



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Simulación de comportamientos de clases de buques a partir de
datos AIS*

Grado en Ingeniería Mecánica

ALUMNO: Javier González Enríquez

DIRECTORES: José Ignacio Valles Cancela

CURSO ACADÉMICO: 2025-2026

Universida_{de}Vigo



Centro Universitario de la Defensa en la Escuela Naval Militar

TRABAJO FIN DE GRADO

*Simulación de comportamientos de clases de buques a partir de
datos AIS*

Grado en Ingeniería Mecánica
Intensificación en Tecnología Naval
Cuerpo General

UniversidadeVigo

RESUMEN

El análisis del tráfico marítimo mediante datos del Sistema de Identificación Automática (AIS) es una herramienta clave para comprender patrones de navegación y apoyar la vigilancia y gestión del entorno marítimo. En este Trabajo de Fin de Grado se aborda la simulación del comportamiento de distintas clases de buques a partir de datos AIS reales, con el objetivo de generar trayectorias sintéticas que reproduzcan de forma realista las dinámicas observadas en un entorno portuario complejo.

El estudio se desarrolla sobre un conjunto de datos AIS correspondiente al entorno del Puerto de Barcelona. A partir de estos datos se caracterizan patrones de movimiento de buques tipo *Cargo* y buques tipo *Fishing*, cuyas dinámicas difieren significativamente en estructura y variabilidad.

Sobre esta base se implementan y comparan distintos enfoques de modelado para la generación de trayectorias sintéticas, incluyendo métodos probabilísticos basados en discretización espacial mediante cuadrículas H3 y cadenas de Markov, así como modelos generativos basados en aprendizaje profundo, entre ellos redes generativas antagónicas condicionales (cGAN) y arquitecturas recurrentes LSTM. Los resultados se evalúan mediante métricas cuantitativas y análisis visual, permitiendo comparar el rendimiento de los modelos y determinar qué metodología resulta más adecuada para cada tipología de buque.

PALABRAS CLAVE

AIS, Tráfico Marítimo, Modelos generativos, Aprendizaje automático, Trayectorias sintéticas

AGRADECIMIENTOS

En primer lugar, me gustaría expresar mi más sincero agradecimiento a mi tutor, el Capitán de Corbeta Valles, por su orientación, incansable ayuda, plena disponibilidad, resiliencia cuando las cosas no salían y esfuerzo para que funcionasen durante todo el desarrollo de este Trabajo Fin de Grado.

También quiero agradecer a mi familia, por su apoyo constante durante todos estos años en la Escuela Naval Militar. Su paciencia y ánimo han sido un pilar fundamental que me ha permitido afrontar esta etapa.

No me quiero olvidar de mis amigos y compañeros, con quienes he compartido estos años de formación tanto dentro como fuera de la Escuela. Su apoyo, nuestras vivencias y sufrimiento juntos han hecho que este camino haya sido mucho más llevadero y, sin duda, inolvidable.

A todos ellos, gracias por haber formado parte de esta etapa y por haber contribuido, de una u otra manera, a la realización de este trabajo y supervivencia en esta Escuela cuando la motivación fallaba.

CONTENIDO

Contenido	1
Índice de Figuras	6
Índice de Tablas.....	9
Índice de ecuaciones.....	10
1 Introducción y objetivos	11
1.1 Motivación	11
1.2 Objetivos	12
2 Estado del arte	13
2.1 El sistema AIS como fuente de datos para el análisis del tráfico marítimo.....	13
2.1.1 Funcionamiento del sistema AIS	13
2.1.2 Conjunto de datos AIS y áreas de estudio	15
2.2 Tipificación de buques y caracterización de comportamientos	16
2.2.1 Buques de carga	16
2.2.2 Petroleros y buques tanque	17
2.2.3 Buques de pasaje.....	18
2.2.4 Veleros y embarcaciones de recreo	18
2.2.5 Buques pesqueros	19
2.2.6 Embarcaciones de servicio y apoyo portuario	19
2.3 Representación espacial del tráfico marítimo	20
2.3.1 Técnicas de discretización espacial	20
2.3.2 Cuadrículas tipo UBER aplicadas al tráfico marítimo	21
2.3.3 Modelos probabilísticos de transición espacial: cadenas de Markov	22
2.4 Inteligencia Artificial y Machine Learning.....	23
2.4.1 Inteligencia Artificial y Aprendizaje Automático: conceptos generales	23
2.4.2 Aplicación del Machine Learning al análisis de datos	24
2.4.2.1 Aprendizaje supervisado.....	24
2.4.2.2 Aprendizaje no supervisado.....	26
2.4.2.3 Aprendizaje semisupervisado	27
2.4.2.4 Aprendizaje por refuerzo	28
2.4.3 Algoritmo de K-nearest neighbors (KNN)	29
2.5 Redes neuronales.....	30
2.5.1 Redes neuronales artificiales: fundamentos.....	30
2.5.2 Redes neuronales para datos secuenciales y temporales	32

2.5.3 Redes neuronales recurrentes (RNN)	32
2.6 Redes neuronales recurrentes avanzadas	33
2.6.1 Redes neuronales recurrentes LSTM.....	33
2.6.2 Transformer	34
2.7 Modelos generativos aplicados a secuencias temporales.....	35
2.7.1 Modelos generativos en aprendizaje profundo	35
2.7.2 Redes Generativas Antagónicas (GAN) para secuencias temporales.....	36
2.7.3 GAN secuenciales basadas en LSTM.....	38
2.7.4 Redes GAN condicionales (cGAN).....	38
2.8 Análisis del entorno marítimo de interés en la Armada.....	40
2.8.1 Centro de Operaciones y Vigilancia de Acción Marítima de la Armada (COVAM).....	40
2.8.2 Uso de Inteligencia Artificial para el Conocimiento del Entorno Marítimo (CEM).....	42
3 Desarrollo del TFG.....	45
3.1 Entorno del trabajo y software empleado	45
3.1.1 Jupyter Lab	45
3.1.2 Lenguaje de programación Python	46
3.1.3 MiniConda	46
3.2 Definición del área de estudio y criterios de filtrado espacial	47
3.2.1 Puerto de Barcelona.....	47
3.2.2 Definición de polígonos y zonas de filtrado portuario	47
3.2.3 Delimitación geográfica y radio operativo de 30 millas náuticas.....	49
3.2.4 Polígono específico para buques pesqueros	51
3.3 Construcción y preprocesamiento del conjunto de datos AIS	52
3.3.1 Selección y análisis de mensajes dinámicos	52
3.3.2 Proceso de filtrado inicial por tipología de buque	53
3.3.3 Construcción del dataset “Cargo entrada”	54
3.3.4 Construcción del dataset “Fishing”	56
3.4 Generación de trayectorias sintéticas mediante discretización espacial H3 y mapas de calor.....	57
3.4.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona.....	57
3.4.1.1 Visualización y validación de trayectorias reales.....	57
3.4.1.2 Generación de cuadrículas H3 tipo UBER.....	59
3.4.1.3 Transiciones entre celdas.....	60
3.4.1.4 Establecimiento de celdas objetivo y celdas iniciales	61
3.4.1.5 Aplicación de mejoras de la trayectoria	63
3.4.1.6 Generador de trayectorias basado en modelo Markoviano	64
3.4.1.7 Transformación de celdas a posiciones geográficas.....	65

3.4.1.8 Suavizado de la trayectoria.....	66
3.4.1.9 Generación de trayectoria final sin velocidad	67
3.4.1.10 Construcción de mapas de calor y asignación de velocidades	68
3.4.2 Buques tipo Fishing que salen, faenan y entran en las inmediaciones del Puerto de Barcelona.....	73
3.4.2.1 Discretización espacial del tráfico pesquero mediante cuadrículas tipo UBER(H3)	73
3.4.2.2 Definición de vecindad y transiciones entre celdas H3	74
3.4.2.3 Cálculo y normalización de la tasa de ocupación de celdas H3	74
3.4.2.4 Identificación de zonas de salida, faena y entrada a puerto.....	75
3.4.2.4.1 Generación de mapa de calor para análisis de zonas de pesca.....	75
3.4.2.4.2 Identificación de las zonas de pesca	76
3.4.2.5 Establecimiento de celdas objetivo y celdas iniciales	78
3.4.2.6 Generación de trayectorias sintéticas	79
3.4.2.7 Asignación de velocidades a las trayectorias	83
3.5 Generación de trayectorias sintéticas mediante redes generativas antagónicas condicionales (cGAN) y estimación cinemática con KNN	84
3.5.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona.....	84
3.5.1.1 Preparación del conjunto de datos para el entrenamiento de la cGAN	84
3.5.1.1.1 Definición de las condiciones	85
3.5.1.1.2 Cálculo de incrementos temporales	86
3.5.1.1.3 Normalización de datos.....	86
3.5.1.2 Arquitectura del generador del cGAN.....	87
3.5.1.3 Arquitectura del discriminador del cGAN.....	88
3.5.1.4 Proceso de entrenamiento adversarial	89
3.5.1.5 Reconstrucción y postprocesado geométrico de trayectorias	92
3.5.1.6 Estimación de velocidades mediante K-Nearest Neighbors (KNN)	93
3.5.1.7 Generación final de trayectorias cGAN + KNN.....	94
3.6 Generación conjunta de trayectoria y velocidad mediante redes generativas antagónicas condicionales (cGAN)	96
3.6.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona.....	96
3.6.1.1 Preparación del conjunto de datos para la cGAN espacio-dinámica.....	96
3.6.1.1.1 Definición de las condiciones	98
3.6.1.1.2 Cálculo de incrementos temporales	98
3.6.1.1.3 Normalización de los datos	98
3.6.1.2 Generador condicional espacio-dinámico	99
3.6.1.3 Arquitectura del discriminador	100

3.6.1.4 Configuración y estrategia de entrenamiento	100
3.6.1.5 Reconstrucción de trayectorias absolutas	101
3.7 Modelado de trayectorias de buques tipo Fishing mediante LSTM	103
3.7.1 Limitaciones del enfoque cGAN en buques pesqueros	103
3.7.2 Justificación del modelo LSTM autorregresivo.....	103
3.7.3 Discretización espacial mediante cuadrículas H3.....	104
3.7.4 Construcción del vocabulario espacial	105
3.7.5 Formulación del modelo LSTM autorregresivo	106
3.7.5.1 Planteamiento secuencial del problema y construcción del dataset de entrenamiento	106
3.7.5.2 Arquitectura del modelo	106
3.7.5.2.1 Capa de embedding.....	106
3.7.5.2.2 Red LSTM (Long Short-Term Memory).....	107
3.7.5.2.3 Capa lineal de salida	107
3.7.6 Proceso de entrenamiento	107
3.7.7 Procedimiento autorregresivo de generación.....	108
3.7.8 Comparación visual entre trayectorias reales y sintéticas	109
3.7.9 Suavizado y remuestreo espacial de trayectorias.....	109
4 Resultados y discusión	113
4.1 Métricas para la comparación de trayectorias.....	113
4.1.1 Distancia de Fréchet	114
4.1.2 Comparación espacial mediante mapas de calor	115
4.1.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos..	115
4.1.3.1 MAE (Mean Absolute Error).....	115
4.1.3.2 DTW (Dynamic Time Warping)	116
4.1.3.3 Cálculos estadísticos de la velocidad.....	117
4.2 Validación en buques tipo Cargo	118
4.2.1 Distancia de Fréchet	118
4.2.1.1 Unidad de cálculo	118
4.2.1.2 Obtención de resultados.....	118
4.2.1.3 Resultados.....	119
4.2.1.4 Método seleccionado según la distancia Fréchet.....	122
4.2.2 Comparación espacial mediante mapas de calor	122
4.2.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos..	124
4.2.3.1 Análisis gráfico de la velocidad.....	124
4.2.3.2 Error absoluto medio (MAE).....	127

4.2.3.3 Alineamiento dinámico del tiempo (DTW).....	127
4.2.3.4 Cálculos estadísticos.....	128
4.2.3.5 Método seleccionado según velocidad	128
4.3 Validación en buques tipo Fishing.....	130
4.3.1 Distancia de Fréchet	131
4.3.1.1 Resultados.....	131
4.3.1.2 Método seleccionado según la distancia Fréchet.....	134
4.3.2 Comparación espacial mediante mapas de calor	134
4.3.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos..	136
4.3.3.1 Análisis gráfico de la velocidad.....	136
4.3.3.2 Error absoluto medio (MAE).....	137
4.3.3.3 Alineamiento dinámico del tiempo (DTW).....	138
4.3.3.4 Cálculos estadísticos.....	138
4.3.3.5 Método seleccionado según velocidad	139
4.3.4 Selección del método óptimo para buques tipo Fishing	139
4.4 Representación simulada del entorno operativo del Puerto de Barcelona	140
4.5 Limitaciones del análisis.....	141
5 Conclusiones y líneas futuras	143
5.1 Conclusiones	143
5.2 Líneas futuras	144
6 Bibliografía.....	145
Anexo I: Implicaciones Sociales, y/o Económicas, y/o Ambientales	152
Anexo II: Reflexiones Éticas y Sociales	153
Anexo III: SIGLAS, ACRÓNIMOS Y ABREVIATURAS	154
Anexo IV: CÓDIGO NÚMÉRICO DE LA TIPIFICACIÓN DE BUQUES EN LOS DATOS AIS	156
Anexo V: ESTRUCTURA DE UN MENSAJE AIS.....	157
Anexo VI: ACCESO AL CÓDIGO DESARROLLADO DE LOS MODELOS GENERATIVOS	161
Anexo VII: DECLARACIÓN DE USO DE INTELIGENCIA ARTIFICIAL	162

ÍNDICE DE FIGURAS

Figura 2-1 Tipos de información AIS (Fuente: propia generada con IA)	14
Figura 2-2 Esquema del funcionamiento del sistema AIS y transmisión de mensajes [33].....	14
Figura 2-3 Distribución global de mensajes AIS registrados a escala mundial [34].....	15
Figura 2-4 Buque portacontenedores entrando en puerto [37]	17
Figura 2-5 Gasero en operaciones portuarias [36]	17
Figura 2-6 Crucero navegando en aguas del Mar Mediterráneo [35]	18
Figura 2-7 Velero navegando a vela [31]	18
Figura 2-8 Pesquero faenando en las Rías Baixas (Fuente: propia generada con IA)	19
Figura 2-9 Embarcación de práctico dirigiéndose a un buque (Fuente: propia generada con IA)...	20
Figura 2-10 Ejemplo conceptual de discretización espacial mediante rejilla regular aplicada a trayectorias de buques (Fuente: propia generada con IA).....	21
Figura 2-11 Ejemplo de discretización espacial mediante cuadrícula hexagonal jerárquica tipo H3 [12]	21
Figura 2-12 Cadena simple de Markov [69]	22
Figura 2-13 Evolución histórica de la IA (Fuente: propia generada con IA).....	23
Figura 2-14 Ejemplo de aprendizaje supervisado [46].....	25
Figura 2-15 CAPTCHA [45].....	26
Figura 2-16 Comparación de aprendizaje no supervisado con aprendizaje supervisado [48]	27
Figura 2-17 Esquema sencillo del aprendizaje semisupervisado[50].....	28
Figura 2-18 Aprendizaje por refuerzo [51]	29
Figura 2-19 Ejemplo de KNN para K=3 (Fuente: propia)	30
Figura 2-20 Estructura básica de una red neuronal artificial [38]	31
Figura 2-21 Esquema del funcionamiento de una neurona artificial [39]	31
Figura 2-22 Esquema de funcionamiento de una red neuronal recurrente desplegada en el tiempo (Fuente: propia generada con IA)	33
Figura 2-23 Comparación entre una RNN y una unidad LSTM, mostrando la estructura interna de la celda de memoria y el funcionamiento de sus puertas de control [32]	34
Figura 2-24 Esquema básico del funcionamiento de un Transformer [40]	35
Figura 2-25 Proceso general de aprendizaje y generación en modelos generativos profundos (Fuente: propia generada con IA).....	36
Figura 2-26 Esquema conceptual de una GAN aplicada a secuencias temporales (Fuente: propia generada con IA).....	37
Figura 2-27 Representación de cGAN (Fuente: propia generada con IA)	39
Figura 2-28 Fragata F-104 "Méndez Núñez" escoltando a un pesquero [41]	40
Figura 2-29 Centro de Operaciones y Vigilancia de la Acción Marítima (COVAM) [30].....	41

Figura 2-30 Zonas interés marítimo nacional [25]	42
Figura 3-1 Interfaz para la escritura de código en Jupyter Lab (Fuente: propia)	46
Figura 3-2 Plano del Puerto de Barcelona [60]	47
Figura 3-3 Ejemplo del uso de Polyline tool (Fuente: propia)	48
Figura 3-4 Polígonos filtradores aplicados al puerto de Barcelona (Fuente: propia)	48
Figura 3-5 Áreas generadas para el filtrado y condiciones (Fuente: propia)	49
Figura 3-6 Línea de costa generada junto al radio de 30mn (Fuente: propia)	50
Figura 3-7 Representación final de los mapas mostrados en pantalla (Fuente: propia)	50
Figura 3-8 Puertos en la costa de Barcelona (Fuente: propia con [62][63])	51
Figura 3-9 Polígono específico para buques pesqueros (Fuente: propia)	51
Figura 3-10 Ejemplo Dataset CSV " <i>Cargo</i> Entrada" (Fuente: propia)	55
Figura 3-11 Trayectorias filtradas con status 7 (Fuente: propia)	57
Figura 3-12 Trayectoria real según serie temporal (Fuente: propia)	58
Figura 3-13 Trayectoria real según la velocidad (Fuente: propia)	58
Figura 3-14 Cuadrícula H3 tipo UBER [67]	59
Figura 3-15 Puntos geográficos asociados a celdas (Fuente: propia con [67])	60
Figura 3-16 Sistema de transiciones entre celdas vecinas (Fuente: propia)	61
Figura 3-17 Celdas objetivo de buques tipo <i>Cargo</i> (Fuente: propia)	62
Figura 3-18 Trayectoria sintética con celdas tipo UBER (Fuente: propia)	65
Figura 3-19 Trayectoria sintética sin suavizar (Fuente: propia)	66
Figura 3-20 Trayectoria sintética suavizada (Fuente: propia)	67
Figura 3-21 Ejemplo trayectoria sintética mediante mapas de calor y cuadrículas tipo UBER (Fuente: propia)	68
Figura 3-22 Generación de 400 trayectorias sintéticas (Fuente: propia)	68
Figura 3-23 Mapa de calor según velocidad (Fuente: propia)	70
Figura 3-24 Trayectoria con velocidad sin suavizar (Fuente: propia)	71
Figura 3-25 Trayectoria con velocidad suavizada (Fuente: propia)	71
Figura 3-26 Perfil de velocidad sintético discretizado (Fuente: propia)	72
Figura 3-27 Resultado final método mapas de calor y celdas tipo H3 (Fuente: propia)	73
Figura 3-28 Mapa de calor según velocidad media de pesqueros (Fuente: propia)	76
Figura 3-29 Función gaussiana de velocidades (Fuente: propia)	77
Figura 3-30 Celdas más probables según su probabilidad de ser zona de pesca (Fuente: propia)	78
Figura 3-31 Conjunto de celdas iniciales y objetivo de los buques de pesca (Fuente: propia)	79
Figura 3-32 Trayectorias sintéticas de tres pesqueros empleando mapas de calor (Fuente: propia)	81
Figura 3-33 200 trayectorias sintéticas mediante mapas de calor y celdas H3 (Fuente: propia)	82

