impacto de las frecuencias de transmisión en todas las resoluciones H3. Dado que la frecuencia de transmisión es el factor clave que determina la distribución de los mensajes en la cuadrícula, cualquier trayectoria que cumpla con un rango típico de navegación es suficiente para observar cómo las transmisiones se agrupan o dispersan según la granularidad del *grid*. Por lo tanto, una única trayectoria bien definida permite extrapolar conclusiones generales sobre la relación entre la resolución H3 y la segmentación de las transmisiones AIS.

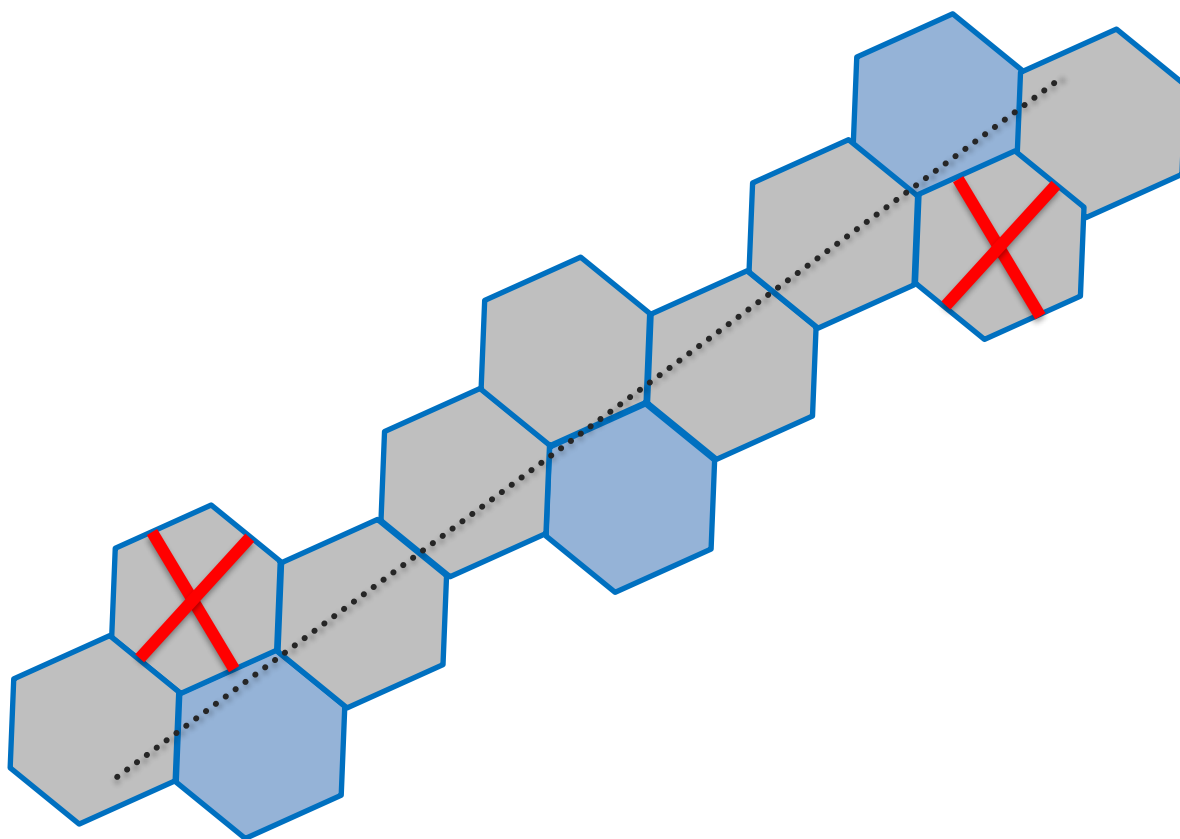


Figura 3-26 Representación de trayectoria mediante *grid* y transmisiones

La Figura 3-26 representa la relación entre la trayectoria de una embarcación, las celdas H3 teóricas por las que debería pasar y las celdas en las que realmente se transmiten los datos AIS. La línea de puntos indica la trayectoria del barco, mientras que los hexágonos representan las celdas de la cuadrícula H3 en distintas categorías:

- Celdas con una cruz roja: Son celdas teóricas por las que el barco debería haber pasado, pero en las que no se registraron transmisiones AIS.
- Celdas grises: Son celdas teóricas en las que sí se transmitieron datos, es decir, forman parte de la ruta esperada y han recibido señales.
- Celdas azules: Son celdas no teóricas, lo que significa que el barco no debería haber transmitido en ellas según la trayectoria estimada, pero en las que sí se registraron transmisiones AIS.

Esta representación ilustra cómo, dependiendo de la resolución del grid H3 y la frecuencia de transmisión, pueden existir discrepancias entre la ruta esperada y la distribución real de las transmisiones. Esto puede deberse a la interpolación de posiciones, variaciones en la señal AIS o diferencias en la asignación de transmisiones a las celdas. Además, la propia librería de H3 no calcula de modo preciso el *grid* y es posible que la trayectoria rectilínea no pase por encima de todas las celdas del propio *grid* y se tengan que añadir más celdas no incluidas en las que sí que se transmite.

Res	Transmisiones			Celdas del <i>grid_path</i>	Celdas con transmisiones			Celdas del <i>grid_path</i> sin transmisiones	Celdas con transmisiones que no están en el <i>grid_path</i>			Media de transmisiones por celda en que se transmitió		
	A	B	B+		A	B	B+		A	B	B+	A	B	B+
0	5810	1162	2324	1	1	1	1	0	0	0	0	5810,00	1162,00	2324,00
1	5810	1162	2324	2	2	2	2	0	0	0	0	2905,00	581,00	1162,00
2	5810	1162	2324	2	2	2	2	0	0	0	0	2905,00	581,00	1162,00
3	5810	1162	2324	4	4	4	4	0	0	0	0	1452,50	290,50	581,00
4	5810	1162	2324	8	9	9	9	1	2	2	2	645,56	129,11	258,22
5	5810	1162	2324	17	21	21	21	2	6	6	6	276,67	55,33	110,67
6	5810	1162	2324	48	50	50	50	12	14	14	14	116,20	23,24	46,48
7	5810	1162	2324	116	143	138	140	27	54	53	53	40,63	8,42	16,60
8	5810	1162	2324	327	332	329	332	273	278	276	278	17,50	3,53	7,00
9	5810	1162	2324	808	958	843	905	718	868	764	818	6,06	1,38	2,57
10	5810	1162	2324	2278	2290	1162	2086	2219	2231	1124	2032	2,54	1,00	1,11
11	5810	1162	2324	5650	5442	1162	2324	5616	5408	1152	2309	1,07	1,00	1,00
12	5810	1162	2324	15940	5810	1162	2324	15915	5785	1156	2313	1,00	1,00	1,00
13	5810	1162	2324	39542	5810	1162	2324	39533	5801	1159	2320	1,00	1,00	1,00
14	5810	1162	2324	111576	5810	1162	2324	111571	5805	1160	2322	1,00	1,00	1,00
15	5810	1162	2324	276783	5810	1162	2324	276781	5808	1160	2322	1,00	1,00	1,00

Tabla 3-9 Resultados de transmisiones por celdas

El comportamiento de la tabla sigue un patrón progresivo que se vuelve evidente al analizar cómo cambian las distintas métricas a medida que la resolución H3 aumenta. En las resoluciones más bajas,

los hexágonos son grandes y agrupan múltiples transmisiones dentro de cada celda. Esto se traduce en un número reducido de celdas en el *grid_path*, que representa el recorrido teórico, y una cantidad similar de celdas con transmisiones, lo que implica que casi todas las celdas en las que debería haber transmisiones efectivamente las tienen. En este rango de resolución, la media de transmisiones por celda es muy alta porque las transmisiones están altamente concentradas dentro de unas pocas celdas.

En los primeros niveles de segmentación, entre las resoluciones 0 y 3, las celdas del *grid_path* son pocas, con valores de 1, 2 o 4, dependiendo de la resolución. Esto significa que en estos niveles la trayectoria se representa con muy pocos hexágonos, lo que explica por qué la cantidad de transmisiones por celda sigue siendo extremadamente alta, con valores de hasta 5810 transmisiones por celda en AIS clase A. Esto ocurre porque todas las transmisiones quedan encapsuladas en un número mínimo de hexágonos, lo que simplifica la representación del trayecto, pero a costa de una gran pérdida de detalle.

A medida que la resolución avanza hacia valores intermedios, entre las resoluciones 4 y 7, comienza a notarse una transición. Aquí, la cantidad de celdas en el *grid_path* crece de forma significativa, llegando a valores superiores a 100 en resolución 7. Esto significa que el trayecto se empieza a segmentar en más celdas, lo que provoca una reducción en la media de transmisiones por celda. Es en este rango donde por primera vez aparecen celdas del *grid_path* que no reciben transmisiones, lo que sugiere que la dispersión de los datos AIS comienza a afectar la continuidad de la representación. Asimismo, comienzan a aparecer celdas con transmisiones que no forman parte del *grid_path* teórico, lo que indica que, a medida que el nivel de detalle aumenta, surgen discrepancias entre la trayectoria esperada y la forma en la que las transmisiones se distribuyen en la malla hexagonal.

En resoluciones más altas, a partir de la resolución 8, la tendencia se intensifica. En estos niveles, el número de celdas teóricas crece rápidamente, superando las 300 en resolución 8 y llegando a miles en resoluciones superiores. Esto significa que el trayecto se fragmenta en una gran cantidad de pequeños hexágonos, lo que lleva a un descenso aún más marcado en la cantidad de transmisiones por celda. Al mismo tiempo, se observa un aumento en el número de celdas del *grid_path* que no contienen transmisiones, alcanzando más de 1000 en resolución 10. También hay un crecimiento significativo en las celdas con transmisiones fuera del *grid_path*, lo que refuerza la idea de que, en estos niveles de segmentación, la precisión en la asignación de transmisiones a celdas específicas se ve afectada por la alta granularidad de la malla. En la Figura 3-27 se ve, mediante una escala de colores la cantidad de transmisiones por celda: los colores más oscuros para celdas donde se ha transmitido menos y los colores claros, al contrario.



Figura 3-27 Ruta con transmisiones en resoluciones medias

En los valores más extremos de resolución, desde la resolución 11 en adelante, el patrón alcanza su punto máximo. Aquí, el número de celdas en el *grid_path* y las celdas con transmisiones se vuelven prácticamente iguales, lo que sugiere que cada transmisión se encapsula en su propia celda. La media de

transmisiones por celda en estos casos se estabiliza en 1, lo que confirma que el nivel de segmentación es tan alto que la trayectoria se convierte en una serie de puntos discretos dentro de la malla H3. En estas resoluciones, la fragmentación es tan elevada que ya no se puede considerar que las transmisiones forman un trayecto continuo, sino más bien una secuencia de eventos aislados dentro de la cuadrícula como se puede ver en la Figura 3-28 donde los puntos azules representan las transmisiones.