Figura 3-34 Comparación de trayectorias generadas mediante mapas de calor y trayectorias reales (Fuente: propia).....	83
Figura 3-35 Representación de trayectoria con velocidad asignada de los pesqueros (Fuente: propia)	83
Figura 3-36 Trayectoria completa de buque pesquero con gráficas de rumbo y velocidad (Fuente: propia)	84
Figura 3-37 Incremento de latitud y longitud (Fuente: propia).....	86
Figura 3-38 Esquema del generador cGAN (Fuente: propia)	88
Figura 3-39 Generación de trayectorias sintéticas con cGAN (Fuente: propia)	93
Figura 3-40 Trayectorias generadas con cGAN junto con KNN (Fuente: propia).....	95
Figura 3-41 Trayectoria generada con cGAN y KNN junto con sus gráficas de velocidad y rumbo (Fuente: propia).....	96
Figura 3-42 Incremento de latitud, longitud y velocidad (Fuente: propia)	98
Figura 3-43 400 trayectorias generadas con cGAN (Fuente: propia).....	102
Figura 3-44 Trayectoria sintética generada con cGAN con gráficos de rumbos y velocidades (Fuente: propia)	103
Figura 3-45 comparación de 50 trayectorias sintéticas con 50 trayectorias reales (Fuente: propia)	109
Figura 3-46 trayectorias sintéticas suavizadas comparadas con trayectorias reales (Fuente: propia)	110
Figura 3-47 50 trayectorias sintéticas en comparación con todas las reales (Fuente: propia)	111
Figura 3-48 Rutas generadas con asignación de velocidad (Fuente: propia)	112
Figura 3-49 Presentación de trayectoria sintética completa (Fuente: propia)	112
Figura 4-1 Ejemplo de distancia Fréchet [72]	114
Figura 4-2 Distancia Fréchet semejada a un perro y una correa[72].....	114
Figura 4-3 Ejemplo de aplicación de DTW entre dos trayectorias[76].....	116
Figura 4-4 Gráfica resultados de distancia de Fréchet (Fuente: propia)	122
Figura 4-5 Comparación visual de la distribución espacial de trayectorias mediante mapas de calor (Fuente: propia).....	123
Figura 4-6 Perfil de velocidades según método (Fuente: propia).....	124
Figura 4-7 Gráfica de las velocidades finales según método (Fuente: propia)	126
Figura 4-8 Trayectorias generadas con los distintos métodos Cargo (Fuente: propia)	130
Figura 4-9 Distribución de la distancia de Fréchet para trayectorias de buques tipo <i>Fishing</i> (Fuente: propia)	133
Figura 4-10 Comparación modelos con mapas de calor (Fuente: propia).....	134
Figura 4-11 Perfil de velocidad buques pesqueros (Fuente: propia).....	136
Figura 4-12 Trayectorias generadas con los distintos métodos Fishing (Fuente: propia).....	140
Figura 4-13 Ejemplo simulado de entorno operativo de Barcelona	141

ÍNDICE DE TABLAS

Tabla 3-1 Parámetros estadísticos	69
Tabla 3-2 Velocidades calculadas	70
Tabla 3-3 Velocidades calculadas	76
Tabla 3-4 Estructura de 3 dimensiones	85
Tabla 3-5 Representación en dos dimensiones.....	85
Tabla 3-6 Estructura de tres dimensiones.....	86
Tabla 3-7 Parámetros elegidos para el entrenamiento.....	90
Tabla 3-8 Parámetros de control.....	97
Tabla 3-9 Parámetros tras filtrado	97
Tabla 3-10 Estructura de dos dimensiones.....	98
Tabla 3-11 Parámetros elegidos entrenamiento	99
Tabla 3-12 Número de celdas.....	104
Tabla 3-13 Palabras especiales.....	105
Tabla 4-1 Resultados de Distancia de Fréchet	119
Tabla 4-2 Cálculo porcentaje diferencia respecto trayectorias reales	120
Tabla 4-3 Clasificación de modelos	121
Tabla 4-4 Resultados MAE	127
Tabla 4-5 Resultados DTW	127
Tabla 4-6 Media y desviación estándar según método.....	128
Tabla 4-7 Evaluación comparativa de los modelos analizados	129
Tabla 4-8 Leyenda.....	129
Tabla 4-9 Resultados distancia Fréchet.....	131
Tabla 4-10 Cálculo porcentaje diferencia respecto trayectorias reales	131
Tabla 4-11 Clasificación de los modelos	132
Tabla 4-12 Resultados MAE pesqueros	137
Tabla 4-13 Resultados DTW pesqueros	138
Tabla 4-14 Cálculos estadísticos de pesqueros	138
Tabla 4-15 Comparación modelos tipo pesqueros	139
Tabla 4-16 Leyenda.....	139
Tabla 0-1 Código numérico según tipo de buque en los mensajes AIS [53]	156
Tabla 0-1 Estructura de un mensaje AIS	160

ÍNDICE DE ECUACIONES

Ecuación 1 54

Ecuación 2 62

Ecuación 3 62

Ecuación 4 62

Ecuación 5 74

Ecuación 6 116

Ecuación 7 117

Ecuación 8 117

Ecuación 9 117

1 INTRODUCCIÓN Y OBJETIVOS

1.1 Motivación

El tráfico marítimo es uno de los elementos fundamentales del comercio internacional y de la actividad económica global, concentrándose en puertos y zonas costeras una elevada densidad de buques de distintos tipos y funciones. La gestión eficiente de este tráfico, así como la garantizar la seguridad marítima, requieren de un conocimiento detallado y actualizado del comportamiento de los buques que operan en estos entornos, especialmente en áreas caracterizadas por una alta complejidad operativa.

El Sistema de Identificación Automática (*Automatic Identification System*, AIS) se ha consolidado como una de las principales fuentes de información para el control y seguimiento del tráfico marítimo a escala mundial. Este sistema proporciona de manera continuada e ininterrumpida datos tanto dinámicos como estáticos de los buques, tales como posición, rumbo, velocidad o tipo de embarcación, generando grandes volúmenes de información. El análisis de estos datos permite el estudio de patrones de movimiento, caracterizar comportamientos habituales y mejorar la comprensión de la dinámica del tráfico marítimo en determinadas zonas geográficas. No obstante, el volumen y la complejidad de los datos AIS hacen necesario el uso de técnicas avanzadas de análisis y procesamiento de datos que permitan extraer información relevante de manera eficiente y reproducible.

En el ámbito de la Ingeniería Naval y de la Defensa, la capacidad de analizar y simular el comportamiento de los buques a partir de datos reales responde a una necesidad creciente asociada al conocimiento y control del entorno marítimo. El análisis sistemático del tráfico marítimo, y en particular de los comportamientos asociados a distintas clases de buques, resulta fundamental para disponer de una visión precisa y actualizada de la situación marítima en nuestro entorno. Esta información constituye la base sobre la que se apoyan los sistemas de vigilancia y control del tráfico, como los empleados por el Centro de Operaciones de Vigilancia y Acción Marítima (COVAM) de la Armada, cuya eficacia depende, en gran medida, de la calidad del análisis previo de la zona de operaciones. La calidad de dicha imagen depende directamente del rigor en el tratamiento de los datos y de la capacidad para modelar con fidelidad las dinámicas observadas en el entorno de operaciones.

Asimismo, la simulación de comportamientos de buques a partir de datos AIS permite generar escenarios de navegación sintéticos, pero estadísticamente coherentes con el tráfico real, que pueden integrarse en los simuladores de navegación empleados en las distintas escuelas de la Armada, contribuyendo a mejorar la formación y el adiestramiento al reproducir situaciones operativas basadas en tráfico real. Este enfoque reduce la dependencia de escenarios teóricos o simplificados y resulta

especialmente relevante en entornos portuarios de gran actividad, donde la diversidad de tráfico y maniobras incrementa notablemente la complejidad del análisis y de la operación.

En este contexto se enmarca el presente Trabajo de Fin de Grado, centrado en la **simulación de comportamientos de clases de buques a partir de datos AIS**, tomando como caso de estudio el Puerto de Barcelona. Este puerto, uno de los principales del Mar Mediterráneo, constituye un entorno real y representativo debido a su elevada densidad de tráfico y a los distintos tipos de buques que operan en él, lo que lo convierte en un escenario adecuado para el análisis y modelado del comportamiento marítimo aplicado a la vida real.

El desarrollo del trabajo se apoya en el uso de herramientas de programación y análisis de datos, permitiendo abordar el tratamiento de grandes volúmenes de información y la generación de modelos de simulación del tráfico marítimo. De este modo, el trabajo integra conocimientos de análisis de datos, modelado computacional y sistemas marítimos, enmarcándose en las competencias propias del Grado en Ingeniería Mecánica con intensificación en Tecnología Naval y alineándose con el interés del Centro Universitario de la Defensa en la Escuela Naval Militar y de la Armada por el desarrollo de capacidades tecnológicas aplicadas al ámbito marítimo.

1.2 Objetivos

El objetivo principal del presente Trabajo de Fin de Grado es desarrollar y comparar distintos modelos de generación de trayectorias a partir de datos AIS reales con el fin de simular de manera realista el comportamiento de diferentes clases de buques en el entorno del Puerto de Barcelona y determinar qué metodología resulta más adecuada para cada tipología analizada.

Por otro lado, los objetivos específicos son los siguientes:

- Procesar y depurar datos AIS reales, seleccionando los mensajes dinámicos relevantes y aplicando criterios de filtrado espacial y temporal adecuados al área de estudio.
- Caracterizar el comportamiento espacial y dinámico de buques tipo *Cargo* y buques tipo *Fishing*, identificando patrones de navegación asociados a cada clase.
- Implementar diferentes enfoques de modelado, incluyendo métodos probabilísticos basados en transiciones espaciales y modelos generativos basados en redes neuronales, tales como cGAN y arquitecturas LSTM.
- Generar trayectorias sintéticas coherentes con los datos reales, preservando tanto la estructura geométrica de las rutas como sus características cinemáticas.
- Evaluar de forma comparativa y cuantitativa el realismo de las trayectorias generadas, contrastándolas con los datos AIS originales.
- Seleccionar la metodología más robusta y adecuada para cada tipología de buque en función de los resultados obtenidos.

2 ESTADO DEL ARTE

2.1 El sistema AIS como fuente de datos para el análisis del tráfico marítimo

2.1.1 Funcionamiento del sistema AIS

El Sistema de Identificación Automática (AIS) es un sistema de comunicación marítima basado en radiofrecuencia que utiliza enlaces VHF para el intercambio de forma automática de información entre buques y estaciones costeras. Su principal objetivo es el de mejorar la seguridad de la navegación y aumentar la conciencia situacional del navegante, permitiendo conocer en tiempo casi real la posición y el estado de las embarcaciones en un radio de, aproximadamente, 20 millas náuticas [1][2].

El funcionamiento del AIS se apoya en la transmisión de forma periódica de mensajes estructurados mediante un mecanismo de acceso múltiple por división de tiempo (TDMA). Este mecanismo permite que múltiples buques compartan los mismos canales de comunicación sin grandes interferencias, asignando intervalos temporales específicos para cada transmisión emitida [3].

Antes de nada, hay que distinguir entre los receptores AIS y los transceptores AIS, también llamados transpondedores AIS. Con un receptor se pueden ver los datos de los barcos que se encuentran cerca de nuestra embarcación, sin embargo, no se pueden enviar los datos propios del buque. Por otro lado, los transpondedores reciben los datos de las embarcaciones cercanas y, al mismo tiempo, transmiten los datos propios. La importancia de los datos varía en función del ámbito de uso del transceptor. Por lo tanto, existen diferentes tipos de transceptores, cuyo uso se divide principalmente en los ámbitos de la navegación marítima profesional y la navegación de recreo [1]:

- **Transpondedores AIS clase A:** Las embarcaciones de navegación internacional de más de 300 GT (arqueo bruto) realizando un trayecto internacional, los de más de 500GT en caso de no operar en ruta internacional, así como todos los barcos de pasajeros, están sujetos al Convenio SOLAS. Estos barcos tienen la obligación de estar equipados con transceptores AIS de clase A. Estos transceptores tienen una potencia de transmisión de hasta 12,5 W, superior a la de los dispositivos de clase B utilizados en las embarcaciones de recreo, esto se traduce en que pueden transmitir y recibir datos a mayores distancias. Así mismo, la transmisión de datos AIS también se produce con mayor frecuencia que con los transceptores de clase B. La frecuencia exacta depende de la velocidad y el estado de maniobra del barco y oscila entre cada 2 y 10 segundos mientras navega y cada 3 minutos en caso de estar fondeado[55][56].
- **Transpondedores AIS clase B:** Las embarcaciones de recreo que sólo se utilizan para el ocio o la pesca no están obligadas a estar equipadas con transceptores. Sin embargo, dado que la combinación de transmisor y receptor a bordo aporta muchas ventajas para la

seguridad, existen los llamados transceptores AIS clase B. Estos transceptores se reducen a los datos más importantes (número MMSI del barco, posición actual, rumbo y el tamaño del barco) y su potencia de transmisión es de un máximo de 2W [55][56].

Tal y como se muestra en la Figura 2-1, los mensajes AIS se clasifican, de forma general, en función del tipo de información transmitida por el buque tal:

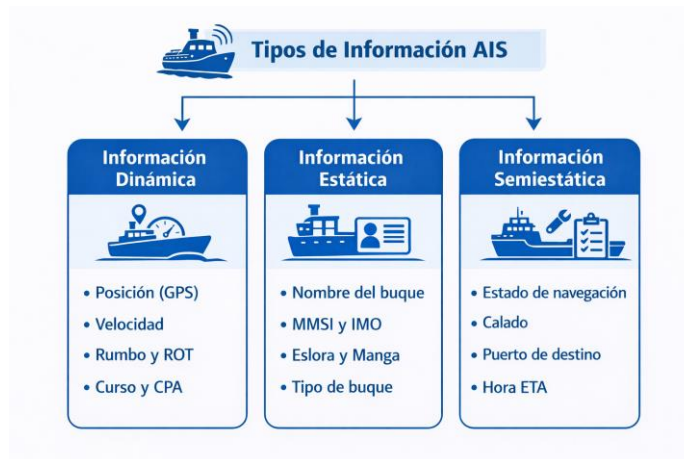


Figura 2-1 Tipos de información AIS (Fuente: propia generada con IA)

La información dinámica (mensaje tipo 1, 2 y 3) es la que está vinculada al movimiento del buque. Es especialmente relevante para este TFG, ya que incluye variables clave para la investigación como la posición geográfica, la velocidad sobre el fondo, el rumbo sobre el fondo, la tasa de giro y el estado de navegación. Estos parámetros se transmiten con una frecuencia dependiente de la velocidad y las maniobras del buque [2][3].

Por su parte, la información estática y semiestática (mensaje tipo 5) están asociadas a la identificación y características físicas del barco, así como los datos operativos que pueden variar durante la navegación. Estos incluyen el identificador MMSI (Número de Identificación del Servicio Móvil Marítimo), el tipo de buque, sus dimensiones principales, el calado estimado y el puerto de destino. Estos datos permiten clasificar los buques en distintas categorías y asociar los patrones de navegación observados a cada clase [1][2].

Como se puede observar en la Figura 2-2, que muestra un esquema simplificado del funcionamiento del sistema AIS, los buques transmiten periódicamente su información, que es recibida por otras embarcaciones y por estaciones costeras para su posterior almacenamiento.

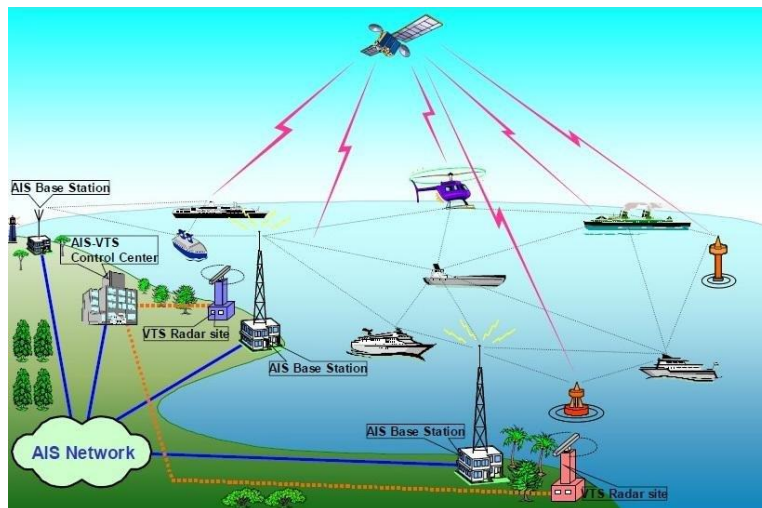


Figura 2-2 Esquema del funcionamiento del sistema AIS y transmisión de mensajes [33]

Desde el punto de vista del análisis del tráfico marítimo, esta estructura de mensajes permite reconstruir trayectorias completas y analizar maniobras e interacciones entre buques, constituyendo la base para la extracción de patrones de comportamiento.

2.1.2 Conjunto de datos AIS y áreas de estudio

El conjunto de datos AIS están formados por el total de mensajes transmitidos por los buques a través del sistema de identificación automática y almacenados por estaciones receptoras costeras o satelitales. Estos conjuntos, conocidos como *datasets*, recogen información temporalmente ordenada sobre la posición y el movimiento de las embarcaciones, así como datos de identificación y características básicas de los buques, de acuerdo con la estructura definida por la Organización Marítima Internacional (OMI)[2][3].

Desde el punto de vista de la investigación académica, los conjuntos de datos AIS se caracterizan por su naturaleza temporal y geoespacial, ya que cada registro incluye una marca de tiempo, una velocidad y una posición geográfica asociada a un buque identificado de forma única mediante su MMSI y tipo de buque. Esta estructura permite agrupar los mensajes correspondientes a una misma embarcación y reconstruir secuencias continuas de navegación a lo largo del tiempo, permitiendo trazar rutas [4][5].

Los *datasets* AIS se clasifican normalmente según su origen y cobertura espacial. Existen conjuntos de datos AIS:

- **Alcance amplio:** cubren grandes regiones marítimas, obtenidos mediante redes de estaciones receptoras distribuidas a lo largo de la costa o mediante sistemas satelitales. Estos *datasets* presentan una cobertura extensa, aunque con menor resolución temporal en comparación con los datos costeros, especialmente en zonas alejadas de tierra firme[2][3]. La Figura 2-3 muestra la distribución global de mensajes AIS registrados a escala mundial.

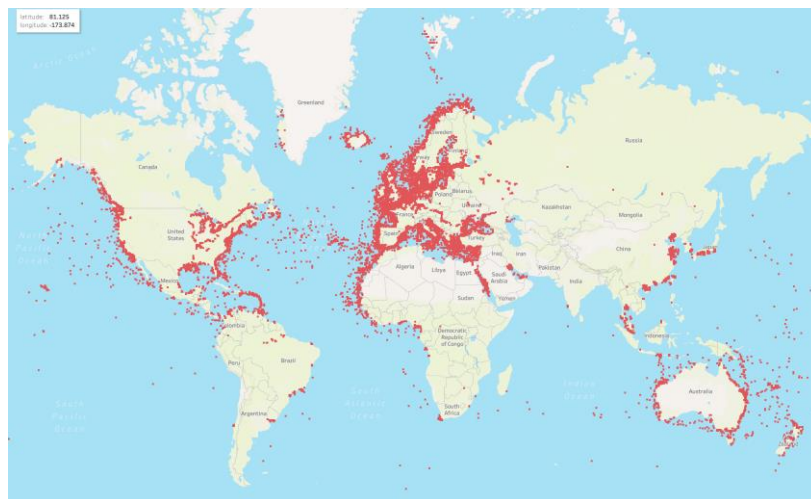


Figura 2-3 Distribución global de mensajes AIS registrados a escala mundial [34]

- **Locales:** centrados en áreas geográficas concretas como puertos, accesos portuarios o zonas costeras delimitadas. En estos casos, los datos suelen proceder de una o varias estaciones receptoras situadas estratégicamente para cubrir el área de interés. Este tipo de *datasets* presenta una mayor densidad de mensajes y una mejor resolución temporal, lo que permite capturar con mayor detalle las maniobras de los buques en espacios reducidos [4].