Figura 3-28 Ruta mediante transmisiones en resoluciones muy altas

Este análisis muestra claramente cómo la discretización en celdas H3 influye en la representación de los datos AIS. En resoluciones bajas, el trayecto se simplifica y la información se agrupa en pocas celdas, lo que permite visualizar patrones generales de navegación. En resoluciones intermedias, aparece un punto de transición donde las transmisiones comienzan a dispersarse y surgen discrepancias entre el *grid_path* teórico y las transmisiones reales. En resoluciones muy altas, la segmentación es tan detallada que la trayectoria se descompone en puntos individuales, lo que puede ser útil para análisis de alta precisión, pero también introduce dificultades en la continuidad del recorrido. Encontrar un equilibrio entre estas representaciones depende del propósito del análisis y de la cantidad de detalle que se necesite preservar (véase Anexo XIII: Código para la comparación de transpondedores AIS en una trayectoria).

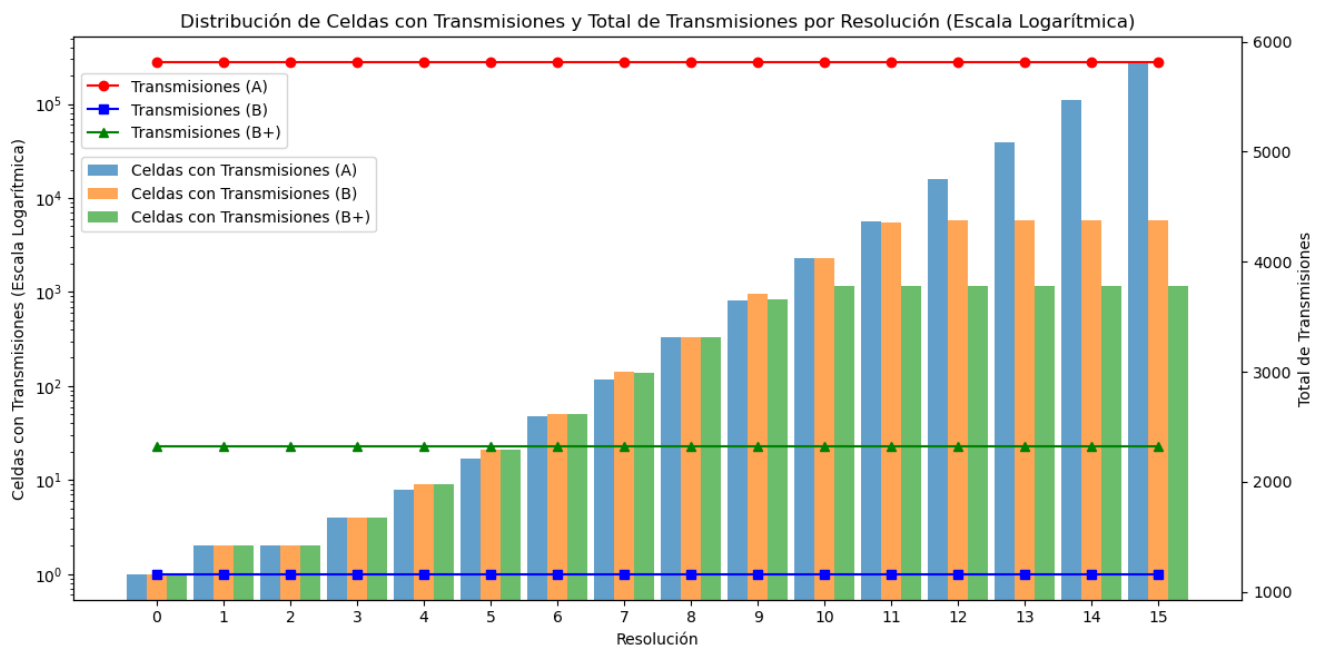


Figura 3-29 Gráfica de los resultados de la Tabla 3-9 Resultados de transmisiones por celdas

En la gráfica de la Figura 3-29 se representa la relación entre la cantidad de celdas con transmisiones y el total de transmisiones registradas a diferentes resoluciones. Se combinan dos tipos de visualización para facilitar la interpretación de los datos. En primer lugar, las barras reflejan la cantidad de celdas en las que se han registrado transmisiones, distinguiendo entre las categorías A, B y B+. La escala

logarítmica aplicada en el eje correspondiente permite observar la evolución de este parámetro de manera más clara, especialmente en los niveles de resolución más altos, donde los valores aumentan de forma considerable.

A su vez, sobre el mismo eje de resolución, se superpone una serie de líneas que representan el total de transmisiones para cada categoría. Este componente de la gráfica permite analizar cómo varía la distribución de transmisiones en relación con el número de celdas utilizadas. El uso de una segunda escala en el eje derecho permite que ambas variables coexistan en la representación sin distorsionar la lectura de los valores más pequeños.

A medida que la resolución se incrementa, como se ha detallado en párrafos anteriores, el comportamiento de ambas magnitudes refleja el impacto de la segmentación en el sistema de indexación. La combinación de ambos enfoques permite visualizar con mayor claridad la manera en que se distribuyen las transmisiones en función de la granularidad del análisis, facilitando la comparación entre diferentes configuraciones y ofreciendo una visión estructurada del fenómeno estudiado.

4 ANÁLISIS DE RESULTADOS

En este capítulo se profundiza en la evaluación de los resultados obtenidos a lo largo del desarrollo del estudio, explorando su validez y pertinencia en un contexto operativo. No se trata solo de exponer cifras y gráficos, sino de interpretar sus implicaciones y valorar hasta qué punto los métodos aplicados han logrado capturar la complejidad del problema analizado. Se abordará la coherencia interna de los datos, su fiabilidad frente a modelos teóricos y la comparación con enfoques alternativos, sin perder de vista las posibles limitaciones que pudieran comprometer su aplicabilidad en escenarios reales.

4.1 Resumen y validación de resultados.

La Figura 4-1 resume los resultados más relevantes de los distintos experimentos del trabajo. Aún así, uno de los aspectos fundamentales en cualquier análisis cuantitativo es determinar si los resultados obtenidos se alinean con lo esperado desde una perspectiva teórica. Para ello, se han contrastado los valores obtenidos con modelos previos, identificando patrones que confirmen la solidez del método empleado. En general, la distribución de errores sigue la lógica esperada, con una reducción progresiva a medida que se incrementa la resolución del sistema de celdas. Sin embargo, se han detectado ciertas anomalías en áreas donde la regularidad geométrica de las celdas se ve afectada, particularmente en regiones cercanas a los límites de indexación del sistema.

Más allá del análisis numérico, la repetición de experimentos ha permitido verificar la estabilidad de los resultados ante variaciones en las condiciones iniciales. En todas las iteraciones, las diferencias entre mediciones han permanecido dentro de márgenes razonables, lo que sugiere que el método utilizado es lo suficientemente robusto como para ofrecer resultados consistentes. No obstante, en algunas resoluciones específicas, sobre todo, en los análisis de trayectoria, se observan ligeras desviaciones que, si bien no invalidan las conclusiones generales, sí requieren un análisis más detallado en estudios futuros.

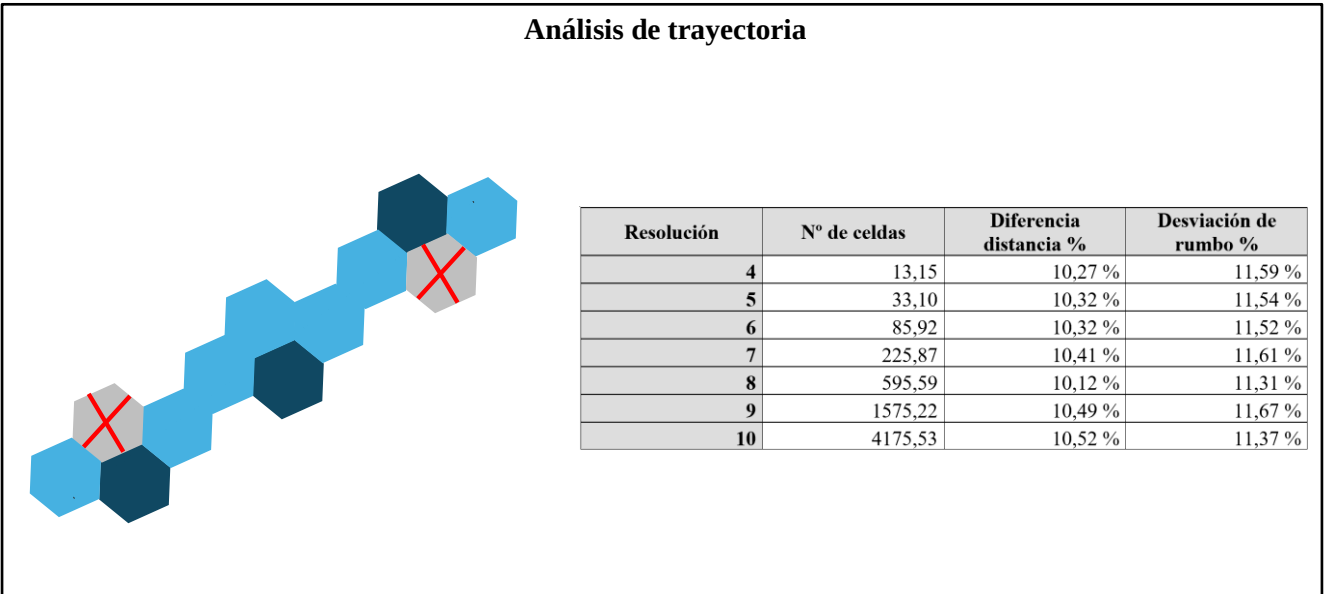
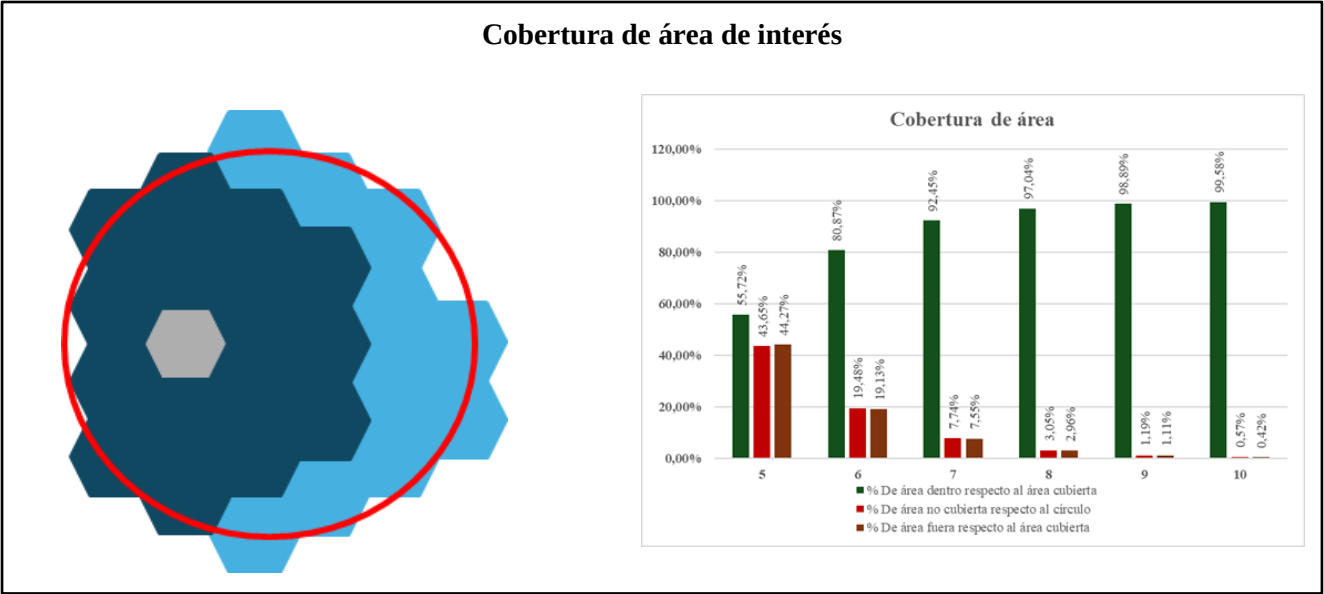
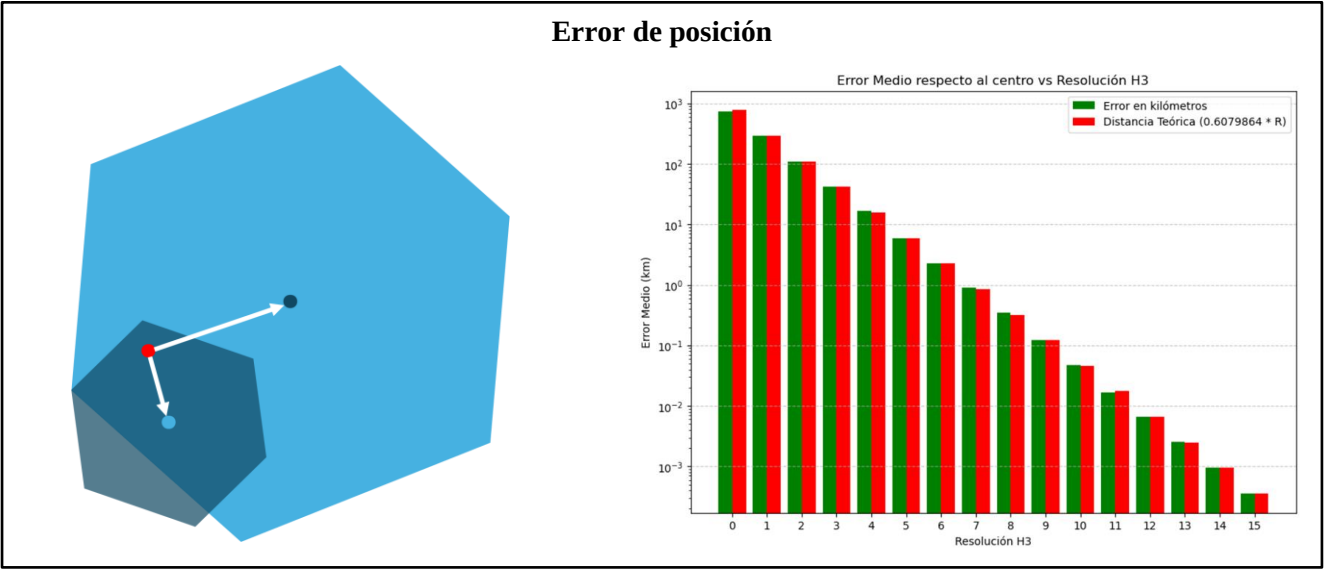


Figura 4-1 Resumen de resultados obtenidos

4.2 Discusión de resultados

Los resultados numéricos obtenidos a partir del análisis del sistema H3 permiten establecer criterios sólidos para determinar qué resolución es más adecuada en aplicaciones marítimas, como el tratamiento de datos AIS. Las gráficas de error medio y cobertura de área, junto con los valores relacionados con la desviación de trayectorias, muestran un comportamiento coherente que facilita la toma de decisiones fundamentadas.

Uno de los aspectos más destacables es la evolución del error medio respecto al centroide de la celda, que disminuye progresivamente a medida que aumenta la resolución. Este error sigue una tendencia decreciente que puede aproximarse teóricamente por una ley proporcional. Esta proporción refleja que, con cada incremento de resolución, el error se reduce por un factor cercano a $\sqrt{7}$. Paralelamente, el número de celdas necesarias para representar una misma área o trayectoria aumenta en un factor de aproximadamente 7 por cada nivel adicional de resolución, lo que implica un coste computacional creciente. No obstante, este coste sigue siendo manejable hasta cierto punto, lo cual es determinante para su viabilidad en contextos operativos.

Respecto a la cobertura de área, los resultados son concluyentes: a partir de la resolución 8, el sistema cubre más del 97% del área teórica y se acerca al 99,86% en resolución 10, con pérdidas de área prácticamente despreciables. Esto indica que, desde el punto de vista de eficiencia espacial, las resoluciones altas aseguran una representación casi completa, reduciendo la necesidad de cálculos complementarios o compensaciones por áreas no cubiertas.

En cuanto al análisis de trayectorias, es importante destacar que los errores relacionados con la diferencia de distancia y la desviación de rumbo se mantienen prácticamente constantes en todas las resoluciones estudiadas, situándose alrededor del 10% en distancia y 11,5% en rumbo. Este comportamiento constante se demuestra mediante la teoría de la distancia Manhattan que la resolución elegida no influye de forma significativa en la precisión de la representación de trayectorias. Por lo tanto, este parámetro no resulta determinante debido a que es sistemático y no se puede reducir mejorando la resolución; entonces, a la hora de elegir la resolución óptima, la decisión puede centrarse en los otros dos factores clave: error medio de posicionamiento y cobertura de área.

Con base en estos criterios, se puede concluir que la resolución 9 es apropiada para análisis de alta resolución y representa una elección equilibrada y eficiente para aplicaciones marítimas. A este nivel, el error de posicionamiento medio se sitúa alrededor de 100 metros, margen adecuado para tareas de seguimiento y control marítimo, mientras que la cobertura de área alcanza valores cercanos al 99%, con un coste computacional aún razonable. Si bien resoluciones superiores, como la 10, ofrecen mejoras marginales en precisión, estas deben valorarse únicamente en contextos donde se requiera un nivel de detalle más estricto. No obstante, en otras aplicaciones puede ser suficiente con resoluciones menores. Por ejemplo, para el proyecto del Sistema de Inteligencia artificial para el Reconocimiento del Entorno Marítimo (SIRENA) [41], se utiliza la resolución 6 ya que para esta aplicación es importante que haya varios mensajes en la misma celda.

En cambio, resoluciones inferiores a 6 no son recomendables en escenarios que requieran cierta precisión, dado que el error de posicionamiento se eleva considerablemente, superando los márgenes aceptables para aplicaciones operativas. Por tanto, si bien el análisis de trayectorias aporta información complementaria, la elección de resolución en el sistema H3 debe fundamentarse prioritariamente en la precisión posicional y la eficiencia en la cobertura del área analizada.

4.3 Limitaciones

A pesar de los resultados favorables, este estudio no está exento de limitaciones. Una de las principales es la dependencia del modelo de indexación con la geometría de las celdas, lo que introduce ciertas imprecisiones en regiones donde la estructura hexagonal se ve alterada. Este efecto podría

mitigarse mediante algoritmos de corrección que ajusten los valores en función de la distribución espacial de los errores (Figura 4-2).

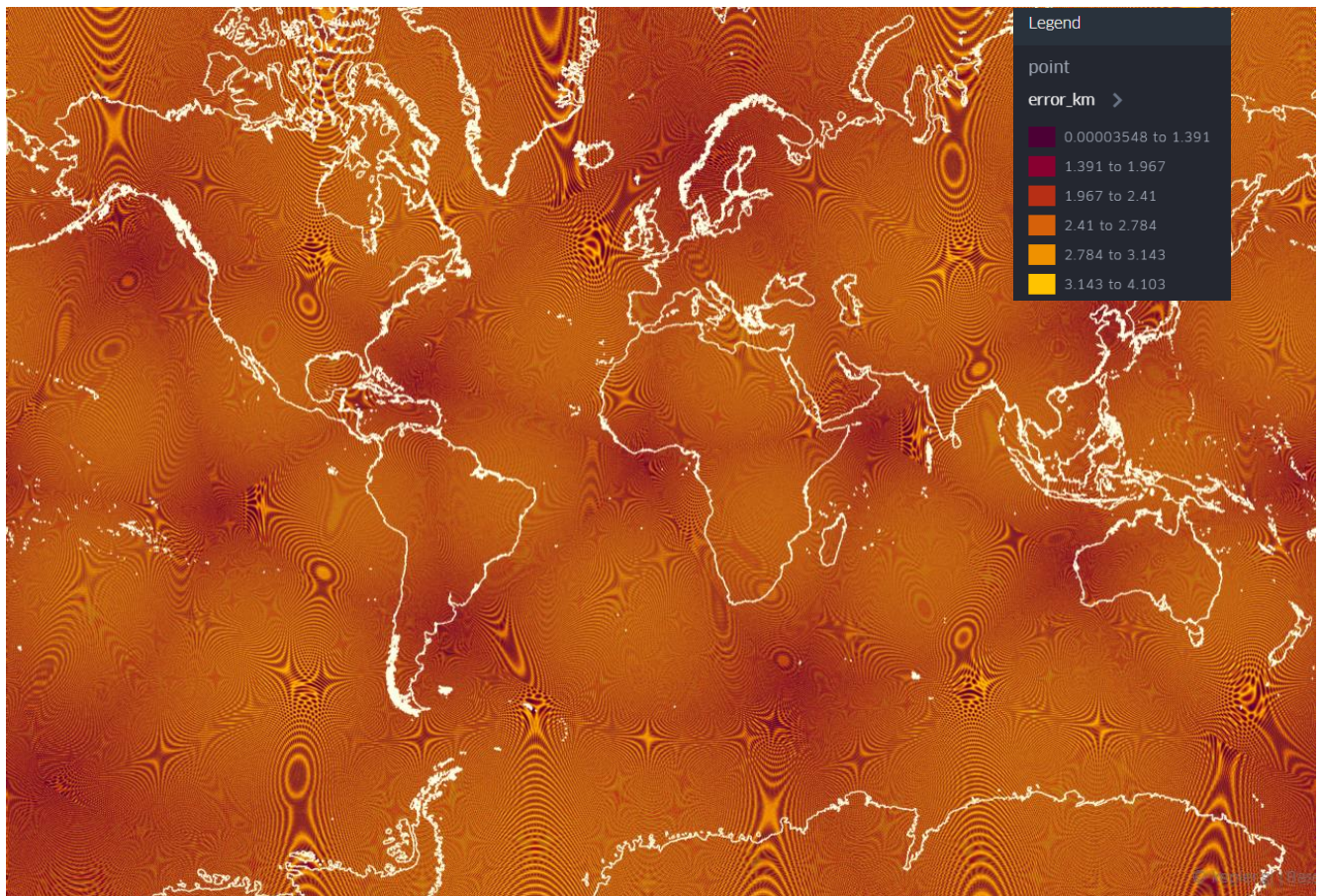


Figura 4-2 Distorsión del sistema en base al error de posición

Otra limitación importante es el costo computacional asociado a resoluciones elevadas. Si bien el sistema permite alcanzar niveles de precisión muy altos, el incremento en el número de celdas puede hacer que el procesamiento de datos se vuelva menos eficiente. Una posible solución sería el desarrollo de un modelo híbrido que combine diferentes resoluciones en función de la densidad de información en cada área, optimizando el uso de recursos sin comprometer la precisión.