En cuanto al volumen de datos, los conjuntos de datos AIS se caracterizan por su gran tamaño, incluso cuando el área de estudio es limitada. La frecuencia de transmisión de los mensajes, que puede ser de pocos segundos en función de la velocidad y maniobra del buque, provoca que en periodos de

semanas se acumulen millones de registros. Estudios empíricos realizados en grandes puertos europeos, como los documentados en *datasets* académicos publicados por universidades y centros de investigación, muestran que un único puerto puede generar decenas o cientos de millones de mensajes AIS en periodos prolongados de observación [4].

2.2 Tipificación de buques y caracterización de comportamientos

La tipificación de buques constituye uno de los pilares fundamentales en el análisis del tráfico marítimo basado en datos AIS. La correcta identificación y clasificación de las embarcaciones resulta esencial para poder describir de forma correcta el tráfico observado y poder interpretar adecuadamente los patrones de navegación registrados. En entornos portuarios y costeros, donde conviven buques con características y funciones completamente diferentes, el comportamiento de navegación está estrechamente ligado al tipo de buque, a sus dimensiones y a su finalidad operativa [9].

En el sistema AIS, la clasificación de los buques se apoya principalmente en la información estática transmitida por las embarcaciones, y en particular en el campo *Ship Type* (tipo de buque). Este campo forma parte del mensaje AIS tipo 5 y sigue la codificación definida por la Organización Marítima Internacional (IMO). Tal y como se muestra en el Anexo IV: CÓDIGO NUMÉRICO DE LA TIPIFICACIÓN DE BUQUES EN LOS DATOS AIS, esta establece una serie de categorías generales para describir el tipo de buque en función de su uso principal[2][3].

Desde el punto de vista del comportamiento, cada clase de buque presenta patrones de movimiento característicos que pueden observarse de forma consistente en los datos AIS. Estas diferencias se manifiestan en variables como la velocidad habitual, la regularidad de las trayectorias, la frecuencia de maniobras, la tasa de giro o la forma en la que los buques interactúan con su entorno. Por ejemplo, los buques mercantes de gran tamaño tienden a mostrar trayectorias suaves y predecibles, mientras que embarcaciones más pequeñas o auxiliares presentan movimientos más irregulares y cambiantes.

La identificación de estas diferencias de comportamiento es especialmente relevante en estudios de tráfico portuario, debido a que la coexistencia de múltiples tipos de buques da lugar a dinámicas complejas. Una tipificación adecuada permite no solo describir el tráfico observado, sino también establecer una relación clara entre el tipo de buque y su forma de navegar, lo que resulta esencial para análisis posteriores más avanzados [9][10].

2.2.1 Buques de carga

Estos buques son denominados en AIS como “*Cargo ships*” e incluyen a los buques de carga general, portacontenedores y graneleros. Los portacontenedores transportan carga unitaria estandarizada (TEU) en contenedores, los graneleros se destinan a mercancías a granel sólido (minerales, cereales, carbón) mientras que los buques de carga general transportan mercancía diversa no contenerizada [2][8].

Los buques de carga constituyen una de las categorías más representativas del tráfico marítimo comercial. En el sistema AIS, estos buques se identifican mediante los códigos comprendidos entre 70 y 79, definidos por la Organización Marítima Internacional (OMI) [2]. Esta categoría agrupa embarcaciones de gran tamaño cuya operativa está fuertemente condicionada por su masa, dimensiones y capacidad de maniobra [8].

Desde el punto de vista del comportamiento de navegación, los buques de carga presentan trayectorias estables y bien definidas, especialmente en rutas de acceso a puerto. Estos mantienen velocidades altas y constantes durante la navegación en tránsito y reducciones progresivas de velocidad en zonas portuarias. Las maniobras suelen ser suaves, con bajas tasas de giro, debido a la inercia asociada a su tamaño.

En entornos portuarios, estos buques siguen corredores de navegación claramente establecidos y muestran una elevada repetitividad en sus trayectorias, lo que permite identificar patrones de

movimiento consistentes en el tiempo. En la Figura 2-4 podemos observar un buque portacontenedores entrando en puerto.



Figura 2-4 Buque portacontenedores entrando en puerto [37]

2.2.2 Petroleros y buques tanque

Estos buques son denominados en AIS como “Tanker ships” y se identifican mediante los códigos 80 a 89 [2]. Los buques tanque están diseñados para el transporte de cargas líquidas a granel, como petróleo crudo, productos refinados, productos químicos o gas natural licuado. Estos barcos tienen un importante peso en el comercio marítimo global, representando una porción significativa del volumen total. Su diseño incluye tanques internos especializados, sistemas de bombeo y estrictas medidas de seguridad estructural [8][9].

Existen subtipos claramente diferenciados: petroleros, quimiqueros y gaseros (Figura 2-5), cada uno con requisitos técnicos específicos. Estos buques suelen operar entre terminales especializadas y están sujetos a normativas medioambientales y de seguridad muy estrictas [9]. Como consecuencia, su comportamiento de navegación es generalmente más conservador que el de otros buques.

Respecto a su comportamiento, estos buques operan a velocidades moderadas, trayectorias muy suaves y amplios radios de giro, especialmente en zonas congestionadas. En áreas portuarias, las maniobras se realizan de forma gradual y controlada, manteniendo distancias de seguridad superiores a las de otros tipos de buque [9]. Estudios realizados por organismos europeos como EMSA (European Maritime Safety Agency) indican que los buques tanque tienden a minimizar maniobras bruscas y presentan patrones de navegación altamente predecibles, lo que los convierte en una clase claramente diferenciable dentro del tráfico marítimo.



Figura 2-5 Gasero en operaciones portuarias [36]

2.2.3 Buques de pasaje

Nombrados como “Passenger ships” en AIS, son identificados mediante los códigos de 60 a 69 e incluyen ferris y cruceros [2]. Su función principal es la del transporte de personas. A diferencia de los buques de carga, su diseño prioriza la seguridad de los pasajeros, el confort y la regularidad del servicio. Esta categoría presenta comportamientos diferenciados en función del tipo de servicio, aunque comparte una elevada regularidad operativa. El tráfico de pasajeros representa un porcentaje reducido del volumen total de tráfico marítimo, sin embargo, tiene gran relevancia económica y social, especialmente en archipiélagos y zonas costeras [9].

Los ferris suelen mostrar trayectorias altamente repetitivas, con rutas y horarios muy definidos. En los datos AIS, esto se traduce en secuencias de trayectorias prácticamente idénticas entre travesías consecutivas. Los cruceros, por su parte, presentan velocidades moderadas y maniobras cuidadosamente controladas durante las fases de entrada y salida de puerto. En la Figura 2-6 se aprecia un crucero navegando en aguas del Mar Mediterráneo.



Figura 2-6 Crucero navegando en aguas del Mar Mediterráneo [35]

2.2.4 Veleros y embarcaciones de recreo

Denominados en AIS como “Sailing vessels” y “Pleasure craft” se identifican mediante los códigos 36 y 37 respectivamente [2]. Su diseño está orientado al ocio y no al transporte comercial. Aunque tradicionalmente no todos estaban equipados con AIS, su uso ha aumentado notablemente en los últimos años, especialmente en Europa [1][9]. En la Figura 2-7 podemos ver un velero navegando a vela.



Figura 2-7 Velero navegando a vela [31]

Desde el punto de vista del comportamiento, estos buques presentan trayectorias irregulares, velocidades generalmente bajas y una elevada variabilidad en rumbo. En los datos AIS es habitual observar cambios frecuentes de dirección, especialmente en zonas de navegación recreativa o fondeo. Además, su comportamiento está influido por factores externos como el viento y las condiciones meteorológicas. Además, en el caso de los veleros, deben adaptar su rumbo a la dirección del viento, lo que conlleva cambios muy grandes de rumbo de forma frecuente.

2.2.5 Buques pesqueros

Estos buques son llamados “*Fishing vessels*” en AIS, y representan el código AIS 30 [2]. Los buques pesqueros constituyen una categoría claramente diferenciada del resto de buques AIS. Su función principal es la captura de recursos pesqueros, lo que se traduce en una gran diversidad de tamaños, diseños y métodos de operación. Dentro de esta categoría se incluyen desde pequeñas embarcaciones artesanales hasta grandes arrastreros y buques factoría [9].

Desde el punto de vista estructural, los pesqueros suelen presentar dimensiones menores que los buques de carga o tanque, aunque con una gran variabilidad en función del tipo de pesca (arrastre, cerco, palangre, etc.). Su diseño prioriza la maniobrabilidad, la capacidad de trabajo en cubierta y el almacenamiento de capturas, más que la eficiencia en navegación a larga distancia.

Aunque el uso de AIS no ha sido históricamente obligatorio para todos los buques pesqueros, en los últimos años su implantación ha aumentado de forma significativa, especialmente en aguas europeas, como consecuencia de regulaciones de seguridad y control pesquero impulsadas por organismos como la Unión Europea y la Organización Marítima Internacional.

Los buques pesqueros presentan un comportamiento en la mar, claramente distinto al de los buques comerciales, ya que su navegación está condicionada por la actividad de pesca. Durante la faena, muestran velocidades bajas, trayectorias irregulares y cambios frecuentes de rumbo, permaneciendo durante largos periodos en zonas concretas del mar. En cambio, en los desplazamientos entre puerto y caladeros adoptan trayectorias más directas y velocidades moderadas, aunque generalmente inferiores a las de los buques mercantes, un patrón ampliamente documentado en estudios europeos basados en datos AIS. Se muestra un ejemplo en la Figura 2-8.



Figura 2-8 Pesquero faenando en las Rías Baixas (Fuente: propia generada con IA)

2.2.6 Embarcaciones de servicio y apoyo portuario

Las embarcaciones de servicio y apoyo portuario incluyen remolcadores, barcos de prácticos y otras embarcaciones auxiliares. En AIS, estas embarcaciones se identifican con el nombre “Tug” o “Port services” habitualmente mediante códigos entre 50 y 59 [2][3]. En la Figura 2-9 se muestra un ejemplo representativo de una embarcación de prácticos en operación.

Estas embarcaciones realizan tareas especializadas que resultan esenciales para el funcionamiento seguro y eficiente de los puertos, como la asistencia en maniobras de atraque y desatraque, el transporte

de prácticos, el remolque de buques, el suministro de servicios logísticos, la señalización marítima y, en algunos casos, labores de mantenimiento, salvamento o control ambiental. Su operativa se concentra principalmente en áreas portuarias y zonas próximas a la costa, donde actúan como elementos de apoyo indispensables para garantizar la seguridad de la navegación y la correcta gestión del tráfico marítimo.

Su comportamiento de navegación es claramente distinto al de los buques comerciales. Estos buques son de tamaño reducido y alta maniobrabilidad, realizando normalmente trayectorias cortas, cambios frecuentes de rumbo y con velocidades muy variables. Estas embarcaciones tienen una interacción constante con otros buques en distintas maniobras de entrada y salida de puerto.



Figura 2-9 Embarcación de práctico dirigiéndose a un buque (Fuente: propia generada con IA)

2.3 Representación espacial del tráfico marítimo

La representación espacial del tráfico marítimo es un aspecto fundamental en el análisis de datos AIS, ya que los registros de posición de los buques se encuentran definidos en un espacio geográfico continuo. Para poder analizar, comparar y modelar estos datos de forma sistemática, resulta necesario transformar el espacio marítimo en una estructura que facilite su tratamiento computacional. En la literatura científica, este proceso se realiza mediante técnicas de discretización espacial, que permiten dividir el área de estudio en regiones manejables, visuales y coherentes.

La elección del método de discretización espacial, influye directamente en la calidad del análisis posterior, ya que condiciona la forma en la que se agregan las trayectorias, se detectan patrones de navegación y se describen las interacciones entre los buques estudiados. Por este motivo, numerosos trabajos sobre tráfico marítimo dedican una atención específica a la definición del espacio de estudio, especialmente en entornos portuarios y costeros, donde la densidad de tráfico es elevada.

2.3.1 Técnicas de discretización espacial

Las técnicas de discretización espacial tienen como objetivo dividir el espacio marítimo continuo en un conjunto finito de celdas o regiones, de forma que los datos AIS puedan analizarse de manera estructurada.

Una de las aproximaciones más sencillas consiste en utilizar rejillas cartesianas regulares, donde el área de estudio se divide en celdas de tamaño fijo definidas por intervalos de latitud y longitud. Este enfoque presenta la ventaja de su simplicidad y facilidad de implementación.

En la Figura 2-10, que ilustra un ejemplo de discretización mediante una rejilla regular.

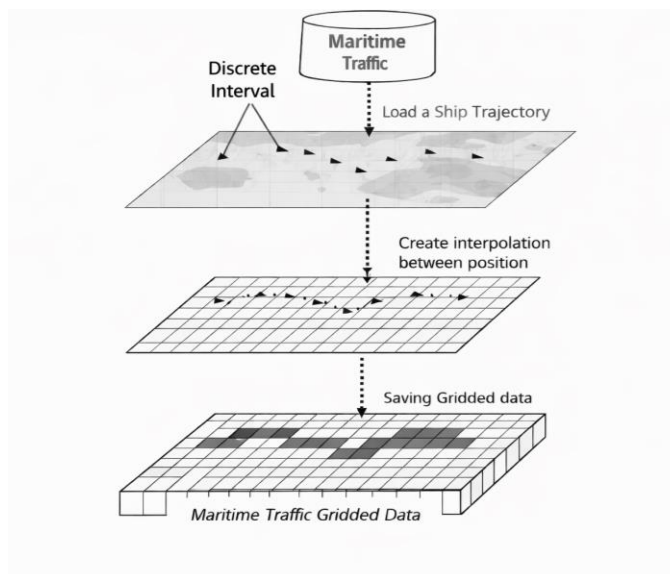


Figura 2-10 Ejemplo conceptual de discretización espacial mediante rejilla regular aplicada a trayectorias de buques (Fuente: propia generada con IA)

2.3.2 Cuadrículas tipo UBER aplicadas al tráfico marítimo

En los últimos años, se ha extendido el uso de cuadrículas espaciales jerárquicas inspiradas en sistemas desarrollados inicialmente para aplicaciones de movilidad urbana, como el método empleado por la plataforma digital UBER. Este enfoque se basa en la división del espacio en celdas hexagonales de tamaño uniforme y organizadas de forma jerárquica, lo que permite representar el espacio de manera eficiente y escalable. La idea fundamental del sistema UBER (H3), como se mencionó, es dividir la superficie de la Tierra en celdas hexagonales, organizadas en distintos niveles de resolución, de forma que cualquier punto geográfico pueda asociarse de manera única a una celda concreta. Cada celda hexagonal puede subdividirse en varias celdas de resolución superior, es decir, en hexágonos más pequeños, manteniendo siempre la estructura hexagonal [12].

Como se puede apreciar en la Figura 2-11, que muestra un ejemplo de discretización mediante una cuadrícula hexagonal, este tipo de representación resulta especialmente adecuada para zonas con alta densidad de tráfico.

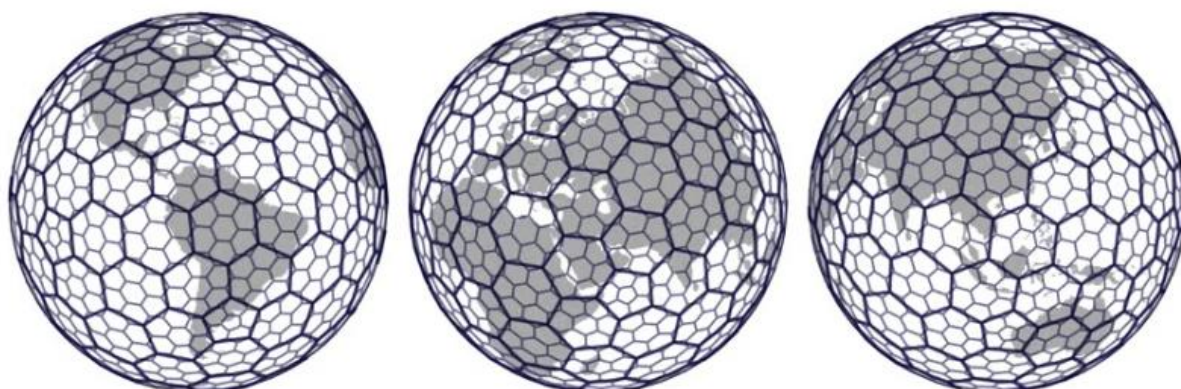


Figura 2-11 Ejemplo de discretización espacial mediante cuadrícula hexagonal jerárquica tipo H3 [12]

La asignación de cada posición AIS a una celda hexagonal facilita la conversión de trayectorias continuas en secuencias discretas, esto significa que cada posición deja de representarse solamente por sus coordenadas y pasa a asociarse a una celda en concreto de la rejilla, lo que puede mejorar la eficiencia

de los modelos de aprendizaje automático al reducir el número de opciones posibles de transición u ocupación espacial.

2.3.3 Modelos probabilísticos de transición espacial: cadenas de Markov

Las cadenas de Markov constituyen un modelo probabilístico utilizado para describir la evolución de un sistema a través de una secuencia de estados discretos. Su característica principal es que la probabilidad de que el sistema pase a un estado futuro depende únicamente del estado en el que se encuentra en el momento actual, y no del conjunto completo de estados anteriores. Esta característica se conoce como propiedad de Markov [68].

Desde un punto de vista conceptual, esto implica que el sistema no “recuerda” el camino completo que ha seguido hasta llegar a su estado actual, sino que su comportamiento futuro está condicionado exclusivamente por su situación presente. Esta simplificación permite modelizar procesos dinámicos de forma más manejable desde el punto de vista matemático y computacional.

Las transiciones entre estados se representan mediante una matriz de transición, que recoge las probabilidades de pasar de un estado a otro en un único paso del proceso. En esta matriz, cada fila corresponde al estado actual del sistema y cada columna a uno de los posibles estados siguientes. El valor de cada elemento indica la probabilidad de que, estando el sistema en un determinado estado, el siguiente estado sea uno concreto. De esta forma, la fila asociada a un estado describe todas las posibles transiciones que pueden producirse a partir de él.

Una propiedad fundamental de la matriz de transición es que la suma de los valores de cada fila es igual a 1. Esto garantiza que se están considerando todas las alternativas posibles desde ese estado, es decir, que el modelo contempla el conjunto completo de escenarios que pueden ocurrir en el siguiente paso. Esta condición asegura la coherencia probabilística del sistema y permite describir su evolución de manera consistente [69]. Esto se observa en la Figura 2-12, donde se muestra una cadena simple de Markov.

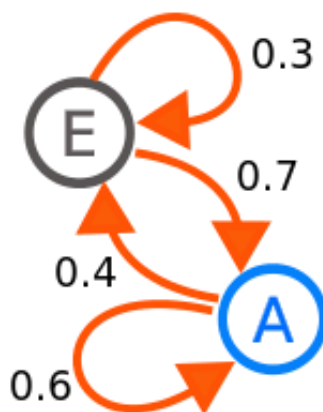


Figura 2-12 Cadena simple de Markov [69]

Este tipo de modelo resulta especialmente útil cuando el comportamiento de un sistema puede interpretarse como una sucesión de cambios entre situaciones bien definidas. En el ámbito espacial, cada estado puede asociarse a una región concreta del entorno. Bajo esta interpretación, una cadena de Markov permite modelar probabilísticamente el paso de una región a otra, describiendo el movimiento del sistema como una secuencia de transiciones entre estados discretos.

2.4 Inteligencia Artificial y Machine Learning

2.4.1 Inteligencia Artificial y Aprendizaje Automático: conceptos generales

La Inteligencia Artificial (IA) es un campo de la ingeniería y de las ciencias computacionales cuyo objetivo es desarrollar sistemas capaces de realizar tareas que, tradicionalmente, requerían de la inteligencia humana. Estas tareas incluyen, entre otras, el razonamiento y aprendizaje a partir de la experiencia, el reconocimiento de patrones y la toma de decisiones en entornos complejos, facilitando los procesos a los usuarios. A lo largo de las últimas décadas, la IA ha evolucionado desde enfoques basados en reglas explícitas básicas hacia métodos más flexibles apoyados en el análisis de datos[13][16].

La IA comenzó a tomar forma como disciplina científica a mediados del siglo XX, cuando investigadores como Alan Turing plantearon, por primera vez, la posibilidad de que las máquinas pudieran imitar ciertos procesos del razonamiento de los humanos. Durante las décadas de los 50 y 60 surgen los primeros programas capaces de resolver problemas lógicos y matemáticos, lo que generó grandes expectativas sobre su potencial. Sin embargo, las limitaciones tecnológicas y computacionales de la época provocaron periodos de estancamiento conocidos como “inviernos de la IA”. A partir de los años 90 y, especialmente, en el siglo XXI, el aumento de la capacidad de cálculo, la disponibilidad masiva de datos y el desarrollo de nuevos algoritmos impulsaron un resurgimiento de la inteligencia artificial. Esto ha dado lugar a avances significativos en áreas como el aprendizaje automático, el reconocimiento de patrones y el procesamiento del lenguaje natural. En la actualidad, la IA se ha consolidado como una tecnología clave en numerosos ámbitos científicos, industriales y sociales, con aplicaciones que van desde la automatización de procesos hasta el apoyo a la toma de decisiones complejas [52]. En la Figura 2-13 se ve la evolución histórica de la Inteligencia Artificial.



Figura 2-13 Evolución histórica de la IA (Fuente: propia generada con IA)

Dentro del amplio ámbito de la Inteligencia Artificial, el Aprendizaje Automático (*Machine Learning*) constituye uno de los enfoques más relevantes en la actualidad. El *Machine Learning* se basa en el desarrollo de algoritmos capaces de aprender automáticamente a partir de datos, sin necesidad de que el comportamiento del sistema sea programado de forma explícita para cada situación requerida. En lugar de definir reglas fijas, estos algoritmos construyen modelos que capturan las relaciones presentes en los datos de entrada[14].

Desde un punto de vista conceptual, el Aprendizaje Automático se apoya en la idea de que muchos problemas pueden abordarse mediante la identificación de patrones en grandes volúmenes de datos. A partir de ejemplos históricos, el sistema va ajustando sus parámetros internos con el objetivo de

representar de la mejor manera posible el comportamiento que ha observado. Este enfoque resulta especialmente útil en escenarios donde las relaciones entre variables son complejas, no lineales o difíciles de describir mediante modelos analíticos clásicos[14][15].

En el contexto de la ingeniería, el interés por la Inteligencia Artificial y el Aprendizaje Automático ha crecido de forma exponencial debido a la disponibilidad de grandes volúmenes de datos y al aumento de la capacidad de cálculo. Frente a los métodos tradicionales, basados en modelos deterministas y en reglas predefinidas, las técnicas de aprendizaje permiten abordar problemas de mayor complejidad y adaptarse mejor a la variabilidad inevitable de todos los sistemas reales [13][16].

Es por este motivo que la Inteligencia Artificial y el Aprendizaje Automático se han consolidado como un marco de referencia en la ingeniería moderna. Su aplicación resulta especialmente útil en el análisis de datos, tanto complejos como dinámicos, donde el comportamiento del sistema no puede describirse fácilmente mediante modelos simples, sentando así las bases para enfoques más avanzados que se abordarán en los apartados posteriores [15][16].

2.4.2 Aplicación del Machine Learning al análisis de datos

Como se expuso en el apartado anterior, el Aprendizaje Automático (*Machine Learning*) se ha consolidado como una herramienta fundamental para el análisis de datos en aquellos ámbitos donde el volumen, la complejidad y la variabilidad de la información superan las capacidades que se pueden llevar a cabo con los métodos tradicionales. En lugar de basarse exclusivamente en modelos analíticos, el *Machine Learning* permite extraer conocimiento directamente a partir de los datos, identificando patrones, regularidades y relaciones subyacentes de forma automática. Esto le otorga al usuario una gran capacidad de análisis.

En el análisis de datos, las técnicas de *Machine Learning* se emplean habitualmente con el objetivo de abordar tareas como la clasificación, la agrupación, la detección de patrones y la modelización de comportamientos [14]. Estos enfoques resultan especialmente adecuados cuando los datos presentan ruido, relaciones no lineales o una elevada dimensionalidad, características que son habituales en muchos problemas de ingeniería actuales.

Sin embargo, de entre todas las ventajas que ofrece, uno de los aspectos clave del *Machine Learning* aplicado al análisis de datos es su capacidad para adaptarse a grandes volúmenes de información heterogénea. A partir de conjuntos de datos históricos que se le otorgan, los algoritmos aprenden representaciones internas que capturan las características más relevantes del sistema analizado. Este enfoque permite analizar de forma conjunta múltiples variables de entrada, capturando mejor la complejidad del sistema estudiado [15].

Así mismo, otra de las ventajas del *Machine Learning* es su capacidad para generalizar a partir de los datos disponibles. Una vez entrenado, un modelo puede aplicarse a nuevos conjuntos de datos con características similares al estudiado. Esto proporcionará resultados coherentes sin necesidad de redefinir el modelo desde cero.

En aplicaciones de ingeniería, el uso del *Machine Learning* en el análisis de datos no tiene por qué sustituir a los métodos clásicos, sino que lo que hace es complementarlos. Mientras que los métodos tradicionales permiten describir y entender fenómenos bien conocidos, las técnicas de aprendizaje mencionadas aportan mayores capacidades para analizar sistemas complejos y dinámicos, ampliando las posibilidades de análisis y proporcionando una base sólida para estudios más avanzados [11][15].

De forma general, las técnicas de *Machine Learning* se agrupan en cuatro grandes categorías que se desarrollarán a continuación.

2.4.2.1 Aprendizaje supervisado

El aprendizaje supervisado, dentro del *Machine Learning* se caracteriza por el uso de datos etiquetados durante la fase de entrenamiento del modelo [14]. En este enfoque de datos etiquetados, el

algoritmo aprende a partir de un conjunto de ejemplos en los que tanto los datos de entrada, como los de salida esperada son conocidos. Esto permite establecer una relación entre ambos mediante un proceso de ajuste progresivo del modelo.

Este tipo de aprendizaje permite a los algoritmos identificar patrones y realizar predicciones fiables cuando se enfrentan a nuevos datos no observados previamente, siempre que estos mantengan características similares a las del conjunto de entrenamiento, ya que si las características fuesen diferentes los patrones aprendidos podrían no ser representativos y afectar al rendimiento del modelo. El proceso de aprendizaje se basa en la evaluación continua del rendimiento del modelo, lo que facilita su optimización y validación.

El aprendizaje supervisado puede compararse con el proceso mediante el cual un niño aprende a reconocer tipos de fruta (Figura 2-14) con la ayuda de un profesor. Durante la fase de aprendizaje, el profesor muestra diferentes frutas y, para cada una de ellas, indica explícitamente su nombre: manzana, plátano o naranja. De este modo, el alumno no solo observa características como el color, la forma o el tamaño, sino que además conoce la respuesta correcta asociada a cada ejemplo. Tras ver suficientes ejemplos etiquetados, el alumno es capaz de identificar una fruta nueva que no había visto anteriormente, esto basándose en las similitudes con los ejemplos aprendidos. Si se le presenta una fruta roja y redondeada, podrá clasificarla como una manzana con un alto grado de acierto.

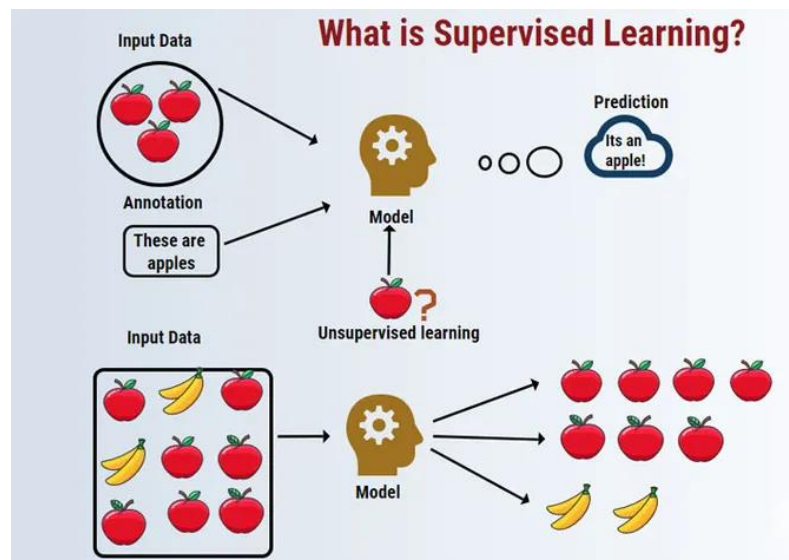


Figura 2-14 Ejemplo de aprendizaje supervisado [46]

Dentro del aprendizaje supervisado encontramos dos tipos principales:

- **Clasificación:** La clasificación tiene como objetivo asignar cada dato de entrada a una categoría o clase discreta previamente definida. El modelo aprende a partir de ejemplos etiquetados y, una vez entrenado, decide a qué clase pertenece una nueva observación. Estas clases pueden ser binarias, cuando solo existen dos posibles resultados (por ejemplo, “sí/no” o “normal/anómalo”), o multiclase, cuando hay más de dos categorías posibles. La clasificación es especialmente útil en problemas de identificación, reconocimiento y detección, donde interesa determinar a qué grupo pertenece un elemento concreto.
- **Regresión:** La regresión, a diferencia de la clasificación, se utiliza cuando el resultado que se desea predecir es un valor continuo dentro de un rango numérico. En este caso, el modelo aprende la relación existente entre las variables de entrada y una variable de salida real, con el objetivo de estimar su valor lo más fielmente posible. La calidad del modelo no se mide por aciertos o errores discretos, sino por la desviación entre el valor predicho y el valor real.

La regresión es adecuada para problemas de predicción y estimación, como el cálculo de tendencias, la previsión de magnitudes o la estimación de parámetros físicos.

Un ejemplo real y común del uso de aprendizaje supervisado, tal y como se muestra en la Figura 2-15 es el CAPTCHA [45]. Estos aparecen en algunas páginas web antes de ingresar en ellas para comprobar que el usuario no es un robot. Sin embargo, a parte de esta comprobación, el usuario está etiquetando fotos que posteriormente serán empleadas en [45], por ejemplo, vehículos autónomos.

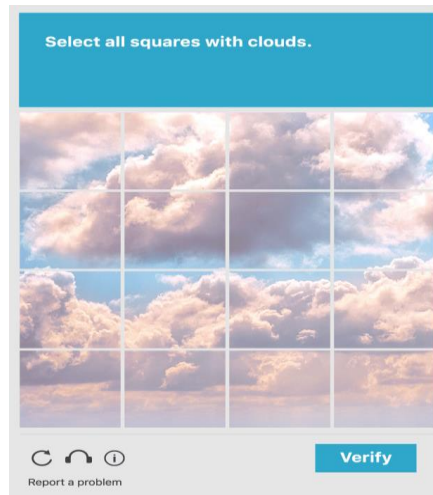


Figura 2-15 CAPTCHA [45]

2.4.2.2 Aprendizaje no supervisado

El aprendizaje no supervisado [47] es una rama del aprendizaje automático que se caracteriza por trabajar con datos no etiquetados, es decir, conjuntos de datos en los que no se conoce previamente el resultado o la categoría asociada a cada observación. A diferencia del aprendizaje supervisado, en este enfoque el algoritmo no dispone de respuestas correctas durante el entrenamiento, sino que debe identificar por sí mismo patrones, estructuras o relaciones internas presentes en los datos.

Este aprendizaje resulta especialmente útil cuando no se dispone de información previa sobre la clasificación de los datos o cuando se desea explorar grandes volúmenes de información para descubrir regularidades ocultas. Este tipo de aprendizaje permite agrupar datos similares, detectar anomalías o reducir la dimensionalidad de los conjuntos de datos, facilitando su análisis posterior y la extracción de conocimiento relevante.

El proceso de entrenamiento del algoritmo se basa en el análisis de las características y regularidades identificadas en el conjunto de datos inicial, dando lugar a un modelo una vez que sus parámetros han sido correctamente ajustados. Posteriormente, cuando se incorpora un nuevo texto, documento o imagen, el modelo extrae sus rasgos más relevantes y los compara con los patrones previamente aprendidos, asignando dicho elemento al grupo que presenta un mayor grado de similitud. En la Figura 2-16 se puede ver un ejemplo de aprendizaje no supervisado en comparación con el aprendizaje supervisado.

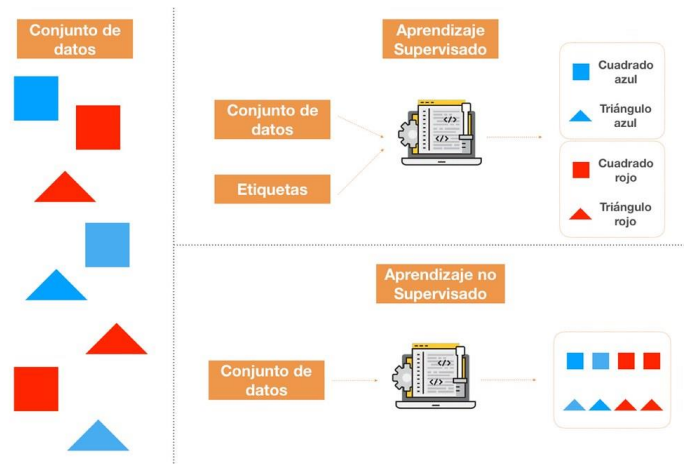


Figura 2-16 Comparación de aprendizaje no supervisado con aprendizaje supervisado [48]

Uno de los usos más habituales del aprendizaje no supervisado es el *clustering*, donde el algoritmo organiza automáticamente los datos en grupos en función de su similitud, sin intervención humana. Este proceso permite identificar comportamientos comunes, segmentar conjuntos de datos complejos o descubrir patrones que no resultarían evidentes mediante métodos tradicionales de análisis (Figura 2-16)[47].

Por otro lado, el aprendizaje no supervisado se emplea en tareas de detección de anomalías. En este caso el objetivo de este es identificar observaciones que se desvían significativamente del comportamiento general del conjunto de datos. Así mismo, se emplea en técnicas de reducción de dimensionalidad, orientadas a simplificar la representación de la información manteniendo sus características más relevantes [47]. Estas mencionadas capacidades lo convierten en una herramienta especialmente valiosa en escenarios donde se manejan grandes volúmenes de datos heterogéneos.

2.4.2.3 Aprendizaje semisupervisado

El aprendizaje semisupervisado es un enfoque del aprendizaje automático que se sitúa entre los mencionados aprendizaje supervisado y no supervisado. Este aprendizaje se caracteriza por la combinación del uso de un conjunto reducido de datos etiquetados con una cantidad considerable de datos sin etiquetar[14].

Este concepto nace de una situación muy común en problemas reales: mientras que la obtención de grandes volúmenes de datos suele ser sencilla, el proceso de etiquetado de estos requiere tiempo, recursos y conocimiento. Esto genera una limitación de disponibilidad de datos completamente etiquetados.

El funcionamiento del aprendizaje semisupervisado empieza generalmente con el entrenamiento de un modelo inicial utilizando únicamente los datos etiquetados que tenga disponibles. A partir de este primer modelo generado, se analizan los datos no etiquetados para identificar patrones y estructuras internas y en función de su similitud con los datos conocidos se le asigna una etiqueta. Estas nuevas muestras, unidas con las originales, permiten ampliar el conjunto de entrenamiento y refinar progresivamente el modelo, mejorando su capacidad de generalización sin necesidad de un etiquetado manual exhaustivo[49]. En la Figura 2-17 se muestra un esquema sencillo del proceso de aprendizaje. En esta se aprecia como se disponen de dos *datasets* distintos, uno con datos etiquetados y otro sin etiquetar. En un primer momento, los datos etiquetados son empleados para entrenar al modelo mediante técnicas de aprendizaje supervisado. Posteriormente, el modelo utilizará lo aprendido para predecir la etiqueta de los datos que no están etiquetados del segundo *dataset*. A partir de ahí, se emplearán ambos, *dataset* etiquetados conocidos y generados para volver a entrenar a un nuevo modelo.

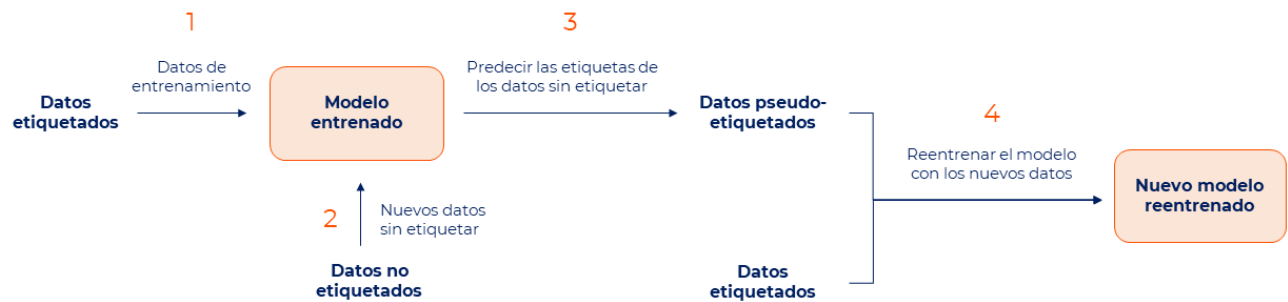


Figura 2-17 Esquema sencillo del aprendizaje semisupervisado[50]

Este enfoque se emplea con frecuencia en aplicaciones que manejan grandes volúmenes de datos, como el análisis de imágenes, texto, vídeo o audio, donde los datos sin etiquetar a menudo son abundantes [49].

2.4.2.4 Aprendizaje por refuerzo

El aprendizaje por refuerzo es una técnica de aprendizaje automático en la que un agente aprende a tomar decisiones mediante la interacción directa con un entorno. A diferencia del aprendizaje supervisado, no se dispone de un conjunto de datos etiquetados ni de respuestas conocidas, sino que el agente aprende a partir de la experiencia. Esto, probando distintas acciones y observando las consecuencias que tienen las mismas. El objetivo principal es encontrar una estrategia de actuación que maximice una recompensa acumulada a lo largo del tiempo [51].

En este tipo de aprendizaje, el proceso se basa en cuatro elementos fundamentales:

- Agente: el que toma las decisiones.
- Entorno: con el que interactúa el agente.
- Acción: las posibles decisiones que puede ejecutar el agente.
- Estado: situación en la que se encuentra el entorno cada vez que recibe una acción del agente.
- Recompensa: señal numérica que indica si la acción tomada por el agente ha sido beneficiosa o perjudicial.

Tras cada acción, el entorno cambia de estado y devuelve una recompensa, lo que permite al agente evaluar la calidad de sus decisiones y ajustar su comportamiento de forma progresiva [51].

Un ejemplo de aprendizaje por refuerzo en la vida real puede compararse con el entrenamiento de un militar durante su instrucción. En este caso, el militar actúa como el agente, mientras que el entorno está formado por el conjunto de ejercicios, normas y situaciones a las que se enfrenta. Cada vez que el militar realiza una acción (ejecutar correctamente un ejercicio o mantener la disciplina) recibe una recompensa, que puede materializarse en reconocimiento o condecoraciones. Por el contrario, cuando su actuación no es adecuada, recibe una penalización, como puede ser un correctivo de 25 flexiones. A lo largo del tiempo, el militar aprende qué comportamientos generan recompensas y cuáles conllevan consecuencias negativas. De esta manera, ajustará progresivamente su forma de actuar para maximizar los resultados positivos y minimizar las negativas. De manera análoga, en el aprendizaje por refuerzo un algoritmo aprende mediante prueba y error, reforzando aquellas decisiones que producen mejores resultados a largo plazo.