Finalmente, cabe mencionar que la calidad de los resultados depende en gran medida de la calidad de los datos AIS utilizados. Si bien el sistema de indexación puede mejorar la organización de la información, no puede corregir errores en los datos de origen. Esto subraya la importancia de contar con fuentes de datos fiables y de aplicar mecanismos de validación adicionales para minimizar el impacto de posibles inconsistencias.

5 CONCLUSIONES Y LÍNEAS FUTURAS

El desarrollo de este Trabajo de Fin de Grado ha permitido profundizar en el análisis y evaluación del sistema de celdas Uber H3 aplicado al procesamiento de datos AIS, explorando su idoneidad y potencial en la gestión de grandes volúmenes de información geoespacial en el ámbito marítimo. A lo largo del estudio, se han contrastado sus capacidades con otros enfoques tradicionales y se han identificado tanto sus ventajas como sus limitaciones. A partir de los resultados obtenidos, es posible extraer una serie de conclusiones clave y proponer líneas de trabajo futuras que podrían enriquecer y ampliar este estudio.

5.1 Conclusiones

Desde una perspectiva global, los resultados han confirmado que Uber H3 es una solución eficiente para la indexación y análisis de datos espaciales derivados del AIS. Su estructura jerárquica de celdas hexagonales permite una representación geoespacial flexible, minimizando distorsiones y favoreciendo la escalabilidad en el tratamiento de grandes volúmenes de información. Este sistema ha demostrado ser especialmente útil para tareas como la segmentación espacial, el análisis de cobertura y la evaluación de trayectorias, proporcionando un marco de referencia homogéneo para la gestión del tráfico marítimo.

Uno de los principales hallazgos del estudio ha sido la cuantificación del error de posicionamiento asociado al uso de celdas H3. Si bien este error es inherente a cualquier sistema de discretización espacial, su magnitud se ha mostrado predecible y controlable mediante la selección adecuada de la resolución. En este sentido, el trabajo ha revelado que, a medida que la resolución aumenta, la precisión en la representación de posiciones mejora considerablemente, aunque a costa de un incremento en la carga computacional y en el número de celdas necesarias para representar un área dada. Por tanto, el análisis de error se ha cumplido satisfactoriamente; ya que los resultados muestran el error decrece de manera predecible siguiendo una proporcionalidad a la resolución y resoluciones 8 y 9 pueden ser útiles en aplicaciones relativas al posicionamiento marítimo debido a que se comenten errores próximos a los 100 metros.

El análisis de cobertura espacial ha puesto de manifiesto la capacidad de H3 para aproximar áreas de interés con un nivel de precisión adecuado para aplicaciones operativas. No obstante, también se han identificado ciertas limitaciones en la representación de regiones irregulares, así como en la transición entre distintas resoluciones. Estas diferencias pueden generar inconsistencias en aplicaciones que requieren una granularidad uniforme en todo el espacio geográfico analizado. Por lo anterior, el análisis de área de cobertura se ha cumplido exitosamente revelando que, de igual manera al de posición, este sistema resulta preciso a la hora de aproximar áreas de interés. En resoluciones medias, como la 8, la precisión de cobertura sobrepasa el 95% lo que hace que el sistema sea preciso con bajo coste computacional.

En lo que respecta al análisis de trayectorias, los resultados han confirmado que la discretización mediante celdas hexagonales introduce una distorsión sistemática en la representación del movimiento de embarcaciones. Sin embargo, esta distorsión sigue un patrón predecible y constante en torno a desviaciones del 10-11% en cuanto a distancia y rumbo se refiere. Este hecho permite corregir desviaciones en el rumbo y en la distancia recorrida mediante algoritmos de interpolación y ajuste de trayectoria. Esto abre la puerta a futuros desarrollos orientados a mejorar la fidelidad de los modelos de navegación basados en H3.

Por último, cabe destacar que la comparación con el sistema de indexación geoespacial Google S2 ha proporcionado información valiosa sobre el desempeño relativo de ambas tecnologías. Si bien S2 ofrece una granularidad más homogénea en ciertas resoluciones, H3 se ha mostrado más eficiente en términos de representación espacial y procesamiento distribuido, lo que refuerza su aplicabilidad en entornos donde la escalabilidad es un factor crítico.

En relación con los objetivos establecidos al inicio del trabajo, se puede afirmar que han sido cumplidos en su totalidad. Se ha realizado un análisis exhaustivo del sistema Uber H3 en el procesamiento de datos AIS, comparándolo con otras metodologías y validando su aplicabilidad en distintos escenarios. Además, los experimentos y pruebas llevadas a cabo han permitido evaluar su precisión, eficiencia y posibles limitaciones, confirmando que H3 es una alternativa viable para la optimización del tratamiento de datos geoespaciales en el ámbito marítimo.

5.2 Aplicaciones en casos reales

Más allá del análisis teórico, es importante considerar cómo estos resultados pueden trasladarse a aplicaciones reales. En el ámbito del tráfico marítimo, la capacidad de representar con precisión la posición de embarcaciones puede ser clave para optimizar rutas de navegación, mejorar la gestión de espacios portuarios y reforzar los sistemas de seguridad. El uso de un sistema de indexación eficiente permite reducir los tiempos de consulta de grandes volúmenes de datos, facilitando la detección de anomalías en el tráfico marítimo o la predicción de posibles incidentes.

Asimismo, en el campo de la gestión de recursos, la indexación geoespacial puede utilizarse para analizar patrones de navegación y evaluar el impacto de ciertas rutas sobre el medio ambiente. Al identificar áreas de alta concentración de tráfico, se pueden implementar estrategias para minimizar la congestión y reducir el consumo de combustible, contribuyendo así a una navegación más sostenible.

5.3 Líneas futuras

A partir de las conclusiones extraídas, se identifican varias líneas de investigación que podrían contribuir a mejorar y ampliar el presente trabajo:

- Optimización del balance entre precisión y eficiencia computacional: Explorar estrategias de adaptación dinámica de la resolución H3 en función de la densidad de datos y la relevancia espacial de cada área podría permitir un uso más eficiente de los recursos computacionales.
- Evaluación del impacto de la discretización en aplicaciones de navegación autónoma: Dado que los sistemas de guiado automático requieren una precisión extremadamente alta, sería interesante investigar hasta qué punto la representación basada en celdas hexagonales puede cumplir con estos requerimientos o si sería necesario combinarla con otros enfoques.
- Extensión del análisis de S2 a nuevos experimentos comparativos: En este trabajo se ha abordado parcialmente la comparación con el sistema de indexación S2. Una extensión natural sería aplicar la misma metodología de evaluación de error y análisis de distorsión espacial a S2, en diferentes resoluciones y contextos geográficos, para contrastar su rendimiento frente a H3. Además, podría explorarse una posible combinación o hibridación de ambos sistemas, aprovechando las ventajas particulares de cada uno en distintas situaciones.

6 BIBLIOGRAFÍA

- [1] «Convenio internacional para la seguridad de la vida humana en el mar, 1974 (Convenio SOLAS)». Accedido: 16 de enero de 2025. [En línea]. Disponible en: [https://www.imo.org/es/About/Conventions/Paginas/International-Convention-for-the-Safety-of-Life-at-Sea-\(SOLAS\),-1974.aspx](https://www.imo.org/es/About/Conventions/Paginas/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx)
- [2] «MarineTraffic: Global Ship Tracking Intelligence | AIS Marine Traffic». Accedido: 26 de febrero de 2025. [En línea]. Disponible en: <https://www.marinetraffic.com/en/ais/home/centerx:5.6/centery:40.7/zoom:4>
- [3] «H3: Uber's Hexagonal Hierarchical Spatial Index», Uber Blog. Accedido: 16 de enero de 2025. [En línea]. Disponible en: <https://www.uber.com/en-ID/blog/h3/>
- [4] M. U. Siddiqui y M. U. Syed, «AI-Based Anomaly Detection in Maritime Traffic», 31 de octubre de 2024. doi: 10.21203/rs.3.rs-5358980/v1.
- [5] «Proyecciones Cartográficas». Accedido: 16 de enero de 2025. [En línea]. Disponible en: <https://www.caee.utexas.edu/prof/maidment/gishyd97/atlas/exercise/ex2/webfiles/ex2afs.htm>
- [6] GISGeography, «How Map Projections Work», GIS Geography. Accedido: 17 de enero de 2025. [En línea]. Disponible en: <https://gisgeography.com/map-projections/>
- [7] W. Jaramillo, «Introducción a los conceptos de superposición en un SIG», Accedido: 16 de enero de 2025. [En línea]. Disponible en: https://www.academia.edu/7720639/Introducci%C3%B3n_a_los_conceptos_de_superposici%C3%B3n_en_un_SIG
- [8] M. Pavo López, «Un nomenclátor en el siglo II: la Geographia de Claudio Ptolomeo», *Bol. Real Soc. Geográfica*, n.º CLII, pp. 203-218, 2017.
- [9] A. Surroca Carrascosa, «Exposición comparativa de mapamundis medievales islámicos y crsitianos», *Bol. Real Soc. Geográfica*, vol. CLII, pp. 363-378, 2017.
- [10] «La Geografía de al Idrisi, Tabula Rogeriana». Accedido: 18 de enero de 2025. [En línea]. Disponible en: <http://www.arauco.org/sapereaude/cartografia/mapasac/tabularogeriana.html>

- [11] pascualsalinas, «Cartografía en el Renacimiento. Proyección Mercator | La Ciencia Renacentista». Accedido: 18 de enero de 2025. [En línea]. Disponible en: <https://blogs.ua.es/cienciarenacentista/2015/12/06/cartografia-en-el-renacimiento-proyeccion-mercator/>
- [12] «Proyecciones cartográficas | Cartografía (ESO)». Accedido: 25 de enero de 2025. [En línea]. Disponible en: https://educativo.ign.es/atlas-didactico/cartografia-eso/proyecciones_cartograficas.html
- [13] «Lambert cónica conforme—ArcMap | Documentación». Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://desktop.arcgis.com/es/arcmap/latest/map/projections/lambert-conformal-conic.htm>
- [14] «Área cilíndrica equivalente—ArcMap | Documentación». Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://desktop.arcgis.com/es/arcmap/latest/map/projections/cylindrical-equal-area.htm>
- [15] «Equidistante de dos puntos—ArcMap | Documentación». Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://desktop.arcgis.com/es/arcmap/latest/map/projections/two-point-equidistant.htm>
- [16] K. T. GmbH (<https://www.klokantech.com/>), «EPSG.io: Coordinate Systems Worldwide». Accedido: 24 de febrero de 2025. [En línea]. Disponible en: <https://epsg.io>
- [17] K. T. GmbH (<https://www.klokantech.com/>), «ETRS89 / Jersey Transverse Mercator - EPSG:3109». Accedido: 24 de febrero de 2025. [En línea]. Disponible en: <https://epsg.io>
- [18] «temariohtml». Accedido: 24 de febrero de 2025. [En línea]. Disponible en: https://www.um.es/geograf/sigmur/temariohtml/node8_tf.html
- [19] S. E. Battersby, M. P. Finn, E. L. Usery, y K. H. Yamamoto, «Implications of Web Mercator and Its Use in Online Mapping», *Cartogr. Int. J. Geogr. Inf. Geovisualization*, vol. 49, n.º 2, pp. 85-101, jun. 2014, doi: 10.3138/carto.49.2.2313.
- [20] «UGRC - The Earth is Not Round! Utah, NAD83 and Web Mercator Projections». Accedido: 24 de febrero de 2025. [En línea]. Disponible en: <https://gis.utah.gov/blog/2015-12-21-nad83-and-webmercator-projections/>
- [21] A. Ábrahánová y M. Vajsáblóvá, «A Comparison of Variational Projection and Cartographic Projection by Ritz's Method», *Slovak J. Civ. Eng.*, vol. 30, n.º 2, pp. 22-29, jun. 2022, doi: 10.2478/sjce-2022-0011.
- [22] «S2 Cells», S2Geometry. Accedido: 14 de febrero de 2025. [En línea]. Disponible en: http://s2geometry.io/devguide/s2cell_hierarchy.html
- [23] «AISLink CA2 Automatic Identification System AIS Class A», ACR Electronics, Inc. Accedido: 27 de enero de 2025. [En línea]. Disponible en: <https://www.acrartex.com/products/automatic-identification-system-aislink-ca2/>
- [24] «Transpondedor AIS - Los mejores transpondedores AIS para barco», Digital Yacht España. Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://digitalyacht.es/categoria-producto/sistemas-ais/ais-transpondedores/>