El entrenamiento de un algoritmo de aprendizaje por refuerzo se produce de manera iterativa. Inicialmente, el agente explora el entorno realizando acciones de forma casi aleatoria. A medida que acumula experiencia, comienza a identificar qué acciones generan mejores recompensas en determinados estados, reforzando esas decisiones y descartando las menos favorables. En la Figura 2-18 se puede observar un esquema de este aprendizaje.

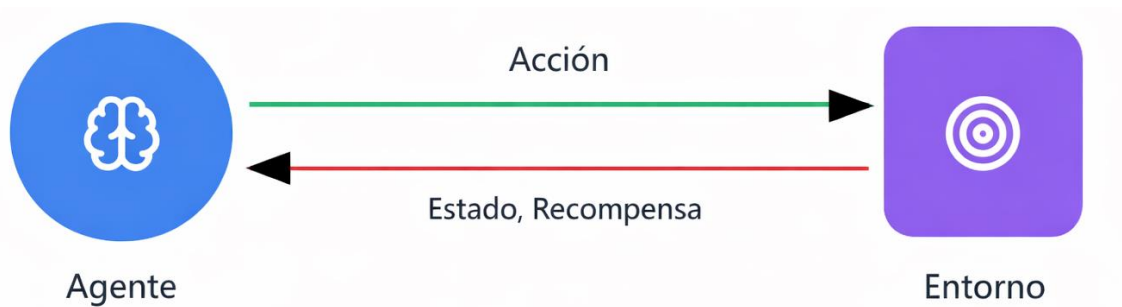


Figura 2-18 Aprendizaje por refuerzo [51]

Este tipo de aprendizaje se emplea en ámbitos donde es necesario aprender estrategias óptimas de control o actuación, como la robótica, los videojuegos o la automatización y control de sistemas, ya que permite adaptarse a entornos dinámicos y complejos [14][51].

2.4.3 Algoritmo de *K-nearest neighbors* (KNN)

El algoritmo *K-Nearest Neighbors* (KNN) es una técnica de Aprendizaje Automático supervisado que se emplea tanto en problemas de clasificación como de regresión. Su principio fundamental consiste en asumir que una muestra nueva puede describirse a partir de las muestras más similares previamente observadas en el conjunto de datos que se emplearon para el entrenamiento [19].

Desde un punto de vista técnico, cada muestra se representa como un vector dentro de un espacio de características, donde cada dimensión corresponde a una variable del problema. Cuando se desea clasificar o estimar el valor asociado a una nueva muestra, el algoritmo calcula la distancia entre esta y todas las muestras del conjunto de entrenamiento. La métrica más empleada es la distancia euclídea, aunque pueden utilizarse otras métricas dependiendo de la naturaleza de los datos analizados [20].

Una vez calculadas las distancias entre la nueva muestra y los datos conocidos, el algoritmo selecciona los K puntos más cercanos. El valor de K es un parámetro importante, ya que influye directamente en el resultado del modelo. Si se elige un valor pequeño de K , el algoritmo se centra mucho en los datos más próximos, lo que puede hacerlo sensible a valores atípicos o ruido. Por el contrario, valores más grandes de K permiten obtener decisiones más estables, aunque pueden suavizar en exceso los patrones presentes en los datos [11].

En problemas de clasificación, KNN asigna a la nueva muestra la clase que aparece con mayor frecuencia entre sus K vecinos más cercanos. En el caso de regresión, el valor final se obtiene calculando la media, o una media ponderada, de los valores asociados a dichos vecinos. De este modo, el algoritmo utiliza directamente la información de los datos más similares para realizar la predicción, sin necesidad de construir un modelo matemático complejo [20].

En la Figura 2-19 se muestra un ejemplo ilustrativo del funcionamiento del algoritmo, donde el dato a clasificar se representa mediante un círculo rojo. En este escenario se consideran dos posibles etiquetas, identificadas por los colores naranja y verde, y se ha fijado el parámetro $K=3$. Los ejemplos utilizados durante el entrenamiento aparecen como círculos, cuyo color indica la etiqueta a la que pertenecen. En este caso, los tres vecinos más próximos al punto de entrada son los que quedan contenidos dentro de la circunferencia dibujada. Dado que la mayoría de estos vecinos pertenecen a la clase verde, el algoritmo asigna finalmente dicha etiqueta (verde) al dato analizado.

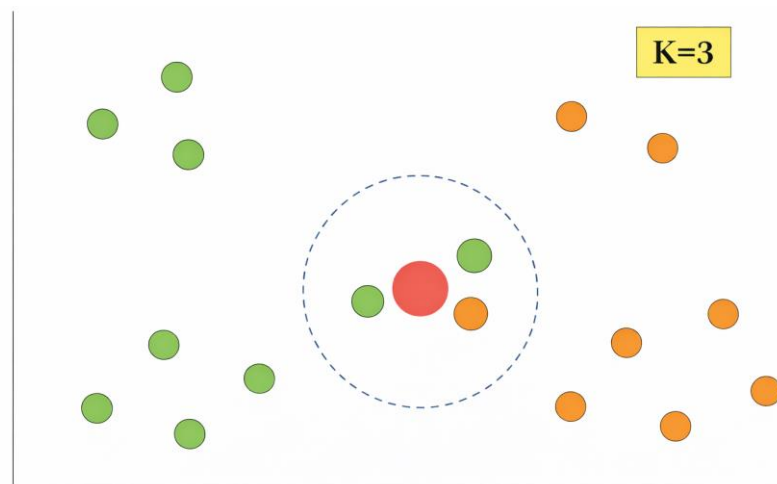


Figura 2-19 Ejemplo de KNN para K=3 (Fuente: propia)

Un problema de regresión es aquel en el que el objetivo del modelo es predecir un valor numérico continuo a partir de un conjunto de variables de entrada. En este tipo de problemas, la salida no pertenece a una clase concreta, sino que puede tomar cualquier valor dentro de un rango determinado. Ejemplos habituales de regresión incluyen la estimación de la velocidad de un buque, la predicción de su posición futura o el cálculo del tiempo estimado de llegada a un puerto.

Un problema de clasificación es aquel en el que el objetivo del modelo es asignar una muestra de entrada a una clase o categoría discreta previamente definida. A partir de un conjunto de variables de entrada, el modelo aprende a diferenciar entre distintas clases utilizando datos etiquetados durante el proceso de entrenamiento. La salida del modelo no es un valor numérico continuo, sino una etiqueta que identifica la categoría a la que pertenece la muestra analizada. Ejemplos habituales de problemas de clasificación incluyen la identificación del tipo de buque, la detección del estado de navegación o la asignación de trayectorias a diferentes patrones de comportamiento, a partir de datos históricos.

2.5 Redes neuronales

2.5.1 Redes neuronales artificiales: fundamentos

Las redes neuronales artificiales son una de las técnicas más utilizadas dentro del Aprendizaje Automático (*Machine learning*) con el objetivo de modelar sistemas complejos a partir de datos. Su desarrollo se inspira en una idea sencilla: imitar, de forma muy simplificada, el funcionamiento del cerebro humano, donde un gran número de neuronas se conectan entre sí para procesar información[14][15].

Desde un punto de vista práctico, una red neuronal es un modelo matemático diseñado para aprender relaciones entre datos de entrada y datos de salida. A diferencia de los modelos tradicionales, en los que las reglas del sistema se definen explícitamente, las redes neuronales aprenden estas relaciones automáticamente a partir de ejemplos, ajustando su comportamiento en función de los datos disponibles [15] [14].

Este tipo de modelos resulta especialmente útil cuando el problema a resolver presenta relaciones complejas o no lineales, es decir, cuando no es posible describir fácilmente el comportamiento del sistema mediante ecuaciones simples [15]. Por este motivo, las redes neuronales se han convertido en una herramienta habitual en numerosos ámbitos de la ingeniería.

De forma general, la red neuronal se compone de:

- **Capa de entrada:** es la que se encarga de recibir las variables iniciales del problema.
- **Capas ocultas:** donde se produce el procesamiento de la información.

➤ **Capa de salida:** proporciona el resultado del modelo

Como se puede observar en la Figura 2-20, la estructura en capas permite que la red transforme los datos de manera gradual, lo que facilita la representación de relaciones complejas.

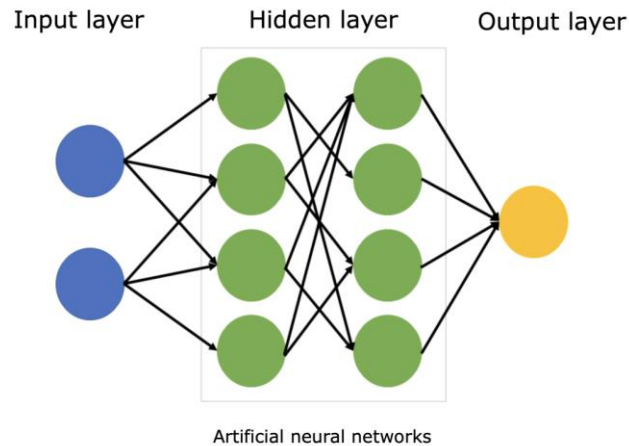


Figura 2-20 Estructura básica de una red neuronal artificial [38]

La unidad básica de una red neuronal artificial es la neurona artificial. Su funcionamiento es sencillo y puede entenderse en tres pasos:

- 1) La neurona recibe varias entradas, que representan las variables del problema que se desea modelar. Estas entradas son valores numéricos que contienen la información necesaria para describir el estado del sistema analizado.
- 2) Cada una de las entradas se multiplica por un peso asociado, que indica la importancia relativa de dicha entrada en el cálculo final. A continuación, todas las contribuciones ponderadas se combinan mediante una suma.
- 3) El resultado de la combinación ponderada se introduce en una función de activación, que genera la salida final de la neurona.

La función de activación introduce no linealidad en el modelo, es decir, permite que la neurona no se limite solamente a realizar simples sumas y multiplicaciones. Gracias a ella, la red puede aprender relaciones más complejas entre los datos de entrada y la salida. Sin esta función, la red solo podría representar relaciones muy simples, lo que reduciría considerablemente su capacidad para modelar el comportamiento real del sistema [15] [14].

En la Figura 2-21 se muestra un esquema típico de una neurona artificial, donde se aprecia claramente cómo se combinan las entradas, los pesos y la función de activación

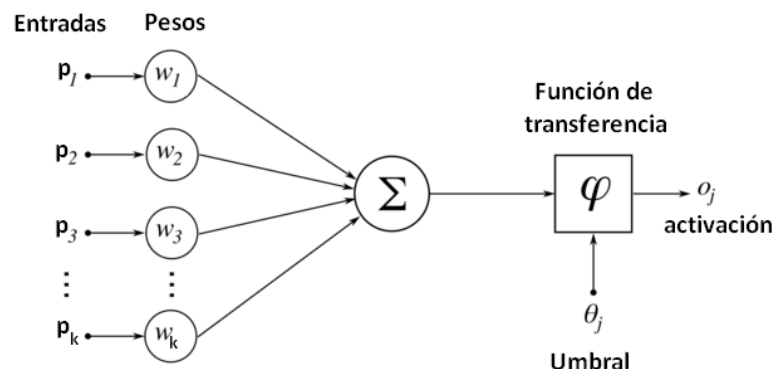


Figura 2-21 Esquema del funcionamiento de una neurona artificial [39]

El aprendizaje de una red neuronal consiste en ajustar los pesos de las conexiones entre neuronas para que la salida del modelo se aproxime a los valores esperados. Este proceso se realiza de forma iterativa, utilizando un conjunto de datos de entrenamiento [11].

En cada iteración, la red produce una salida que se compara con el valor real. A partir de esta comparación se calcula un error, que indica cuánto se ha desviado el resultado del modelo. El objetivo del aprendizaje es reducir progresivamente este error, modificando los pesos de la red [15].

2.5.2 Redes neuronales para datos secuenciales y temporales

En numerosos problemas de ingeniería, los datos no aparecen de forma aislada, sino organizados como secuencias en las que el orden de los valores es relevante. Este tipo de información se conoce como datos secuenciales o temporales, y se caracteriza porque cada observación está relacionada con las anteriores y, en muchos casos, con las posteriores. Ejemplos habituales de este tipo de datos incluyen series temporales, señales, trayectorias o registros obtenidos de forma continua en el tiempo [14].

A diferencia de los datos estáticos, en los que cada muestra puede analizarse de manera independiente, los datos temporales contienen dependencias entre instantes consecutivos. Esto implica que el valor actual no solo depende de las variables observadas en ese momento, sino también del historial previo. Por este motivo, el análisis de datos secuenciales requiere modelos capaces de tener en cuenta la evolución temporal de la información [15].

Las redes neuronales tradicionales, tratan cada entrada como un vector independiente y no incorporan información sobre el orden de los datos. En consecuencia, surge la necesidad de modelos específicamente diseñados para trabajar con secuencias [15].

Las redes neuronales para datos secuenciales y temporales están orientadas a procesar información en la que el tiempo y el orden de los datos recibidos tienen un papel fundamental. Estos modelos permiten incorporar el contexto temporal, es decir, utilizar información pasada para interpretar correctamente el estado presente del sistema. De este modo, es posible representar patrones dinámicos, cambios de comportamiento y tendencias que no serían detectables mediante un análisis tradicional [11].

Un aspecto clave de este tipo de redes es su capacidad para manejar secuencias de longitud variable, lo que las hace especialmente adecuadas para problemas en los que la duración de los registros es variable [15].

En el análisis de datos complejos, las redes neuronales orientadas a secuencias proporcionan una base sólida para la modelación de fenómenos dinámicos. Como se ha expuesto, su uso resulta especialmente útil cuando el comportamiento del sistema evoluciona con el tiempo y presenta regularidades temporales. Esto sienta las bases para el empleo de arquitecturas neuronales más específicas que se abordarán en los apartados siguientes.

2.5.3 Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes (*Recurrent Neural Networks*, RNN) son un tipo de arquitectura neuronal diseñada de forma específica para el tratamiento de datos secuenciales y temporales, en los que el orden de las observaciones, como se mencionó en el apartado anterior, es relevante. A diferencia de las redes neuronales tradicionales, las RNN incorporan mecanismos que permiten utilizar información de instantes anteriores para procesar el estado actual, lo que las hace especialmente adecuadas para modelar secuencias y series temporales [15][14].

La característica fundamental de las redes neuronales recurrentes (RNN) es la presencia de conexiones recurrentes, que permiten que la salida de una neurona en un instante determinado influya en el procesamiento de los instantes posteriores. De este modo, la red mantiene una forma de memoria interna, que almacena información sobre el pasado reciente de la secuencia. Esta memoria permite capturar dependencias temporales que no pueden representarse adecuadamente mediante redes neuronales sin recurrencia [11].

Desde un punto de vista conceptual, una RNN procesa una secuencia de datos de manera iterativa. En cada instante temporal, la red recibe una nueva entrada y la combina con el estado oculto previo, generando un nuevo estado que se transmite al siguiente instante. De este modo, la salida de la red en un momento dado depende tanto de la entrada actual como del contexto proporcionado por los valores anteriores de la secuencia [15] [14].

Como se puede observar en la Figura 2-22, que muestra el esquema general de una red neuronal recurrente desplegada en el tiempo, la misma estructura neuronal se aplica repetidamente a cada elemento de la secuencia, compartiendo los mismos pesos en todos los instantes temporales. Esta propiedad permite procesar secuencias de longitud variable manteniendo una arquitectura compacta y coherente.

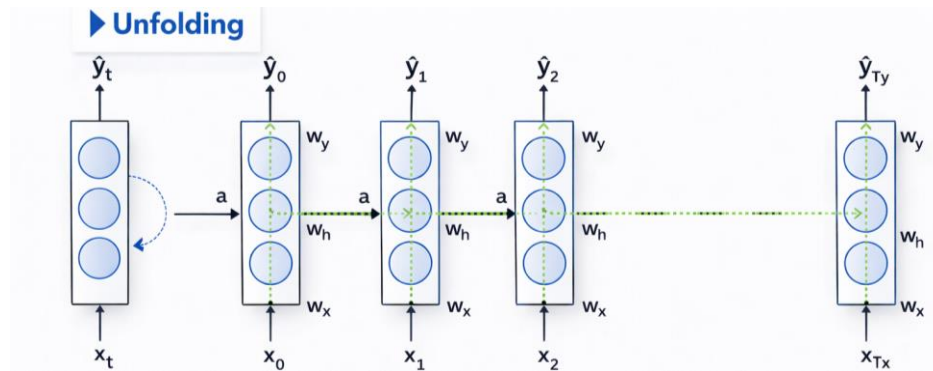


Figura 2-22 Esquema de funcionamiento de una red neuronal recurrente desplegada en el tiempo (Fuente: propia generada con IA)

Una de las principales ventajas de las RNN es su capacidad para modelar dependencias temporales y patrones dinámicos. Esto resulta esencial en problemas donde el comportamiento del sistema evoluciona progresivamente. Estas redes han sido ampliamente utilizadas en tareas como el análisis de series temporales, el procesamiento de señales y el modelado de secuencias, debido a su capacidad para incorporar el contexto temporal en el proceso de aprendizaje.

No obstante, las RNN clásicas presentan limitaciones cuando se aplican a secuencias demasiado largas. En particular, pueden experimentar dificultades para conservar información relevante durante muchos pasos temporales, lo que afecta a su capacidad para aprender dependencias a largo plazo. Este problema, conocido y ampliamente documentado en la literatura, ha motivado el desarrollo de arquitecturas recurrentes avanzadas que se abordarán en los apartados siguientes [15].

2.6 Redes neuronales recurrentes avanzadas

2.6.1 Redes neuronales recurrentes LSTM

Las redes neuronales recurrentes de tipo LSTM (*Long Short-Term Memory*) se diseñaron para mejorar el modelado de dependencias temporales en secuencias, sobre todo cuando existe información relevante que se debe mantener durante muchos pasos de tiempo en el proceso. Las redes LSTM se propusieron inicialmente para solucionar los problemas que presentan las redes neuronales recurrentes tradicionales cuando se trabaja con secuencias largas, ya que estas tienen dificultades para aprender y mantener la información relevante que aparece a lo largo del tiempo durante el proceso de entrenamiento del modelo [17].

De forma general, una LSTM es una arquitectura recurrente que incorpora una celda de memoria interna. Esta celda permite almacenar y mantener información relevante durante periodos prolongados. De esta manera, facilita que el modelo tenga en cuenta acontecimientos pasados que siguen siendo importantes para interpretar el estado actual del sistema. Este enfoque, como en las redes recurrentes tradicionales, resulta especialmente adecuado para datos secuenciales y temporales, como series

temporales, señales o trayectorias, donde el contexto histórico influye directamente en el comportamiento observado. Sin embargo, las LSTM tienen una mayor capacidad a la hora de gestionar procesos que se prolongan más durante el tiempo [17].

El elemento diferencial de las redes LSTM es el uso de mecanismos de control, conocidos como puertas (gates), que regulan el flujo de información dentro de la celda de memoria. Estas puertas permiten decidir qué información se conserva, cuál se descarta y cuál se utiliza para generar la salida en cada instante temporal. De forma simplificada, una celda LSTM incluye tres puertas principales [17]:

- **Puerta de olvido (forget gate):** decide qué parte de la información almacenada previamente deja de ser relevante y debe eliminarse.
- **Puerta de entrada (input gate):** controla qué nueva información procedente de la entrada es importante se incorpora a la memoria.
- **Puerta de salida (output gate):** determina qué parte de la información almacenada se utiliza para generar la salida del modelo en el instante actual.

Como se puede observar en la Figura 2-23, las redes LSTM introducen una estructura interna más compleja que las redes neuronales recurrentes tradicionales, incorporando mecanismos adicionales (puertas) que permiten gestionar de forma más eficiente la información temporal.

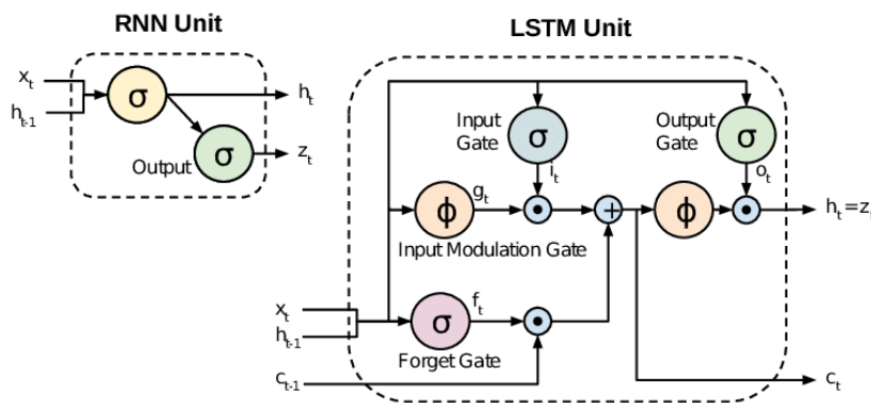


Figura 2-23 Comparación entre una RNN y una unidad LSTM, mostrando la estructura interna de la celda de memoria y el funcionamiento de sus puertas de control [32]

Gracias a esta nueva estructura, las redes LSTM son capaces de aprender dependencias temporales de largo alcance y mantener información importante a lo largo del tiempo sin que el proceso de entrenamiento se degrade. Esto permite capturar patrones dinámicos, tendencias y cambios progresivos de comportamiento que serían difíciles de representar mediante modelos que solo consideran información local o ventanas temporales cortas [17].

En comparación con las redes neuronales recurrentes tradicionales, las LSTM presentan una mayor complejidad estructural y un mayor número de parámetros, lo que implica un coste computacional superior. No obstante, esta complejidad adicional se ve compensada por una mayor estabilidad durante el entrenamiento y una capacidad de modelado significativamente mejor en problemas secuenciales exigentes.

2.6.2 Transformer

Los *Transformers* son una arquitectura de redes neuronales profundas ampliamente utilizada en Inteligencia Artificial y *Machine learning*. Estas se emplean especialmente en tareas que involucran el procesamiento de datos secuenciales, como texto, señales o series temporales. A diferencia de modelos anteriores (RNN o LSTM), los *transformers* no procesan los datos de forma secuencial, sino en paralelo, lo que permite entrenarlos mucho más rápido y capturar dependencias a largo plazo de forma más eficiente [24].

La principal innovación de los *transformers* es el uso del mecanismo de atención, concretamente la denominada *self-attention*. Este mecanismo permite que el modelo evalúe de forma dinámica la importancia relativa de cada elemento de la secuencia con respecto a los demás, independientemente de su posición. De este modo, el modelo tiene la capacidad de capturar dependencias a largo plazo de manera más eficiente que las arquitecturas recurrentes tradicionales, las cuales procesan la información de forma estrictamente secuencial [24].

Un transformer se divide en dos bloques principales:

- **Encoder:** Se encarga de analizar la información de entrada y transformarla en una representación interna que el modelo pueda entender.
- **Decoder:** Usa la información generada por el encoder para producir una salida, como una traducción, una predicción o un nuevo dato generado.

Este mecanismo a su vez se compone en varias capas apiladas con el objetivo de refinar la información recibida y aprender relaciones cada vez más complejas. De esta forma, irá obteniendo resultados más precisos y sutiles. Además, cada una de estas capas integra mecanismos de atención múltiple (*multi-head attention*) y redes neuronales *feed-forward*, permitiendo al modelo extraer relaciones complejas entre los elementos de la secuencia de entrada. La atención múltiple lo que nos otorga es la capacidad de prestar atención en paralelo, fijándose en un tipo de relación distinta. Una vez realizada la atención múltiple, cada elemento pasa por una red neuronal sencilla sin memoria, llamada *feed-forward*, cuyo objetivo es ajustar y afinar los valores. Además, se incorporan técnicas como las conexiones residuales y la normalización de capas, cuyo objetivo es evitar problemas durante el entrenamiento y facilitar que el modelo aprenda correctamente cuando se emplean arquitecturas profundas [24]. En la Figura 2-24 se aprecia el esquema básico de funcionamiento de un *Transformer*.

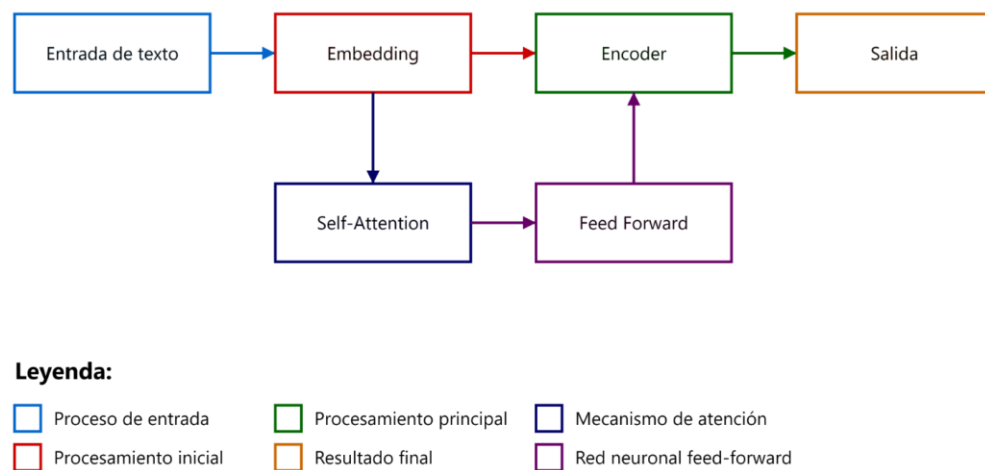


Figura 2-24 Esquema básico del funcionamiento de un Transformer [40]

Actualmente las arquitecturas *Transformer* han desplazado a los modelos recurrentes en muchas tareas de procesamiento de lenguaje natural, convirtiéndose en la arquitectura predominante en el estado del arte actual, siendo además la base fundamental de los grandes modelos de lenguaje (LLM) como puede ser Chat GPT[54].

2.7 Modelos generativos aplicados a secuencias temporales

2.7.1 Modelos generativos en aprendizaje profundo

Los modelos generativos constituyen una familia de técnicas dentro del Aprendizaje Profundo cuyo objetivo principal es aprender la distribución subyacente de los datos a partir de un conjunto de ejemplos, con el fin de generar nuevas muestras similares a las observadas. A diferencia de los modelos

discriminativos, que se centran en predecir una salida a partir de una entrada concreta, los modelos generativos tratan de capturar cómo se distribuyen los datos en el espacio de características, permitiendo así la creación de datos sintéticos coherentes con los reales [11][14].

En términos generales, un modelo generativo aprende una representación interna de los datos de entrenamiento que le permite estimar la probabilidad de aparición de determinadas configuraciones. Esto hace posible no solo generar nuevas muestras, sino también completar datos incompletos, simular escenarios plausibles o analizar la variabilidad inherente a un sistema complejo. Este enfoque resulta especialmente relevante en contextos donde los datos presentan una estructura compleja y multidimensional, como ocurre en series temporales, trayectorias o secuencias dinámicas.

En el ámbito del aprendizaje profundo, los modelos generativos se apoyan en arquitecturas neuronales capaces de representar relaciones no lineales de alta complejidad. Estas arquitecturas permiten transformar un espacio latente, generalmente de menor dimensión, en muestras que siguen las características estadísticas de los datos originales. El proceso de entrenamiento ajusta los parámetros del modelo para que las muestras generadas sean lo más similares posible a los datos reales, según un criterio de optimización definido [11].

Los modelos generativos han ganado un papel destacado en aplicaciones de ingeniería debido a su capacidad para simular comportamientos realistas y ampliar conjuntos de datos cuando la información disponible es limitada. En problemas relacionados con el análisis de sistemas dinámicos, como el que se estudiará, estos modelos permiten generar secuencias sintéticas que conservan las propiedades temporales y espaciales observadas en los datos reales, lo que resulta de gran interés para tareas de simulación y validación de algoritmos [14].

Como se puede observar en la Figura 2-25, que muestra un esquema conceptual del funcionamiento de un modelo generativo profundo, el modelo aprende a transformar una representación latente en datos con características similares a las del conjunto original, cerrando así el ciclo entre aprendizaje y generación.

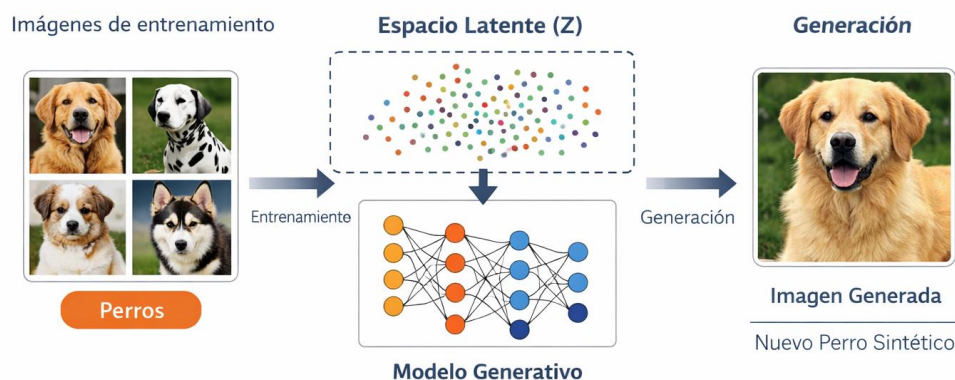


Figura 2-25 Proceso general de aprendizaje y generación en modelos generativos profundos (Fuente: propia generada con IA)

2.7.2 Redes Generativas Antagónicas (GAN) para secuencias temporales

Las Redes Generativas Antagónicas (GAN) constituyen una de las arquitecturas más relevantes dentro del aprendizaje profundo generativo. Su principio fundamental se basa en el entrenamiento simultáneo de dos redes neuronales con objetivos opuestos: un generador, encargado de producir datos sintéticos, y un discriminador, cuya función es distinguir entre datos reales y datos generados artificialmente. Ambos modelos se entrenan de forma competitiva mediante un proceso adversarial que permite mejorar progresivamente la calidad de las muestras generadas [22].

En el contexto de datos secuenciales y temporales, como series temporales, trayectorias o señales continuas, las GAN han sido adaptadas para aprender no solo la distribución de los valores individuales, sino también las dependencias temporales entre instantes consecutivos. Esto supone un reto adicional frente a los datos estáticos, ya que el modelo debe preservar la coherencia temporal y las relaciones dinámicas presentes en los datos originales [21].

A diferencia de los enfoques clásicos de modelado temporal, las GAN aplicadas a secuencias temporales permiten generar secuencias completas de datos sintéticos que reproducen patrones realistas observados en los datos reales. Para ello, el generador aprende a producir secuencias que imitan la evolución temporal de las variables, mientras que el discriminador evalúa si una secuencia completa es coherente desde el punto de vista temporal o si ha sido generada artificialmente [22].

Como se puede observar en la Figura 2-26, que muestra el esquema general de una GAN aplicada a datos temporales, el generador recibe como entrada un vector aleatorio o una representación latente y produce una secuencia completa, mientras que el discriminador analiza dicha secuencia considerando su estructura temporal.

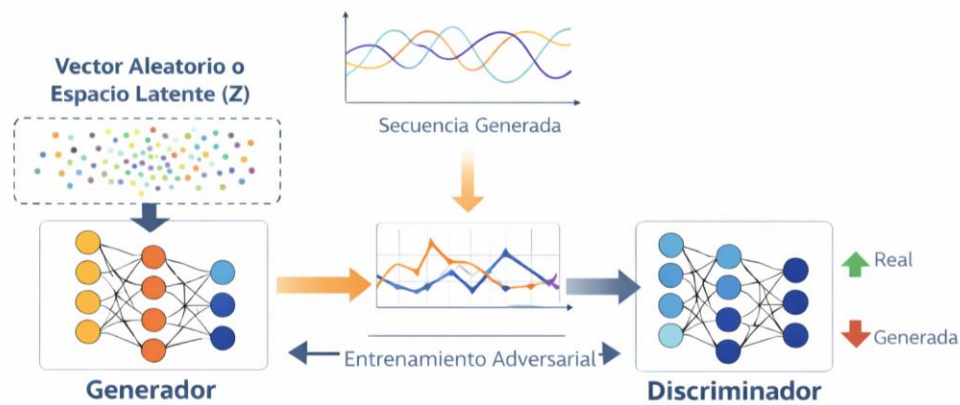


Figura 2-26 Esquema conceptual de una GAN aplicada a secuencias temporales (Fuente: propia generada con IA)

Diversos trabajos recientes destacan que las GAN para secuencias temporales resultan especialmente adecuadas para la generación de datos sintéticos, la ampliación de conjuntos de datos y la simulación de escenarios complejos, especialmente en dominios donde los datos reales son escasos, sensibles o difíciles de obtener [21]. En este sentido, estas arquitecturas han sido aplicadas con éxito en ámbitos como datos financieros, sensores industriales, tráfico, señales biomédicas o series temporales multivariantes.

Sin embargo, el entrenamiento de GAN en datos temporales presenta una serie de desafíos específicos. Entre los más relevantes se encuentran la inestabilidad durante el entrenamiento, el colapso de modos y la dificultad para evaluar la calidad de las secuencias generadas, ya que no existe una métrica única que capture simultáneamente realismo, diversidad y coherencia temporal [22]. Estos aspectos han motivado el desarrollo de variantes específicas y mejoras arquitectónicas orientadas a datos secuenciales.

En conjunto, las GAN aplicadas a secuencias temporales proporcionan un marco potente y flexible para la generación de datos sintéticos dinámicos, constituyendo una herramienta clave en estudios de simulación, análisis avanzado y aprendizaje automático basado en datos temporales. Este enfoque resulta especialmente relevante como base para modelos más especializados, como las GAN secuenciales basadas en redes recurrentes o LSTM, que se abordarán en los apartados siguientes.

2.7.3 GAN secuenciales basadas en LSTM

Las Redes Generativas Antagónicas (GAN) secuenciales basadas en LSTM representan una extensión natural de las GAN tradicionales cuando los datos a modelar presentan una estructura temporal explícita, como ocurre en series temporales, trayectorias o secuencias multivariantes. En este tipo de problemas, no basta con generar valores individuales realistas, sino que es necesario reproducir correctamente la dinámica temporal y las dependencias existentes entre observaciones consecutivas [22].

El generador basado en LSTM es la parte del modelo que se encarga de crear secuencias nuevas de datos que se parezcan a las secuencias reales. Para empezar a generar una secuencia, el generador necesita un punto de partida, que no es más que una señal inicial que le indica que debe comenzar el proceso de generación. Esta señal inicial no contiene información concreta sobre los datos reales, sino que sirve simplemente para iniciar la generación de la secuencia.

A partir de ese punto inicial, la red LSTM va generando la secuencia paso a paso. En cada uno de estos, la LSTM utiliza la información que ha generado en los pasos anteriores para decidir cuál debe ser el siguiente valor de la secuencia. Gracias a su memoria interna, la LSTM es capaz de recordar lo que ha ocurrido antes y mantener una evolución coherente en el tiempo. De este modo, el generador no produce valores aislados, sino una secuencia completa que sigue un patrón temporal similar al observado en los datos reales [22]. De este modo, el generador aprende a producir secuencias que respetan la continuidad temporal y reproducen patrones dinámicos similares a los observados en los datos reales.

El discriminador basado en LSTM es el componente encargado de evaluar las secuencias generadas y decidir si parecen reales o no. Para ello, recibe como entrada secuencias completas, que pueden proceder del conjunto de datos reales o haber sido generadas por el generador.

La red LSTM del discriminador analiza la secuencia teniendo en cuenta el orden de los datos y cómo van cambiando a lo largo del tiempo. En lugar de fijarse solo en valores individuales, el discriminador observa la coherencia temporal de la secuencia: si las transiciones entre un instante y el siguiente son razonables y si el comportamiento global se parece al de las secuencias reales utilizadas durante el entrenamiento. Al finalizar el análisis de la secuencia, el discriminador emite una decisión indicando si considera que la secuencia es real o generada. Esta decisión se utiliza durante el entrenamiento para mejorar tanto al discriminador como al generador, haciendo que el generador aprenda progresivamente a producir secuencias cada vez más realistas [21][22].

El entrenamiento de una GAN secuencial sigue el principio de aprendizaje adversarial, en el que el generador trata de producir secuencias cada vez más realistas con el objetivo de engañar al discriminador, mientras que este último mejora su capacidad para distinguir entre secuencias reales y generadas. Este proceso iterativo conduce progresivamente a un equilibrio en el que las secuencias sintéticas presentan características estadísticas y temporales similares a las reales [21][22].

Una de las principales ventajas del uso de LSTM dentro de GAN secuenciales es su capacidad para manejar secuencias de longitud variable y conservar información relevante durante intervalos temporales prolongados. Esto resulta especialmente útil en aplicaciones donde el comportamiento actual depende de eventos pasados, como en trayectorias, señales temporales o procesos dinámicos complejos.

2.7.4 Redes GAN condicionales (cGAN)

Las Redes Generativas Adversarias Condicionales (Conditional Generative Adversarial Networks, cGAN) son una extensión de las GAN tradicionales que permiten introducir control en el proceso de generación de datos sintéticos mediante el uso de información adicional, denominada condición. Este tipo de arquitectura surge para superar una de las principales limitaciones de las GAN convencionales: la imposibilidad de decidir de forma explícita qué tipo de datos se desean generar.

En una GAN clásica, el generador crea muestras sintéticas a partir de ruido aleatorio, mientras que el discriminador intenta distinguir entre datos reales y generados. Aunque este enfoque ha demostrado ser eficaz para generar datos realistas, presenta dificultades cuando se trabaja con conjuntos de datos

etiquetados o cuando es necesario generar ejemplos asociados a una clase concreta. Esta falta de control reduce la utilidad práctica de los datos sintéticos en determinadas aplicaciones técnicas[22].

Las cGAN abordan este problema incorporando una condición tanto en el generador como en el discriminador. Dicha condición puede representar una etiqueta de clase, un atributo específico o cualquier información relevante del conjunto de datos. De este modo, el generador aprende a producir datos sintéticos coherentes con la condición impuesta, mientras que el discriminador evalúa las muestras teniendo en cuenta esa misma información. Este enfoque permite generar datos sintéticos más estructurados y orientados a un objetivo concreto [21][22]. En la Figura 2-27 se puede observar el funcionamiento de este condicionante.

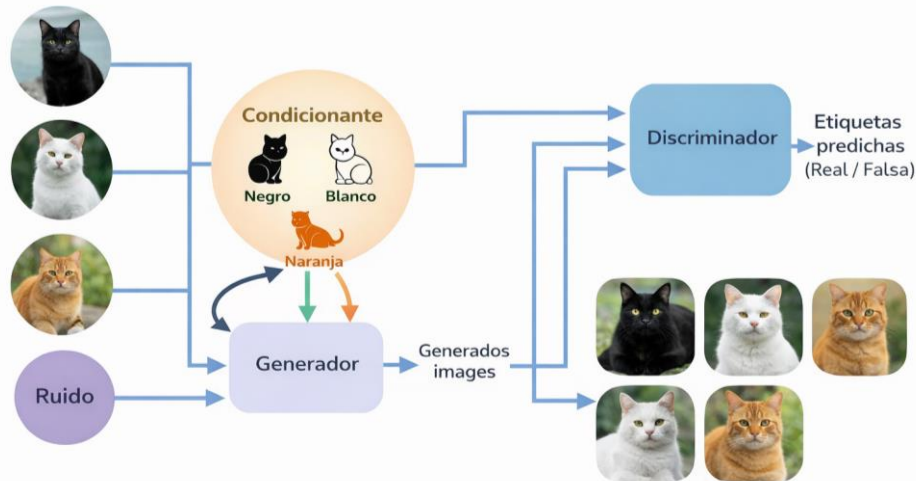


Figura 2-27 Representación de cGAN (Fuente: propia generada con IA)

El proceso de entrenamiento mantiene el esquema adversario característico de las GAN, pero añade la condición como parte de la entrada de ambos modelos. El generador intenta crear muestras que no solo parezcan reales, sino que además respeten la condición establecida, mientras que el discriminador aprende a detectar incoherencias entre los datos y dicha condición. Este mecanismo mejora el control sobre la generación y facilita el uso de los datos sintéticos en tareas como validación de modelos, simulación de escenarios o aumento de datos.