- [25] J. G.-A. Lacalle, «Consideraciones sobre el AIS para un velero», Navegantes Oceánicos. Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://navegantesoceanicos.com/ais-en-un-velero/>
- [26] S. Bhattacharjee, «What is Automatic Identification System (AIS)- Types And Working (FAQs)», Marine Insight. Accedido: 25 de enero de 2025. [En línea]. Disponible en: <https://www.marineinsight.com/marine-navigation/automatic-identification-system-ais-integrating-and-identifying-marine-communication-channels/>
- [27] ITU, «Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band». Accedido: 26 de enero de 2025. [En línea]. Disponible en: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I%21%21PDF-E.pdf
- [28] V. Alonso Aller, «Análisis de los sistemas de indexado geoespacial para el Conocimiento del Entorno Marítimo», Centro Universitario de la Defensa, Universidad de Vigo. Accedido: 25 de enero de 2025. [En línea]. Disponible en: <http://calderon.cud.uvigo.es:8080/api/core/bitstreams/3a46ec76-a820-46c1-af2f-5b37a8c665cf/content>
- [29] M. Prvan, J. Ožegovic, y A. Burazin, «A Review of Embedding Hexagonal Cells in the Circular and Hexagonal Region of Interest», *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, n.º 7, 2019, doi: 10.14569/IJACSA.2019.0100747.
- [30] «Project Jupyter». Accedido: 15 de febrero de 2025. [En línea]. Disponible en: <https://jupyter.org>
- [31] «Welcome to Python.org». Accedido: 15 de febrero de 2025. [En línea]. Disponible en: <https://www.python.org/>
- [32] «API Quick Reference — h3-py». Accedido: 11 de marzo de 2025. [En línea]. Disponible en: https://uber.github.io/h3-py/api_quick.html
- [33] «Anaconda | Built to Advance Open Source AI», Anaconda. Accedido: 11 de marzo de 2025. [En línea]. Disponible en: <https://www.anaconda.com/>
- [34] «Large-scale WebGL-powered Geospatial Data Visualization Tool», kepler.gl. Accedido: 15 de febrero de 2025. [En línea]. Disponible en: <http://kepler.gl/>
- [35] «¿Qué es la simulación de Montecarlo? | IBM». Accedido: 24 de febrero de 2025. [En línea]. Disponible en: <https://www.ibm.com/es-es/topics/monte-carlo-simulation>
- [36] «Natural Earth» Downloads - Free vector and raster map data at 1:10m, 1:50m, and 1:110m scales». Accedido: 19 de febrero de 2025. [En línea]. Disponible en: <https://www.naturalearthdata.com/downloads/>
- [37] «Welcome to GeoPy's documentation! — GeoPy 2.4.1 documentation». Accedido: 12 de marzo de 2025. [En línea]. Disponible en: <https://geopy.readthedocs.io/en/stable/#module-geopy.distance>
- [38] R. E. Stone, «Technical Note—Some Average Distance Results», *Transp. Sci.*, vol. 25, n.º 1, pp. 83-90, feb. 1991, doi: 10.1287/trsc.25.1.83.
- [39] J. V. Gómez Urgellés, «Euclides no vivió en Manhattan: Geometría urbana», *Model. Sci. Educ. Learn.*, vol. 12, n.º 1, p. 59, feb. 2019, doi: 10.4995/msel.2019.10808.

- [40] A. J. Patel, «Hexagonal Grids», Red Blob Games, 2013. Accedido: 22 de marzo de 2025. [En línea]. Disponible en: <https://www.redblobgames.com/grids/hexagons/>
- [41] M. Rodelgo Lacruz, B. Barragáns Martínez, P. Sendín Raña, A. García Suárez, N. Fernández García, y G. Higgins, «Real-time AIS data analysis for anomaly detection in the maritime environment».

ANEXO I: DICCIONARIO DE SIGLAS, ACRÓNIMOS Y ABREVIATURAS

AIS: *Automatic Identification System*

COG: *Course Over Ground*

d. C.: *Después de Cristo*

ETA: *Estimated Time of Arrival*

GNSS: *Global Navigation Satellite System*

GPS: *Global Positioning System*

IMO: *International Maritime Organisation* (Organización Marítima Internacional (OMI))

MMSI: *Maritime Mobile Service Identity*

SAR: *Search And Rescue*

SIG: *Sistema de Información Geográfico*

SOG: *Speed Over Ground*

SOLAS: *Safety Of Life At Sea*

SOTDMA: *Self Organizing Time-Division Multiple Access*

UTC: *Coordinated Universal Time*

ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

El uso avanzado de datos AIS combinado con la indexación espacial Uber H3 aporta grandes capacidades en el ámbito marítimo, pero también genera importantes consideraciones éticas y sociales. Entre ellas destacan la privacidad de la información, la vigilancia marítima y seguridad, la equidad en el acceso a la tecnología y la transparencia y fiabilidad de los sistemas inteligentes.

El AIS transmite información detallada de la posición e identidad de los buques, lo que suscita preocupaciones de privacidad. Una vez recopilados, estos datos permiten rastrear rutas y actividades de buques o incluso embarcaciones de recreo, potencialmente revelando información sensible. Algunos profesionales marítimos han cuestionado la exposición pública de estos datos, advirtiendo que podría poner en riesgo a las tripulaciones si actores malintencionados los usan para ubicar objetivos vulnerables. Sin embargo, la filosofía de seguridad cooperativa de la OMI primó sobre la privacidad al diseñar el AIS: al emitir datos sin cifrado por radio VHF, cualquier parte interesada (desde autoridades hasta otros buques) puede monitorear el tráfico. Invertir esta dinámica requeriría un cambio tecnológico global (por ejemplo, aplicar cifrado en los transpondedores), algo poco viable económicamente. En la práctica, para proteger la privacidad se recurre más bien a la anonimización

y agregación de datos en los análisis; por ejemplo, trabajando sobre cuadrículas H3 se pueden extraer patrones generales sin necesidad de divulgar cada posición individual.

A su vez, la disponibilidad masiva de datos AIS habilita una vigilancia marítima eficaz a escala global. Organismos civiles y militares emplean estos datos para supervisar el tráfico, detectar pesca ilegal, piratería u otras amenazas. Tecnologías como H3 potencian esta labor al facilitar la integración de diversas fuentes de datos en un marco común. Ahora bien, el AIS abierto también puede ser explotado con fines ilícitos: se han documentado tácticas de “apagado” del AIS para hacer que un buque desaparezca de los radares civiles, práctica común de flotas pesqueras ilegales o buques en actividades clandestinas. Asimismo, la manipulación deliberada de las señales AIS también se emplea para ocultar contrabando o eludir sanciones, aprovechando la ausencia de cifrado en el sistema. Estas vulnerabilidades suponen un reto ético en términos de seguridad internacional: si bien el intercambio abierto de datos fortalece la colaboración y el conocimiento situacional, también obliga a desarrollar salvaguardas (técnicas y legales) contra usos maliciosos. El equilibrio entre transparencia y control se vuelve crítico para que la vigilancia marítima no derive ni en desprotección ante amenazas ni en una intromisión excesiva en la soberanía o en las actividades legítimas.

ANEXO III: IMPLICACIONES AMBIENTALES Y ECONÓMICAS

El sistema de indexación geoespacial H3, desarrollado por Uber, divide la superficie terrestre en una cuadrícula jerárquica de celdas hexagonales, facilitando el análisis y la visualización de datos espaciales. En el ámbito marítimo, esta herramienta ofrece aplicaciones significativas para el monitoreo de emisiones y la gestión ambiental, permitiendo una representación precisa de datos geoespaciales esenciales para supervisar la contaminación generada por el tráfico de embarcaciones.

Al integrar datos del Sistema de Identificación Automática (AIS) con la cuadrícula de H3, es posible rastrear las rutas de las embarcaciones y asociarlas con niveles de emisiones específicos. Esta integración facilita la identificación de áreas marítimas con alta concentración de contaminantes, permitiendo a las autoridades implementar medidas de control más efectivas. Además, la capacidad de H3 para agregar datos a diferentes niveles de resolución permite un análisis detallado de las emisiones en zonas costeras sensibles, contribuyendo a la protección de ecosistemas marinos vulnerables.

La aplicación de H3 en el monitoreo de emisiones marítimas también se alinea con normativas internacionales que buscan reducir la contaminación en los océanos. Por ejemplo, la Organización Marítima Internacional (OMI) ha establecido límites estrictos para las emisiones de óxidos de azufre y nitrógeno provenientes de buques. La implementación de sistemas de monitoreo basados en H3 puede ayudar a garantizar el cumplimiento de estas regulaciones al proporcionar datos precisos y actualizados sobre las emisiones de cada embarcación. Asimismo, el sistema H3 es compatible con el marco de la Directiva INSPIRE de la Unión Europea, que establece un modelo para la armonización y el intercambio de información geográfica entre los Estados miembros. Su aplicación en la vigilancia ambiental marítima facilitaría la cooperación entre administraciones y reforzaría la toma de decisiones fundamentadas en datos espaciales estandarizados.

Además, la capacidad de H3 para integrarse con otras tecnologías de monitoreo ambiental permite una gestión más completa del entorno marítimo. Al combinar datos de calidad del aire y del agua con la información de tráfico marítimo indexada por H3, es posible evaluar el impacto ambiental de las operaciones portuarias y desarrollar estrategias para mitigar efectos negativos en las comunidades costeras. Estos esfuerzos se alinean con la Agenda 2030 para el Desarrollo Sostenible de las Naciones Unidas, en particular con los Objetivos de Desarrollo Sostenible (ODS) relacionados con la conservación de los océanos (ODS 14) y la acción climática (ODS 13). La capacidad de este sistema para proporcionar análisis espaciales detallados contribuye al diseño de políticas más sostenibles y al establecimiento de zonas de control de emisiones, ayudando a reducir la huella ambiental del transporte marítimo.

Además de los factores ambientales, la adopción del sistema de celdas Uber H3 para procesar datos AIS conlleva notables implicaciones económicas en el sector marítimo. Entre ellas se incluyen los costes de implementación frente a los beneficios operativos, las mejoras en la eficiencia (especialmente en rutas y combustible) y los impactos en la competitividad y reducción de riesgos.