Entre las principales ventajas de las GAN condicionales se encuentran el mayor control sobre el proceso generativo, la mejor adaptación a conjuntos de datos multiclase y la posibilidad de generar datos sintéticos más útiles desde el punto de vista práctico. Estas características resultan especialmente relevantes en contextos donde los datos reales son escasos, sensibles o costosos de obtener, manteniendo además beneficios como la preservación de la privacidad [22].

El empleo de redes generativas adversariales condicionales (cGAN) en este trabajo, que será el empleado, se justifica por la necesidad de generar trayectorias sintéticas de buques que no solo reproduzcan patrones espaciales realistas, sino que además respeten condiciones operativas específicas, como el punto y la velocidad de inicio y destino. A diferencia de modelos generativos no condicionados, la cGAN permite guiar explícitamente la generación mediante variables físicas relevantes, asegurando coherencia global en la aproximación al puerto y consistencia dinámica en términos de evolución espacial y velocidad. Este enfoque resulta especialmente adecuado en el contexto del análisis AIS, donde el objetivo no es únicamente imitar distribuciones estadísticas, sino producir trayectorias plausibles desde el punto de vista náutico, capaces de preservar tanto la variabilidad local del movimiento como las restricciones geométricas y operativas propias del entorno portuario.

2.8 Análisis del entorno marítimo de interés en la Armada

El análisis o Conocimiento del Entorno Marítimo (CEM) se trata del proceso mediante el cual se integran y evalúan datos procedentes de múltiples fuentes de información, con el objetivo de construir una visión fiable y detallada de lo que ocurre en los espacios marítimos de interés nacional. A través de este análisis continuo y sistemático, es posible disponer de una representación actualizada, prácticamente en tiempo real, de las actividades que se desarrollan en dichas aguas. La presentación de este método es conocida como RMP (Recognized Maritime Picture).

Miles de buques navegan cada año a través de las aguas de soberanía española. Es por esta razón que es fundamental la defensa de estas líneas de tráfico marítimo, tanto para la economía nacional como la internacional. El ámbito marítimo constituye un entorno especialmente atractivo para la realización de actividades ilícitas de diversa naturaleza, que van desde el terrorismo y el tráfico ilegal de armamento hasta la explotación no autorizada del patrimonio subacuático o los flujos irregulares de migración, fenómenos que continúan representando un desafío relevante para la seguridad de la Unión Europea y de la OTAN. Ante este escenario, resulta imprescindible ejercer una supervisión permanente y eficaz de los espacios marítimos de interés. En este contexto, la Armada asume la responsabilidad de garantizar la protección, control y vigilancia de dichos espacios, contribuyendo de forma directa a la seguridad marítima nacional y a la salvaguarda de los intereses estratégicos del Estado, tal y como se ve en la Figura 2-28.



Figura 2-28 Fragata F-104 "Méndez Núñez" escoltando a un pesquero [41]

Debido a este objetivo, se hace necesario disponer de capacidades avanzadas de vigilancia y control de los espacios marítimos de interés, en las que el Conocimiento del Entorno Marítimo (CEM) desempeña un papel esencial. En la actualidad, la disponibilidad de una representación precisa, coherente y actualizada de la situación marítima constituye un elemento crítico para el planeamiento, conducción y seguimiento de las operaciones de seguridad marítima desarrolladas por las Fuerzas Armadas a través de la Fuerza de Acción Marítima (FAM). Esta capacidad de obtención, fusión y análisis de información procedente de múltiples fuentes se articula fundamentalmente a través del Centro de Operaciones y Vigilancia de la Acción Marítima (COVAM), que actúa como el principal sistema de apoyo a la decisión para la generación y mantenimiento del Conocimiento del Entorno Marítimo.

2.8.1 Centro de Operaciones y Vigilancia de Acción Marítima de la Armada (COVAM)

La Península Ibérica ocupa una posición geoestratégica de primer orden en el contexto global, situándose en un punto de convergencia esencial de las principales rutas marítimas internacionales. Esta

circunstancia confiere a España un papel destacado en el mantenimiento y la seguridad del tráfico marítimo mundial. Asimismo, su emplazamiento como nexo natural entre Europa, África y América, junto con la proyección estratégica que aportan las Islas Canarias, refuerza su relevancia como uno de los países con mayor capacidad para influir y destacar en el ámbito del transporte marítimo internacional.

En su permanente lucha por mantener control y seguimiento de las zonas de interés para España, la Armada, representada en este caso por el Almirante de Acción Marítima (ALMART), tiene entre sus principales cometidos ser el Comandante del Mando Operativo Marítimo (CMOM). Este organismo tiene como cometido principal la planificación, dirección y control de las actuaciones orientadas a la vigilancia y protección de los espacios marítimos bajo soberanía, responsabilidad y especial interés nacional. Para el cumplimiento eficaz de estas funciones, el CMOM mantiene una coordinación permanente con múltiples entidades y autoridades, tanto de carácter civil como militar, en los ámbitos nacional e internacional. Con el objetivo de ayudar al CMOM en su cometido, la Armada cuenta con el COVAM (Centro de Operaciones y Vigilancia de Acción Marítima de la Armada). Este se ubica en la ciudad de Cartagena y se encuadra dentro del Estado Mayor de la Fuerza de Acción Marítima. Véase en la Figura 2-29.



Figura 2-29 Centro de Operaciones y Vigilancia de la Acción Marítima (COVAM) [30]

El COVAM actúa de manera coordinada con las unidades de la Fuerza de Acción Marítima (FAM) en la protección y vigilancia de los espacios marítimos considerados de interés nacional, que incluyen tanto las aguas bajo soberanía española como áreas estratégicas de carácter internacional, tales como el Golfo de Guinea, el Cuerno de África o el Mar Mediterráneo, tal y como se representa en la Figura 2-30. Para el adecuado desempeño de estas funciones, uno de los objetivos fundamentales del COVAM es la generación y distribución de la Imagen Marítima Reconocida (RMP) a las unidades de la Armada, estaciones terrestres y organismos civiles que lo requieran. Asimismo, el centro mantiene una operatividad continua, funcionando de manera ininterrumpida las 24 horas del día, los siete días de la semana, con el fin de atender el amplio conjunto de misiones que tiene encomendadas, entre las que destacan [25]:

- **Obtención y mantenimiento del Conocimiento del Entorno Marítimo (CEM):** El COVAM integra y analiza información procedente de múltiples sensores y fuentes (AIS, radares, satélites, informes de unidades, etc.) para generar una imagen actualizada y coherente de la situación marítima, conocida como Recognized Maritime Picture (RMP).
- **Vigilancia y control de los espacios marítimos de interés nacional:** Supervisa de forma permanente tanto las aguas bajo soberanía y jurisdicción española como otras áreas estratégicas de interés, contribuyendo a la detección temprana de riesgos, amenazas o actividades anómalas.

- **Apoyo al planeamiento, conducción y seguimiento de operaciones marítimas:** Proporciona apoyo directo al Mando Operativo Marítimo (CMOM) en el seguimiento de las operaciones de la Fuerza de Acción Marítima, facilitando la toma de decisiones operativas.
- **Coordinación con organismos civiles y militares, nacionales e internacionales:** Actúa como punto de enlace con otras autoridades civiles (como Salvamento Marítimo) y con organizaciones militares nacionales y aliadas, favoreciendo la cooperación y el intercambio de información en el ámbito marítimo.

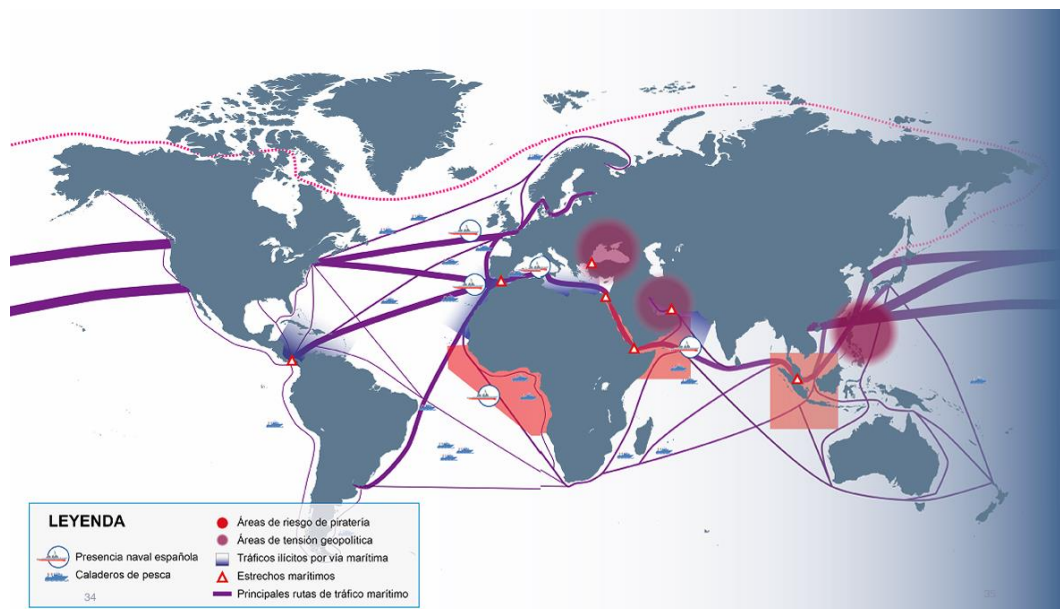


Figura 2-30 Zonas interés marítimo nacional [25]

2.8.2 Uso de Inteligencia Artificial para el Conocimiento del Entorno Marítimo (CEM)

El Conocimiento del Entorno Marítimo (CEM) es una capacidad fundamental para la Armada, al permitir disponer de una visión precisa y actualizada de las actividades que se desarrollan en los espacios marítimos de interés nacional. Esta función resulta esencial para la vigilancia marítima, la seguridad de las rutas de comunicación y la detección de actividades ilícitas. En este ámbito, el Centro de Operaciones y Vigilancia de Acción Marítima (COVAM) constituye el elemento central encargado de integrar la información procedente de múltiples fuentes y generar la Imagen Marítima Reconocida (RMP) [26].

El crecimiento exponencial del volumen y la heterogeneidad de los datos disponibles (procedentes de sistemas AIS, sensores costeros, radares y satélites) ha puesto de manifiesto las limitaciones de los métodos de análisis con los que se contaba. Ante este escenario, la Armada ha reconocido de forma explícita la Inteligencia Artificial (IA) y la analítica avanzada como tecnologías clave para mejorar el tratamiento de la información, apoyar la toma de decisiones y reforzar el mantenimiento del CEM [27].

Esta apuesta por la IA se refleja en la orientación doctrinal y estratégica de la Armada, donde se identifican estas tecnologías como habilitadoras para el análisis automatizado de grandes volúmenes de datos y la detección de patrones de comportamiento anómalos en el tráfico marítimo, capacidades directamente alineadas con las funciones del COVAM [27].

De forma paralela, se han impulsado programas de modernización de los sistemas de mando y control, entre los que destaca la actualización tecnológica del COVAM, gestionada por la Dirección General de Armamento y Material (DGAM). Estas iniciativas tienen como objetivo mejorar la integración de sistemas, la gestión de la información y sentar las bases para la incorporación progresiva de herramientas basadas en Inteligencia Artificial aplicadas al CEM [28].

Desde el punto de vista operativo, el empleo de técnicas de IA permite avanzar hacia un modelo de vigilancia marítima más proactivo y predictivo, facilitando la identificación de comportamientos

anómalos y la mitigación de las limitaciones inherentes a los datos AIS, como la ausencia o incorrección de determinados campos. De este modo, la Armada avanza hacia un CEM más robusto y eficiente, enmarcado en el proceso de transformación digital asociado al concepto de Armada 4.0 [29].

3 DESARROLLO DEL TFG

Una vez establecido el marco teórico relativo al sistema AIS, la discretización espacial y los modelos de aprendizaje automático aplicables al análisis de trayectorias, en este capítulo se aborda el desarrollo práctico del trabajo. El objetivo es trasladar los conceptos expuestos anteriormente, a una implementación real que permita procesar datos AIS, estructurarlos adecuadamente y generar modelos capaces de reproducir comportamientos de navegación representativos de distintas clases de buques.

El desarrollo se articula de forma progresiva. En primer lugar, se define el entorno de trabajo y el área de estudio, estableciendo los criterios de filtrado espacial y selección de datos. A continuación, se describe la construcción y preprocesamiento del conjunto de datos AIS, incluyendo su discretización mediante cuadrículas espaciales. Sobre esta base se implementan distintos enfoques para la generación de trayectorias sintéticas, combinando modelos probabilísticos de transición espacial y modelos basados en aprendizaje automático, con el fin de analizar su capacidad para reproducir patrones de movimiento coherentes con los observados en los datos reales.

3.1 Entorno del trabajo y software empleado

Como se ha mencionado previamente en la introducción, la herramienta seleccionada para el tratamiento de los datos y la aplicación de técnicas de aprendizaje automático ha sido *Jupyter Lab*, mientras que el lenguaje de programación empleado para el desarrollo del código ha sido *Python*. En este apartado se justificará la elección de ambas tecnologías y se presentarán los primeros pasos realizados para su utilización en el desarrollo del presente trabajo.

3.1.1 *Jupyter Lab*

Como se mencionó con anterioridad, el entorno de trabajo empleado es *Jupyter Lab*. Este es accesible a través de un navegador web que permite el desarrollo interactivo de cuadernos, código y conjuntos de datos en un mismo espacio. Su diseño modular y configurable facilita la creación y gestión de flujos de trabajo orientados a tareas como el análisis y visualización de datos, la computación científica, el periodismo de datos y la aplicación de técnicas de aprendizaje automático [42].

La principal ventaja que ofrece este entorno es la capacidad de organizar el código creado en celdas separadas tal y como se puede observar en la Figura 3-1. Esto supone un gran avance a la hora de facilitar el trabajo al usuario que lo emplea, ya que permite ejecutar la programación de cada celda de manera independiente, permitiendo detectar errores con mayor facilidad sin tener que actualizar todo el código. Así mismo, el entorno *Jupyter Lab* no restringe el desarrollo a un único lenguaje de programación, lo que proporciona una mayor versatilidad tanto en la implementación del código como en el análisis de los resultados obtenidos. Por otro lado, la plataforma es compatible con múltiples lenguajes, ampliando así sus posibilidades de uso [42][43].

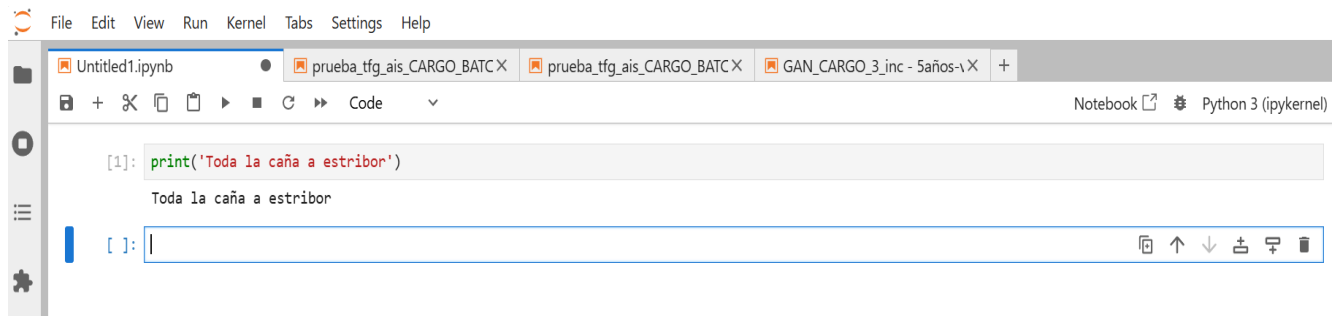


Figura 3-1 Interfaz para la escritura de código en Jupyter Lab (Fuente: propia)

Además de las mencionadas principales ventajas, la aplicación ofrece una serie de funcionalidades adicionales que refuerzan su utilidad en entornos de desarrollo y análisis de datos [43]:

- Es un Software gratuito
- Funciona desde un navegador designado.
- Permite trabajar conjuntamente con otros usuarios, ya que los notebooks y proyectos se pueden compartir.
- Permite visualizar directamente los datos, al mostrarse gráficos, tablas y resultados de manera inmediata.

Es por estas ventajas por lo cual la aplicación *Jupyter Lab* ha sido la seleccionada para la confección de este trabajo, destacando sobre el resto de las ventajas, la capacidad de poder ejecutar y probar las celdas de manera independiente. Esta ventaja resulta imprescindible a la hora de proyectar tareas de aprendizaje automático.

3.1.2 Lenguaje de programación Python

Dentro de los distintos lenguajes de programación que permite emplear *Jupyter Lab* para el desarrollo del código, la opción seleccionada para este Trabajo Fin de Grado ha sido *Python*. Esta elección se fundamenta en criterios técnicos directamente relacionados con la naturaleza del problema abordado, el tipo de datos empleados (datos AIS) y los objetivos de análisis y modelado planteados en el proyecto.

Python es un lenguaje de programación de alto nivel ampliamente utilizado en ámbitos como la ciencia de datos, el análisis de grandes volúmenes de información y el aprendizaje automático [57]. Este crecimiento está estrechamente ligado a la amplia comunidad de desarrolladores que lo respalda y a una sintaxis clara y legible, lo que facilita tanto la escritura como la interpretación del código y reduce la probabilidad de errores durante el desarrollo.

Así mismo dispone de un conjunto muy amplio y consolidado de librerías orientadas al tratamiento de datos estructurados en formato tabular, como es el caso de los datos AIS. Librerías como *NumPy* y *Pandas* permiten realizar de manera eficiente tareas fundamentales para este trabajo, tales como la limpieza de datos, el filtrado de registros erróneos, la transformación de variables y la agregación de información, etapas imprescindibles antes de aplicar cualquier técnica de aprendizaje automático.

3.1.3 MiniConda

Para el desarrollo del trabajo se ha empleado *Miniconda*, una distribución mínima del sistema *Conda* que permite al usuario crear entornos virtuales y seleccionar e instalar los paquetes necesarios[58][59].

3.2 Definición del área de estudio y criterios de filtrado espacial

3.2.1 Puerto de Barcelona

El Puerto de Barcelona (Figura 3-2) es uno de los principales puertos del Mediterráneo y un nodo logístico de referencia tanto a nivel nacional como internacional. Situado en la costa noreste de la península ibérica, desempeña un papel clave en el tráfico de mercancías y pasajeros, concentrando un elevado volumen de operaciones comerciales, industriales y de transporte marítimo. En 2024, atracaron en este puerto un total de 8381 buques mercantes, con buques que suman un arqueo bruto total de 334 376 531 GT[60]. Estas estadísticas no hacen más que mostrar la importancia y densidad de este puerto.



Figura 3-2 Plano del Puerto de Barcelona [60]

La elevada densidad de tráfico marítimo registrada en el puerto de Barcelona, junto con la presencia de buques de distinta tipología y tamaño, lo convierte en un escenario especialmente adecuado para el estudio de datos AIS. Este entorno permite analizar de forma consistente trayectorias, velocidades y patrones de maniobra bajo condiciones reales de operación portuaria, proporcionando una base de datos rica y representativa para el desarrollo y validación de los métodos propuestos en este Trabajo Fin de Grado.

3.2.2 Definición de polígonos y zonas de filtrado portuario

Debido a la variabilidad de comportamientos que tienen los barcos en el área de Barcelona, sumado a la gran cantidad de datos recibidos, se ha buscado una manera de filtrar correctamente aquellos datos que sí son útiles para el estudio, así como para diferenciar aquellos buques que se encuentran en maniobra de salida de puerto y los que se encuentran entrando.