El efecto económico directo más notable es la reducción del consumo de combustible. Dado que el combustible puede suponer entre el 50% y 70% de los costes operativos de un buque, incluso una mejora modesta tiene gran impacto. Estudios y proyectos piloto han evidenciado ahorros cercanos al 10% en consumo de combustible gracias a la optimización basada en datos AIS e inteligencia artificial, lo que en un gran ahorro por travesía. Además del ahorro en combustible, las rutas óptimas reducen el desgaste de maquinaria y minimizan costes asociados a demoras (esperas en puerto, sobretiempos de tripulación, etc.). En conjunto, la eficiencia operativa alcanzada con estas técnicas repercute en menores gastos logísticos y en una operación más rentable y predecible.

Desde una perspectiva estratégica, incorporar Uber H3 en el análisis AIS puede mejorar la competitividad de las empresas marítimas. Aquellas navieras que logran bajar sus costes (por menor consumo y mejor utilización de la flota) pueden ofrecer tarifas más atractivas o aumentar sus márgenes,

ganando ventaja en un mercado global muy disputado. Asimismo, la fiabilidad añadida en las operaciones (p. ej., entregas más puntuales gracias a rutas optimizadas) refuerza la reputación y las relaciones comerciales de la compañía. Por otro lado, la analítica avanzada de AIS contribuye a la gestión del riesgo operacional, lo cual tiene también un reflejo económico. Al facilitar un monitoreo proactivo, es posible anticipar eventos anómalos y prevenir accidentes costosos. Por ejemplo, integrar datos de navegación con H3 permite identificar áreas de alta probabilidad de incidentes (como encallamientos) y alertar con antelación suficiente, evitando siniestros que acarrearían gastos enormes en reparaciones, indemnizaciones o interrupciones del servicio. En suma, los avances apoyados en AIS y H3 impulsan una eficiencia económica integral: menos costes operativos diarios y menor exposición a pérdidas extraordinarias, fortaleciendo la sostenibilidad financiera del negocio marítimo.

ANEXO IV: CÓDIGO DE DISTRIBUCIÓN DE DISTANCIAS EN UN HEXÁGONO

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.patches as patches

# Parámetros
n_points = 10000000
radius = 1
points = []

# Generar puntos aleatorios directamente
while len(points) < n_points:
    x = np.random.uniform(-radius, radius)
    y = np.random.uniform(-radius, radius)

    # Comprobar si el punto está dentro del hexágono
    if np.abs(x) <= radius and np.abs(y) <= np.sqrt(3) * radius / 2:
        if np.sqrt(3) * abs(x) + abs(y) <= np.sqrt(3) * radius:
            points.append((x, y))

points = np.array(points)

# Calcular distancias desde el centro
distances = np.sqrt(points[:, 0]**2 + points[:, 1]**2)
mean_distance = np.mean(distances)

# Radio del círculo inscrito
inscribed_radius = np.sqrt(3) * radius / 2

# Graficar los puntos con un gradiente de color por distancia y el círculo
inscrito
plt.figure(figsize=(10, 7))
plt.scatter(points[:, 0], points[:, 1], c=distances, cmap='viridis', s=0.1,
            alpha=0.6)

# Dibujar el círculo inscrito
circle = patches.Circle((0, 0), radius=inscribed_radius, edgecolor='red',
                        facecolor='none', linewidth=2, linestyle='--')
plt.gca().add_patch(circle)

plt.colorbar(label='Distancia desde el centro')
plt.title('Distancia de puntos aleatorios dentro de un hexágono perfecto')
plt.xlabel('X')
plt.ylabel('Y')
plt.axis('equal')
plt.show()

# Graficar el histograma de las distancias con la línea de distancia media
plt.figure(figsize=(10, 6))
plt.hist(distances, bins=100, color='blue', alpha=0.7, edgecolor='black')
plt.axvline(mean_distance, color='red', linestyle='dashed', linewidth=2,
            label=f'Distancia media = {mean_distance:.6f}')
plt.title('Distribución de la distancia desde el centro en un hexágono perfecto')
plt.xlabel('Distancia desde el centro')
plt.ylabel('Frecuencia')
plt.legend()
plt.grid(True)
plt.show()
```

ANEXO V: CÓDIGO DE DISTRIBUCIÓN DE ERROR PARA UNA RESOLUCIÓN

```
import h3
import numpy as np
import matplotlib.pyplot as plt
from geopy.distance import geodesic
import random

resolucion = 6
puntos = []
while len(puntos) < 1000000:
    lat = np.degrees(np.arcsin(random.uniform(-1, 1)))
    lon = random.uniform(-180, 180)
    puntos.append((lat, lon))

indices_h3 = [h3.latlng_to_cell(lat, lon, resolucion) for lat, lon in puntos]
centroides = [h3.cell_to_latlng(h) for h in indices_h3]

distancias = [geodesic(punto, centroide).kilometers for punto, centroide in
zip(puntos, centroides)]
error_medio = np.mean(distancias)

plt.figure(figsize=(10, 6))
plt.hist(distancias, bins=90, edgecolor='k', alpha=0.7, label="Distribución real",
density=True)

plt.axvline(error_medio, color='r', linestyle='dashed', linewidth=2, label=f'Error
medio: {error_medio:.4f} km')
plt.text(error_medio, plt.ylim()[1] * 0.9, f'{error_medio:.4f} km', color='r',
ha='center')

# Detalles de la gráfica
plt.title(f'Distribución de errores para resolución H3 {resolucion}')
plt.xlabel('Distancia al centroide (km)')
plt.ylabel('Frecuencia (densidad)')
plt.legend()
plt.grid(True)
plt.show()

print(f'Error medio real para resolución {resolucion}: {error_medio:.4f} km')
```

ANEXO VI: CÓDIGO DE ERROR TEÓRICO VS REAL POR RESOLUCIÓN

```
import h3
import random
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from geopy.distance import geodesic

# Generar 1000000 puntos aleatorios una vez
points = []
while len(points) < 1000000:
    lat = np.degrees(np.arcsin(random.uniform(-1, 1)))
    lon = random.uniform(-180, 180)
    points.append((lat, lon))

mean_errors_km = []
resolutions = list(range(16))

# Tabla de longitudes de lado medio de celda por resolución
side_lengths = [
    1281.256011, 483.0568391, 182.5129565, 68.97922179, 26.07175968,
    9.854090990, 3.724532667, 1.406475763, 0.531414010, 0.200786148,
    0.075863783, 0.028663897, 0.010830188, 0.004092010, 0.001546100,
    0.000584169
]

# Calcular la distancia teórica
theoretical_distances = [0.6079864 * R for R in resolutions]

# Iterar a través de las resoluciones
for res in resolutions:
    errors_km = []
    for lat, lon in points: # Utilizar los puntos generados para calcular errores
        h3_cell = h3.latlng_to_cell(lat, lon, res)
        center_lat, center_lon = h3.cell_to_latlng(h3_cell)
        error_km = geodesic((lat, lon), (center_lat, center_lon)).kilometers
        errors_km.append(error_km)
    mean_error_km = np.mean(errors_km) # error medio para resolución
    mean_errors_km.append(mean_error_km)

print("Errores medios por resolución:")
for res, error_km in zip(resolutions, mean_errors_km):
    print(f"Resolución {res}: {error_km:.4f} km --> {error_km/1.852:.4f} mn ")

# Graficar los resultados con barras
plt.figure(figsize=(10, 7))
bar_width = 0.4
plt.bar(np.array(resolutions) - bar_width/2, mean_errors_km, width=bar_width,
color='g', label='Error en kilómetros')
plt.bar(np.array(resolutions) + bar_width/2, theoretical_distances,
width=bar_width, color='r', label='Distancia Teórica (0.6079864 * R)')
plt.title('Error Medio respecto al centro vs Resolución H3')
plt.xlabel('Resolución H3')
plt.ylabel('Error Medio (km)')
plt.xticks(resolutions)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.show()
```

ANEXO VII: CÓDIGO PARA VISUALIZAR DISTRIBUCIÓN DE ERROR DE UN ÁREA EN UN MAPA EN KEPLER.GL

```
import h3
import numpy as np
import pandas as pd
from geopy.distance import geodesic
import geopandas as gpd
from shapely.geometry import Point
from kepler.gl import KeplerGl

# Definir la zona de interés
lat_min = 42.0
lat_max = 42.7
lon_min = -10.0
lon_max = -8.59

# Crear una cuadrícula de puntos dentro de la zona de interés
n_lat = 100
n_lon = 100
latitudes = np.linspace(lat_min, lat_max, n_lat)
longitudes = np.linspace(lon_min, lon_max, n_lon)
puntos = [(lat, lon) for lat in latitudes for lon in longitudes]

# Cargar shapefile de tierra
ruta_shp = r"C:\Users\Mario\Downloads\ne_10m_land\ne_10m_land.shp"
land = gpd.read_file(ruta_shp)

# Filtrar solo puntos oceánicos
points = []
for lat, lon in puntos:
    point = Point(lon, lat)
    is_in_land = any(geom.contains(point) for geom in land.geometry)
    if not is_in_land:
        points.append((lat, lon))

# Calcular el error en cada punto usando H3
resolucion = 6
datos = []
for lat, lon in points:
    h3_index = h3.latlng_to_cell(lat, lon, resolucion)
    centroide_lat, centroide_lon = h3.cell_to_latlng(h3_index)
    distancia = geodesic((lat, lon), (centroide_lat, centroide_lon)).kilometers
    datos.append({'latitude': lat, 'longitude': lon, 'error_km': distancia})

df_errores = pd.DataFrame(datos)

# Visualizar los datos en Kepler.gl
map_kepler = KeplerGl(height=600)
map_kepler.add_data(data=df_errores, name='Errores H3 Oceánicos')
html_file = "mapa_error_kepler.html"
map_kepler.save_to_html(file_name=html_file)

print(f"Mapa interactivo guardado como {html_file}")
```

ANEXO VIII: CÓDIGO DE COMPARACIÓN UBER H3 CON GOOGLE S2

```
import random
import h3
import s2sphere
import geopy.distance
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# Selección de resoluciones equivalentes en tamaño de celda entre H3 y S2
resolutions_h3 = [1, 4, 7, 10, 13] # Equivalentes en tamaño con S2
resolutions_s2 = [2, 5, 8, 11, 14] # Equivalentes en tamaño con H3

# Número de puntos a evaluar
num_points = 1000000

# Generación de puntos aleatorios en el océano con corrección de latitud usando arcsen
points = [(np.degrees(np.arcsin(random.uniform(-1, 1))), random.uniform(-180, 180)) for _ in range(num_points)]

# Lista para almacenar los resultados
results = []

# Comparación de H3 y S2
for h3_res, s2_res in zip(resolutions_h3, resolutions_s2):
    h3_errors = []
    s2_errors = []

    for lat, lon in points:
        # ---- Uber H3 ----
        h3_index = h3.latlng_to_cell(lat, lon, h3_res)
        h3_center = h3.cell_to_latlng(h3_index)
        h3_error = geopy.distance.geodesic((lat, lon), h3_center).km

        # ---- Google S2 ----
        cell = s2sphere.CellId.from_lat_lng(s2sphere.LatLng.from_degrees(lat, lon)).parent(s2_res)
        s2_center = cell.to_lat_lng()
        s2_error = geopy.distance.geodesic((lat, lon), (s2_center.lat().degrees, s2_center.lng().degrees)).km

        # Guardamos los datos
        h3_errors.append(h3_error)
        s2_errors.append(s2_error)

    # Calculamos estadísticas
    results.append({
        "Resolución H3": h3_res,
        "Resolución S2": s2_res,
        "H3_Error_Promedio (km)": np.mean(h3_errors),
        "S2_Error_Promedio (km)": np.mean(s2_errors),
    })

# Convertimos los resultados a un DataFrame
df_results = pd.DataFrame(results)

# Mostramos los resultados en consola
print(df_results)
```

```
# Graficamos los errores como barras con valores numéricos
plt.figure(figsize=(10, 5))
bar_width = 0.4
x = np.arange(len(df_results))

bars1 = plt.bar(x - bar_width / 2, df_results["H3_Error_Promedio (km)"],
bar_width, label="H3 Error")
bars2 = plt.bar(x + bar_width / 2, df_results["S2_Error_Promedio (km)"],
bar_width, label="S2 Error")

plt.xticks(x, df_results["Resolución H3"].astype(str) + " / " +
df_results["Resolución S2"].astype(str))
plt.xlabel("Resoluciones H3 / S2")
plt.ylabel("Error Promedio (km)")
plt.title("Comparación de Error de Posicionamiento")
plt.legend()
plt.grid(axis='y')

# Agregar valores numéricos encima de las barras
for bar in bars1 + bars2:
    plt.text(bar.get_x() + bar.get_width() / 2, bar.get_height(),
f'{bar.get_height():.2f}', ha='center', va='bottom')
plt.show()
```

ANEXO IX: CÓDIGO PARA COMPARACIÓN DE ÁREA REAL VS ÁREA CUBIERTA POR CELDAS UBER H3

```
import h3
import math
from geopy.distance import geodesic

# Ubicación de referencia
lat = 40
lon = -10
print(f"Ubicación: Latitud {lat:.6f}, Longitud {lon:.6f}")

# Radio del área de interés en km
radio = 15
resolucion = int(input("Introduce la resolución H3 deseada (0-15): "))

# Obtener la celda H3 central
celda_central = h3.latlng_to_cell(lat, lon, resolucion)
print(f"Celda central H3: {celda_central}")

# Calcular el área del círculo de referencia
area_circulo = math.pi * (radio ** 2)

# Calcular el área de una celda H3 a la resolución dada
area_hex = h3.cell_area(celda_central, unit='km^2')

# Lista de celdas seleccionadas (comenzamos con la celda central)
celdas_finales = {celda_central}
k = 1 # Iniciamos expansión en k=1

while True:
    anillo = h3.grid_ring(celda_central, k) # Obtener celdas del anillo k
    celdas_validas = set() # Guardará solo las celdas cuyo centro esté dentro del
    radio

    for celda in anillo:
        centroide = h3.cell_to_latlng(celda)
        distancia_centro = geodesic((lat, lon), centroide).km # Distancia al
        centro

        if distancia_centro <= radio: # Si el centro de la celda está dentro del
        radio, la añadimos
            celdas_validas.add(celda)

        if not celdas_validas: # Si ningún centro de celda en el anillo está dentro
        del círculo, se detiene la expansión
            break

    celdas_finales.update(celdas_validas) # Agregar las celdas filtradas a las
    finales
    k += 1 # Pasar al siguiente anillo

# Calcular área aproximada final con las celdas seleccionadas
area_aprox = len(celdas_finales) * area_hex

# Resultados finales
print(f"Área del círculo real: {area_circulo:.2f} km²")
print(f"Área cubierta por H3 ajustada: {area_aprox:.2f} km²")
print(f"Número de celdas utilizadas: {len(celdas_finales)}")
```

ANEXO X: CÓDIGO PARA OBTENER PORCENTAJE DE COBERTURA RESPECTO AL CÍRCULO REAL

```
import h3
import math
import numpy as np
import pandas as pd
from geopy.distance import geodesic
from shapely.geometry import Point, Polygon
import scipy.integrate as spi
from tqdm import tqdm

# Parámetros generales
radio = 15 # Radio del área de interés en km
resolucion = 10 # Resolución H3 deseada
num_ubicaciones = 50 # Número de ubicaciones aleatorias a generar

# Rango de latitudes y longitudes para generar ubicaciones aleatorias
min_lat, max_lat = -80, 80 # Evitamos áreas polares
min_lon, max_lon = -180, 180

# Inicializar diccionario para almacenar resultados
resultados = {
    "num_celdas": [],
    "area_circulo": [],
    "area_total_hexagonos": [],
    "area_dentro_circulo": [],
    "area_no_cubierta": [],
    "area_fuera_circulo": []
}

# Área del círculo es constante para todas las ubicaciones
area_circulo = math.pi * (radio ** 2)

# Iteramos sobre las ubicaciones aleatorias con barra de progreso
for _ in tqdm(range(num_ubicaciones)):
    # Generar ubicación aleatoria
    lat = np.random.uniform(min_lat, max_lat)
    lon = np.random.uniform(min_lon, max_lon)

    # Obtener celda central para esta ubicación
    celda_central = h3.latlng_to_cell(lat, lon, resolucion)

    # Calcular área del hexágono (puede variar según la ubicación)
    area_hex = h3.cell_area(celda_central, unit='km^2')

    # Encontrar todas las celdas con centros dentro del círculo
    celdas_finales = {celda_central}
    k = 1

    # Optimización: Preestimar el número máximo de anillos necesarios
    # basado en el radio y el tamaño aproximado de las celdas
    while True:
        anillo = h3.grid_ring(celda_central, k)
        if not anillo: # Si el anillo está vacío, salir del bucle
            break

        celdas_validas = set()
        for celda in anillo:
            centroide = h3.cell_to_latlng(celda)
```

```
        if geodesic((lat, lon), centroide).km <= radio:
            celdas_validas.add(celda)

    if not celdas_validas:
        break

    celdas_finales.update(celdas_validas)
    k += 1

# Calcular áreas para esta ubicación
area_dentro_del_circulo = 0
area_fuera_del_circulo = 0

for celda in celdas_finales:
    # Obtener vértices y crear polígono
    vertices = h3.cell_to_boundary(celda)
    poly = Polygon([(v[1], v[0]) for v in vertices])
    min_lon_poly, min_lat_poly, max_lon_poly, max_lat_poly = poly.bounds

    # Optimización: Verificar si el polígono está completamente dentro o fuera
del círculo
    # antes de realizar la integración
    vertices_dentro = 0
    vertices_fuera = 0
    for v in vertices:
        if geodesic((lat, lon), (v[0], v[1])).km <= radio:
            vertices_dentro += 1
        else:
            vertices_fuera += 1

    # Si todos los vértices están dentro del círculo, no necesitamos integrar
    if vertices_dentro == len(vertices):
        area_dentro_del_circulo += area_hex
        continue

    # Si todos los vértices están fuera y la distancia mínima es mayor que el
radio,
    # no necesitamos integrar
    if vertices_fuera == len(vertices):
        # Verificar si el centroide está fuera y lejos del círculo
        centroide = h3.cell_to_latlng(celda)
        if geodesic((lat, lon), centroide).km > radio:
            area_fuera_del_circulo += area_hex
            continue

    # Para los casos mixtos, usamos integración
    try:
        area_integrada, _ = spi.dblquad(
            lambda y, x: 1 if poly.contains(Point(x, y)) and geodesic((lat,
lon), (y, x)).km <= radio else 0,
            min_lon_poly, max_lon_poly,
            lambda x: min_lat_poly, lambda x: max_lat_poly,
            epsabs=1.49e-03, epsrel=1.49e-03
        )
    except Exception:
        # En caso de error en la integración, estimamos con la proporción de
vértices
        area_integrada = area_hex * (vertices_dentro / len(vertices))

# Calcular proporción dentro del círculo
area_bbox = (max_lon_poly - min_lon_poly) * (max_lat_poly - min_lat_poly)
proporcion_dentro = area_integrada / area_bbox if area_bbox > 0 else 0
```

```

# Actualizar áreas
area_dentro_del_circulo += area_hex * proporcion_dentro
area_fuera_del_circulo += area_hex * (1 - proporcion_dentro)

# Calcular área no cubierta
area_no_cubierta = max(0, area_circulo - area_dentro_del_circulo)

# Guardar resultados de esta ubicación
resultados["num_celdas"].append(len(celdas_finales))
resultados["area_circulo"].append(area_circulo)
resultados["area_total_hexagonos"].append(len(celdas_finales) * area_hex)
resultados["area_dentro_circulo"].append(area_dentro_del_circulo)
resultados["area_no_cubierta"].append(area_no_cubierta)
resultados["area_fuera_circulo"].append(area_fuera_del_circulo)

# Convertir a DataFrame para análisis estadístico
df_todos = pd.DataFrame(resultados)

# Calcular media y desviación estándar de cada métrica
medias = df_todos.mean()

# Crear DataFrame de resultados finales
df_resultados = pd.DataFrame({
    "Métrica": [
        "Número de celdas utilizadas",
        "Área del círculo (km²)",
        "Área total de hexágonos (km²)",
        "Área efectiva dentro del círculo (km²)",
        "Área no cubierta dentro del círculo (km²)",
        "Área de hexágonos fuera del círculo (km²)"
    ],
    "Valor medio": [
        medias["num_celdas"],
        medias["area_circulo"],
        medias["area_total_hexagonos"],
        medias["area_dentro_circulo"],
        medias["area_no_cubierta"],
        medias["area_fuera_circulo"]
    ]
})

# Calcular relaciones porcentuales
df_porcentajes = pd.DataFrame({
    "Métrica": [
        "Proporción de área no cubierta respecto al círculo (%)",
        "Proporción de área fuera del círculo respecto al total de hexágonos (%)"
    ],
    "Valor medio": [
        100 * medias["area_no_cubierta"] / medias["area_circulo"],
        100 * medias["area_fuera_circulo"] / medias["area_total_hexagonos"]
    ]
})

# Mostrar resultados
print("Resultados para múltiples ubicaciones aleatorias:")
print(df_resultados)
print("\nPorcentajes relativos:")
print(df_porcentajes)

```

ANEXO XI: CÓDIGO PARA EL CÁLCULO DE DESVÍOS EN UNA TRAYECTORIA CONCRETA

```
import h3
import numpy as np
import pandas as pd
from geopy.distance import geodesic

# Coordenadas de inicio y fin
lat_inicial, lon_inicial = 41.33, -11.95
lat_final, lon_final = 42.33, -9.00

# Solicitar resolución H3 al usuario
resolucion = int(input("Ingrese la resolución H3 (5-10): "))

# Calcular distancia real en línea recta
distancia_real = geodesic((lat_inicial, lon_inicial), (lat_final, lon_final)).km

# Calcular rumbo teórico
rumbo_teorico = np.arctan2(lon_final - lon_inicial, lat_final - lat_inicial) * 180 / np.pi
rumbo_teorico = (rumbo_teorico + 180) % 360 - 180

# Obtener celdas H3 para los puntos de inicio y fin
celda_inicial = h3.latlng_to_cell(lat_inicial, lon_inicial, resolucion)
celda_final = h3.latlng_to_cell(lat_final, lon_final, resolucion)

# Generar ruta H3
try:
    path = h3.grid_path_cells(celda_inicial, celda_final)
    trayectoria = [h3.cell_to_latlng(celda) for celda in path]
except h3.H3FailedError:
    print("No se pudo generar la trayectoria en celdas H3.")
    exit()

# Calcular distancia total en la cuadrícula H3
distancia_grid = sum(geodesic(trayectoria[i], trayectoria[i + 1]).km for i in range(len(trayectoria) - 1))

# Calcular rumbos intermedios y desviaciones respecto al teórico
datos_rumbos = []
for i in range(1, len(trayectoria)):
    lat1, lon1 = trayectoria[i - 1]
    lat2, lon2 = trayectoria[i]
    rumbo_intermedio = np.arctan2(lon2 - lon1, lat2 - lat1) * 180 / np.pi
    rumbo_intermedio = (rumbo_intermedio + 180) % 360 - 180
    diferencia = abs((rumbo_intermedio - rumbo_teorico + 180) % 360 - 180)
    desviacion_porcentaje = (diferencia / abs(rumbo_teorico)) * 100 if rumbo_teorico != 0 else 0
    datos_rumbos.append([path[i], rumbo_intermedio, desviacion_porcentaje])

# Crear DataFrame para mostrar la información
df_rumbos = pd.DataFrame(datos_rumbos, columns=["Celda H3", "Rumbo Intermedio (°)", "Desviación (%)"])

# Calcular métricas generales
media_desviacion_rumbo = np.mean([fila[2] for fila in datos_rumbos]) if datos_rumbos else 0
diferencia_distancia = ((distancia_grid - distancia_real) / distancia_real) * 100

# Mostrar resultados
```

```
print(f"Número de celdas en la ruta: {len(path)}")
print(f"Distancia real (km): {distancia_real:.2f}")
print(f"Distancia en celdas H3 (km): {distancia_grid:.2f}")
print(f"Diferencia de distancia (%): {diferencia_distancia:.2f}%")
print(f"Rumbo teórico (°): {rumbo_teorico:.2f}")
print(f"Desviación media de rumbo (%): {media_desviacion_rumbo:.2f}%")

# Mostrar tabla con rumbos intermedios y desviaciones
print(df_rumbos)
```

ANEXO XII: CÓDIGO PARA CÁLCULO DE DESVÍOS DE TRAYECTORIA PARA UNA RESOLUCIÓN

```
import h3
import numpy as np
import pandas as pd
import random
from geopy.distance import geodesic
import math

# Configuración
num_trayectorias = 10000 # Número de trayectorias aleatorias
max_dist_km = 500 # Máxima distancia permitida en km
resolucion = int(input("Ingrese la resolución H3 (5-10): "))

# Almacenar datos
desviaciones_totales = []
desviaciones_distancia = []
celdas_totales = []

for _ in range(num_trayectorias):
    # Generar puntos aleatorios sin restricción de dirección
    while True:
        lat_inicial = random.uniform(-80, 80) # Evitar latitudes extremas
        lon_inicial = random.uniform(-180, 180)

        # Generar un rumbo aleatorio de 0 a 360 grados
        rumbo_deseado = random.uniform(0, 360)

        # Calcular segundo punto basado en una distancia y rumbo
        dist_km = max_dist_km

        # Aproximación para calcular el segundo punto dado el rumbo y distancia
        earth_radius = 6371.0 # Radio de la Tierra en km

        # Convertir distancia a radianes
        dist_rad = dist_km / earth_radius

        # Convertir rumbo a radianes
        rumbo_rad = math.radians(rumbo_deseado)

        # Convertir coordenadas a radianes
        lat1_rad = math.radians(lat_inicial)
        lon1_rad = math.radians(lon_inicial)

        # Calcular nuevo punto
        lat2_rad = math.asin(math.sin(lat1_rad) * math.cos(dist_rad) +
                             math.cos(lat1_rad) * math.sin(dist_rad) *
                             math.cos(rumbo_rad))

        lon2_rad = lon1_rad + math.atan2(math.sin(rumbo_rad) * math.sin(dist_rad)
                                         * math.cos(lat1_rad),
                                         math.cos(dist_rad) - math.sin(lat1_rad) *
                                         math.sin(lat2_rad))

        # Convertir de vuelta a grados
        lat_final = math.degrees(lat2_rad)
        lon_final = math.degrees(lon2_rad)

        # Normalizar longitud
        lon_final = ((lon_final + 180) % 360) - 180
```

```

        # Verificar distancia
        distancia_real = geodesic((lat_inicial, lon_inicial), (lat_final,
lon_final)).km

        if distancia_real <= max_dist_km:
            break

    # Calcular rumbo teórico real (sin usar función)
    # Convertir a radianes
    lat1_rad = math.radians(lat_inicial)
    lon1_rad = math.radians(lon_inicial)
    lat2_rad = math.radians(lat_final)
    lon2_rad = math.radians(lon_final)

    # Calcular rumbo
    y = math.sin(lon2_rad - lon1_rad) * math.cos(lat2_rad)
    x = math.cos(lat1_rad) * math.sin(lat2_rad) - math.sin(lat1_rad) *
math.cos(lat2_rad) * math.cos(lon2_rad - lon1_rad)
    rumbo_rad = math.atan2(y, x)

    # Convertir a grados y normalizar
    rumbo_teorico = (math.degrees(rumbo_rad) + 360) % 360

    # Convertir a celdas H3
    celda_inicial = h3.latlng_to_cell(lat_inicial, lon_inicial, resolucion)
    celda_final = h3.latlng_to_cell(lat_final, lon_final, resolucion)

    # Generación de la trayectoria usando H3
    try:
        path = h3.grid_path_cells(celda_inicial, celda_final)
        trayectoria = [h3.cell_to_latlng(celda) for celda in path]
    except h3.H3FailedError:
        continue

    # Calcular distancia en la cuadrícula
    distancia_grid = sum(geodesic(trayectoria[i], trayectoria[i + 1])).km for i in
range(len(trayectoria) - 1))

    # Calcular desviaciones
    desviaciones = []
    for i in range(1, len(trayectoria)):
        lat1, lon1 = trayectoria[i - 1]
        lat2, lon2 = trayectoria[i]

        # Calcular rumbo intermedio (sin usar función)
        # Convertir a radianes
        lat1_rad = math.radians(lat1)
        lon1_rad = math.radians(lon1)
        lat2_rad = math.radians(lat2)
        lon2_rad = math.radians(lon2)

        # Calcular rumbo
        y = math.sin(lon2_rad - lon1_rad) * math.cos(lat2_rad)
        x = math.cos(lat1_rad) * math.sin(lat2_rad) - math.sin(lat1_rad) *
math.cos(lat2_rad) * math.cos(lon2_rad - lon1_rad)
        rumbo_rad = math.atan2(y, x)

        # Convertir a grados y normalizar
        rumbo_intermedio = (math.degrees(rumbo_rad) + 360) % 360

```

```
# Calcular diferencia angular considerando la naturaleza circular de los
ángulos
diferencia = min((rumbo_intermedio - rumbo_teorico) % 360, (rumbo_teorico
- rumbo_intermedio) % 360)

# Calcular desviación como porcentaje respecto a 180 grados (máxima
desviación posible)
desvio_porcentaje = (diferencia / 180) * 100

desviaciones.append(desvio_porcentaje)

desviaciones_totales.extend(desviaciones)
desviaciones_distancia.append((distancia_grid - distancia_real) /
distancia_real * 100)
celdas_totales.append(len(path))

# Calcular métricas
media_desviacion_rumbo = np.mean(desviaciones_totales) if desviaciones_totales
else 0
media_desviacion_distancia = np.mean(desviaciones_distancia) if
desviaciones_distancia else 0
media_celdas = np.mean(celdas_totales) if celdas_totales else 0

# Mostrar métricas
print(f"Desviación media de rumbo: {media_desviacion_rumbo:.2f}%")
print(f"Desviación media de distancia: {media_desviacion_distancia:.2f}%")
print(f"Número medio de celdas por trayectoria: {media_celdas:.2f}")
```

ANEXO XIII: CÓDIGO PARA LA COMPARACIÓN DE TRANSPONEDORES AIS EN UNA TRAYECTORIA

```
# Reimportar las librerías tras el reinicio del entorno
import numpy as np
import pandas as pd
import h3
from geopy.distance import geodesic

# Definir los puntos de inicio y fin de la trayectoria
start_point = (41.33, -11.95)
end_point = (42.33, -9.00)

# Parámetros generales
speed = 15 # Velocidad del barco en nudos
ais_classes = ["A", "B", "B+"] # Tipos de transpondedores AIS

# Calcular la distancia total del viaje
total_distance = geodesic(start_point, end_point).km # Distancia total en km

# Crear una lista para almacenar los resultados
results = []

# Analizar todas las resoluciones H3 de 0 a 15 para cada tipo de transpondedor AIS
for ais_class in ais_classes:
    # Determinar la frecuencia de transmisión según la velocidad y el tipo de AIS
    if speed <= 2:
        interval = 180 # 3 minutos
    elif 2 < speed <= 14:
        interval = {"A": 10, "B": 30, "B+": 30}[ais_class] # segundos
    elif 14 < speed <= 23:
        interval = {"A": 6, "B": 30, "B+": 15}[ais_class] # segundos
    else:
        interval = {"A": 2, "B": 30, "B+": 5}[ais_class] # segundos

    for h3_resolution in range(16):
        # Calcular el número de transmisiones
        num_transmissions = int((total_distance / (speed * 1.852)) / (interval /
3600)) # Convertir nudos a km/h

        # Generar las posiciones de transmisión a lo largo de la ruta
        lats = np.linspace(start_point[0], end_point[0], num_transmissions)
        lons = np.linspace(start_point[1], end_point[1], num_transmissions)
        transmission_positions = list(zip(lats, lons))

        # Determinar las celdas en las que realmente se transmitió
        transmission_cells = {}
        for lat, lon in transmission_positions:
            cell = h3.latlng_to_cell(lat, lon, h3_resolution)
            if cell not in transmission_cells:
                transmission_cells[cell] = 0
            transmission_cells[cell] += 1 # Contamos transmisiones en cada celda

        # Obtener las celdas teóricas por las que debería pasar el barco
        start_cell = h3.latlng_to_cell(start_point[0], start_point[1],
h3_resolution)
        end_cell = h3.latlng_to_cell(end_point[0], end_point[1], h3_resolution)
        theoretical_cells = h3.grid_path_cells(start_cell, end_cell)

        # Calcular el número medio de transmisiones por celda
```

```
total_transmissions = sum(transmission_cells.values())
num_transmitting_cells = len(transmission_cells)
avg_transmissions_per_cell = total_transmissions / num_transmitting_cells
if num_transmitting_cells > 0 else 0

# Guardar los resultados en la lista
results.append({
    "AIS_Class": ais_class,
    "H3_Resolution": h3_resolution,
    "Num_Transmissions": num_transmissions,
    "Num_Theoretical_Cells": len(theoretical_cells),
    "Num_Transmitting_Cells": len(transmission_cells),
    "Missing_Transmissions": sum(1 for cell in theoretical_cells if cell
not in transmission_cells),
    "Extra_Transmissions": sum(1 for cell in transmission_cells if cell
not in theoretical_cells),
    "Avg_Transmissions_Per_Cell": avg_transmissions_per_cell
})

# Convertir los resultados en un DataFrame y guardarlo como CSV
df_results = pd.DataFrame(results)
csv_file_path = "h3_resolution_analysis_ais_class.csv"
df_results.to_csv(csv_file_path, index=False)

# Mostrar la ubicación del archivo generado
csv_file_path
```