Para la creación de todos los polígonos de una forma visual, y su posterior traspaso a lenguaje *Python* en *JupyterLab* mediante coordenadas geográficas, se ha empleado *Polyline tool*[64], una herramienta desarrollada por el Keene State College que permite poner *waypoints* sobre un mapa del globo y, de este modo, crear los polígonos que se describirán a continuación de manera precisa con sus respectivas coordenadas geográficas, tal y como se muestra en la Figura 3-3.

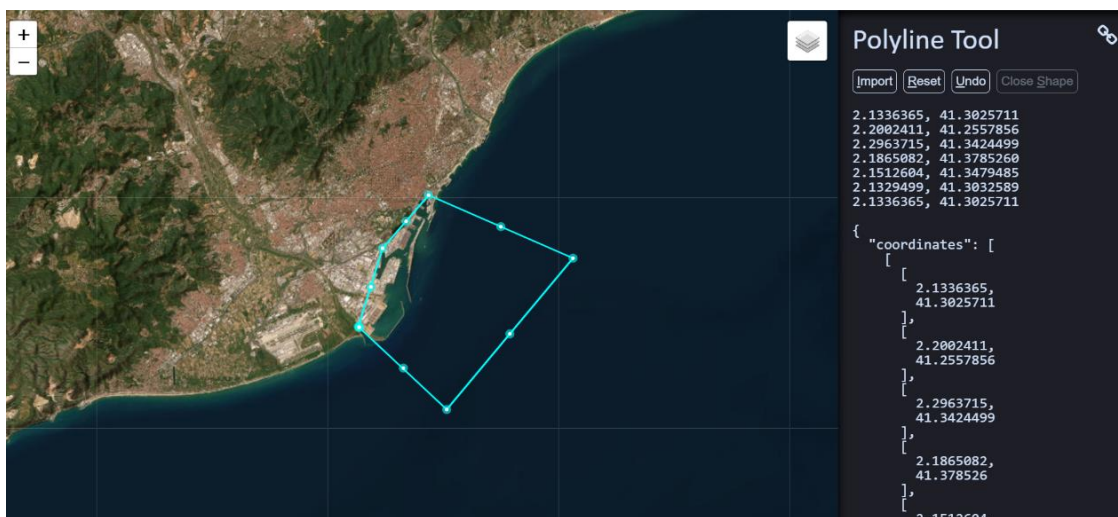


Figura 3-3 Ejemplo del uso de Polyline tool (Fuente: propia)

Para identificar a los barcos tipo *Cargo* que entraron en puerto, se crea una zona poligonal que abarca la extensión del interior del puerto de Barcelona para eliminar los barcos que se desplazaron por el interior del puerto y solo realizar el estudio a partir de la bocana de acceso al puerto y sus zonas adyacentes. Con este paso se descarta tanto el tráfico que se desarrolla exclusivamente en el interior del puerto (embarcaciones auxiliares y de apoyo), como las partes de las trayectorias de los buques que se desarrollan en el interior del puerto, descartando así trayectorias erráticas como puede ser una atracada, que se realiza a muy baja velocidad y haciendo cambios bruscos de rumbo debido a las maniobras con remolcadores.

Por otro lado, y de manera similar al anterior caso, se estableció otro polígono geográfico para filtrar los barcos tipo *Cargo* cuyas trayectorias finalizan en las inmediaciones del puerto de Barcelona y que no llegan a entrar en el recinto. Estos buques, que finalizaron sus trayectorias antes ya sea por fondear o por errores de transmisión de datos, se considerarán como entradas en puerto o recaladas hasta sus proximidades con el objetivo de enriquecer el estudio y entrenamiento. En la Figura 3-4 se muestran estos polígonos aplicados al puerto de Barcelona.

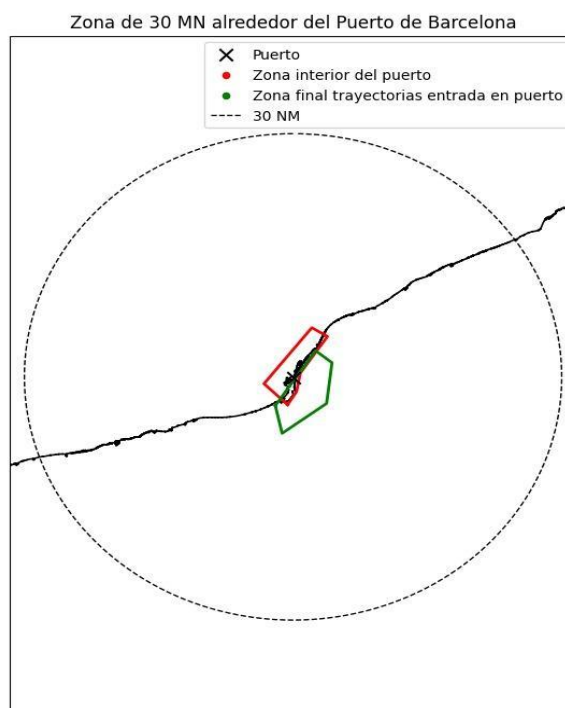


Figura 3-4 Polígonos filtradores aplicados al puerto de Barcelona (Fuente: propia)

Una vez definidas las dos áreas de filtrado, se introduce una tercera región destinada a la generación de trayectorias mediante celdas de tipo UBER (H3). A diferencia de las anteriores, esta zona no se utiliza para el cribado de datos, sino que actúa como región objetivo para las trayectorias sintéticas generadas por el algoritmo. De este modo, dicha área se establece como condición de parada o destino final en los modelos basados en mapas de calor, asegurando que las trayectorias obtenidas concluyan en una región coherente con los patrones de llegada de buques reales observados. Esta nueva área se muestra en verde en la Figura 3-5.

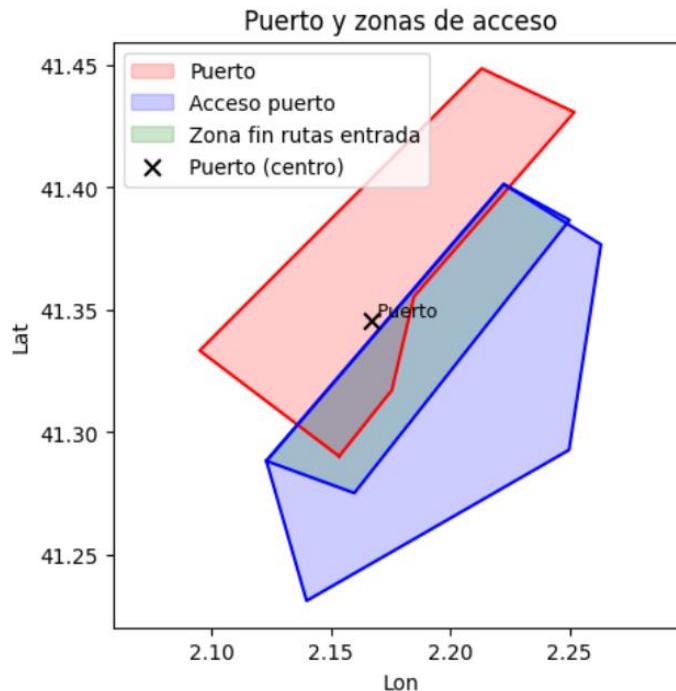


Figura 3-5 Áreas generadas para el filtrado y condiciones (Fuente: propia)

3.2.3 Delimitación geográfica y radio operativo de 30 millas náuticas

Adicionalmente, es necesario señalar que el alcance efectivo de los dispositivos de transmisión y recepción AIS depende en gran medida de las condiciones meteorológicas, cambiantes en cada día y época del año. Esta circunstancia provoca que, en determinados días, el receptor del puerto de Barcelona capte señales procedentes de áreas considerablemente alejadas, incluyendo zonas próximas al archipiélago balear o a la Comunidad Valenciana.

Es por esta razón que se tomó la decisión de crear un radio de 30 millas náuticas alrededor del puerto de Barcelona, con el objetivo de que solo se utilicen aquellos datos AIS que se reciban dentro de ese círculo marcado. Este filtro es clave a la hora de obtener datos útiles para el presente trabajo, descartando procesar información muy lejana a la zona de acción. En la Figura 3-6 se muestra la representación de este círculo frontera de 30 millas náuticas alrededor del Puerto de Barcelona.

Con el objetivo de hacer más visuales todos los mapas generados en este trabajo, se toma la decisión de cargar la línea de costa catalana en la zona de estudio. La línea de costa se define como la línea que representa la intersección entre la superficie terrestre y el mar en un instante determinado. Su posición no es fija, ya que puede experimentar variaciones como consecuencia de procesos naturales (mareas, la erosión o la sedimentación) así como debido a actuaciones de origen antrópico.

Gracias a la visualización de la línea de costa se puede apreciar con mayor facilidad la situación general del puerto, y cómo interactúan los barcos con el litoral. Los puntos generados fueron descargados

directamente del ICGC (Instituto Cartográfico y Geológico de Cataluña) [61] que los recoge mediante vuelos fotográficos periódicos que realiza sobre el litoral catalán. Estos datos descargados fueron posteriormente introducidos en el entorno *Python* del trabajo, viéndose en la Figura 3-6 el resultado de la línea de costa cargada junto al radio de 30 millas náuticas que delimita el procesamiento de datos.

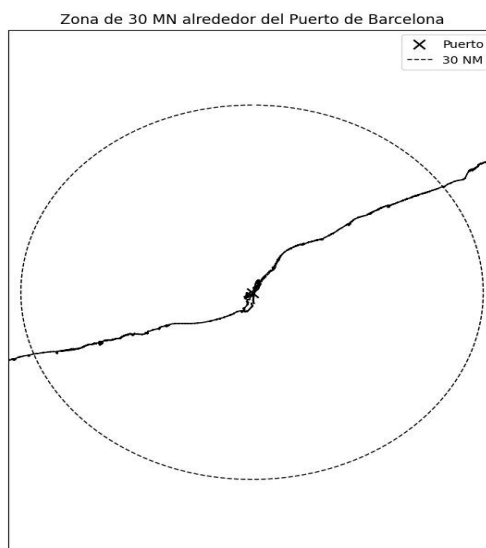


Figura 3-6 Línea de costa generada junto al radio de 30mn (Fuente: propia)

Una vez definidas e integradas todas las características que serán representadas en los mapas generados, con el objetivo de facilitar su interpretación visual, así como las distintas áreas destinadas al filtrado y al condicionamiento de los datos, se procede a la visualización del mapa resultante en pantalla, tal y como se muestra en la Figura 3-7. Este proceso de configuración y representación no solo resulta esencial para garantizar la coherencia y validez de las trayectorias obtenidas, sino que constituye también una fase clave para el análisis crítico de los resultados alcanzados. En consecuencia, la correcta definición de estos elementos permite evaluar de forma rigurosa el comportamiento del sistema propuesto y comprobar la adecuación del modelo desarrollado respecto a los patrones reales observados.

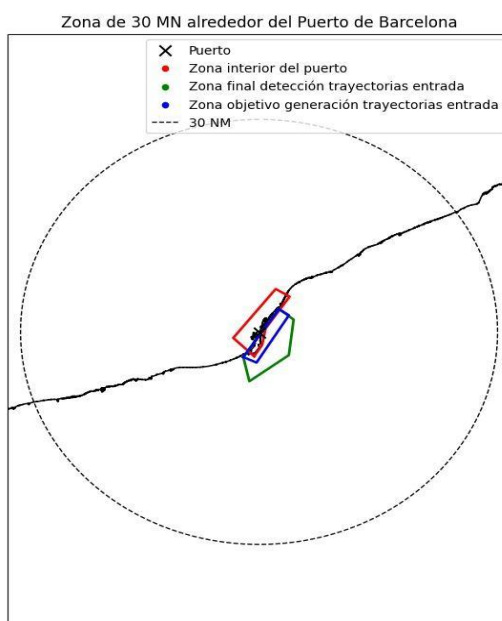


Figura 3-7 Representación final de los mapas mostrados en pantalla (Fuente: propia)

3.2.4 Polígono específico para buques pesqueros

Debido a la existencia de diversos puertos menores con actividad pesquera en los alrededores del puerto de Barcelona[63](Figura 3-8), así como la existencia de buques que se encuentran navegando hacia sus respectivos caladeros transmitiendo como que están pescando (Status 7), se ha visto necesaria la creación de un nuevo polígono específico. Este servirá para el filtrado de todos aquellos buques que, al comportarse de manera distinta a un buque pesquero que sale de puerto, faena y vuelve, aportaría ruido a los entrenamientos de los algoritmos de aprendizaje automático, empeorando las trayectorias sintéticas que genere y, por lo tanto, el resultado esperado.

Además, se observó que, en caso de no aplicar esta área de llegada y salida, el sistema recibía una gran cantidad de señales que provenían de puertos adyacentes, generando patrones de movimiento completamente distintos al deseado y, por lo tanto, imposibilitando un filtrado adecuado y un entrenamiento eficaz de los algoritmos.



Figura 3-8 Puertos en la costa de Barcelona (Fuente: propia con [62][63])

El objetivo del mencionado polígono no es otro que acotar el área donde salen y entran los pesqueros correspondientes al puerto de Barcelona. De esta forma, se filtrarán aquellos buques que salen del puerto, pero que no vuelven a entrar. Por lo tanto, este polígono descarta aquellos buques que se encuentran de paso por la zona dirigiéndose a sus caladeros sin actuar como un buque pesquero, así como aquellos que salen de puertos menores de la zona. Este polígono es el que aparece dibujado en la Figura 3-9 que, como se dijo anteriormente, representa el área de donde deben salir y retornar los pesqueros.

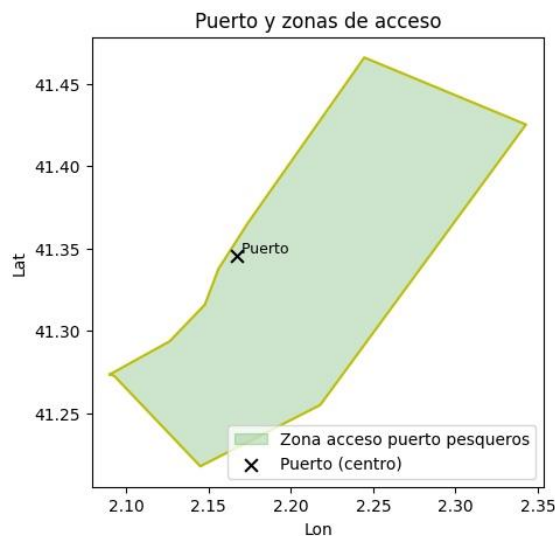


Figura 3-9 Polígono específico para buques pesqueros (Fuente: propia)

3.3 Construcción y preprocesamiento del conjunto de datos AIS

Tal y como se ha descrito previamente, el desarrollo de un sistema basado en aprendizaje automático requiere partir de un conjunto de datos suficientemente representativo, que le permita al modelo aprender a partir de ejemplos reales. En este trabajo, el objetivo principal es diseñar un algoritmo capaz de desarrollar e imitar el comportamiento de buque a partir de información AIS para, más tarde, generar trayectorias simuladas que se correspondan a las reales.

Dado que se ha optado por un enfoque de aprendizaje supervisado, el proceso de entrenamiento se apoya en datos previamente etiquetados, donde cada trayectoria AIS de un buque registrada incluye el tipo de buque y comportamiento correspondiente. Esta información actúa como referencia durante el aprendizaje del modelo, lo que le permite identificar relaciones y patrones entre las variables disponibles y la clase que se desea imitar.

Asimismo, en el caso de los algoritmos orientados a la asignación de coordenadas y velocidades mediante mapas de calor, los datos AIS se emplearán para la construcción de mapas de densidad basados en la cuadrícula Uber (H3) definida para la zona de estudio. Estos mapas permitirán, por un lado, identificar las celdas con mayor nivel de ocupación y, por lo tanto, aquellas trayectorias más comunes. Por otro lado, estima las velocidades características asociadas a cada celda, proporcionando así una representación espacial y dinámica del comportamiento de los buques en el área analizada.

En el marco de este Trabajo Fin de Grado, los conjuntos de datos utilizados se han construido a partir de mensajes AIS reales transmitidos por buques en el entorno del puerto de Barcelona. Estos mensajes constituyen la fuente principal de información para el análisis del comportamiento de los buques y para el desarrollo de los algoritmos propuestos.

Por otro lado, conviene recordar que el sistema AIS define distintos tipos de mensajes, cada uno asociado a un conjunto específico de información. De entre todos ellos, para los objetivos de este trabajo resulta especialmente relevante un grupo: los mensajes dinámicos. Estos mensajes contienen información dinámica relacionada con la navegación, correspondientes a los tipos 1, 2 y 3, que incluyen datos como la posición, la velocidad o el rumbo del buque.

El conjunto de datos utilizado en este Trabajo Fin de Grado está compuesto por mensajes AIS decodificados y depurados. Estos fueron obtenidos a partir de las transmisiones captadas por las antenas de la Facultad de Náutica de Barcelona (FNB-UPC), perteneciente a la Universidad Politécnica de Cataluña. Los datos, procedentes de transmisiones AIS públicas registradas entre 2020 y 2024, se encuentran organizados por años y meses y distribuidos en archivos CSV, donde cada fichero recoge la información correspondiente a uno o dos días, agrupados en paquetes comprimidos para facilitar su gestión y tratamiento[65].

3.3.1 Selección y análisis de mensajes dinámicos

Como se explicó en capítulos anteriores, los mensajes AIS de tipo 1, 2 y 3 contienen información asociada al comportamiento dinámico del buque. Estos contienen aquellas variables que evolucionan con el tiempo y dependen del estado de navegación en cada instante. Los mensajes dinámicos se caracterizan por una alta frecuencia de transmisión, que habitualmente se sitúa en intervalos de entre dos y diez segundos, lo que permite realizar un seguimiento continuo del movimiento del buque con fines principalmente relacionados con la seguridad marítima. En el Anexo V: ESTRUCTURA DE UN MENSAJE AIS, se puede ver la estructura del mensaje que se recibe en la estación. En el contexto de este Trabajo Fin de Grado, esta información dinámica se emplea como fuente de datos a partir de la cual se entrenarán los algoritmos de aprendizaje automático, con el objetivo de que aprenda a generar trayectorias y comportamientos iguales a los de un buque real.

A partir de los mensajes recibidos y guardados en el mencionado *dataset*, se generan archivos del tipo CSV (Comma-Separated Values) que se encargan de recoger la información de las trazas AIS

transmitidas en cada instante. Estos archivos CSV son el punto de partida para poder tener una fuente de información, entrenamiento y diseño que sea útil para el algoritmo elegido.

Primeramente, se descarga localmente el conjunto completo de datos AIS correspondiente al entorno del puerto de Barcelona para el periodo de estudio comprendido entre 2020 y 2024. Estos datos, registrados por las antenas y sistemas de la Facultad de Náutica de Barcelona, se encuentran originalmente organizados en carpetas mensuales, existiendo una carpeta independiente para cada mes de cada año analizado.

Con el fin de facilitar su gestión y tratamiento posterior, todos los ficheros correspondientes a un mismo año se agrupan en una única carpeta anual. De este modo, se genera una estructura final compuesta por cinco carpetas, una por cada año considerado en el estudio (2020, 2021, 2022, 2023 y 2024). Cada una de estas carpetas contiene la totalidad de los mensajes AIS captados por las antenas del puerto de Barcelona durante el año correspondiente, constituyendo la base de datos inicial empleada en el desarrollo del trabajo.

Estos conjuntos de datos, organizados por años, se importarán al entorno de *Python*, donde se llevará a cabo su análisis y filtrado en función de los requisitos específicos del estudio.

3.3.2 Proceso de filtrado inicial por tipología de buque

Una vez incorporados los conjuntos de datos al entorno de *Python*, se procede a definir y ejecutar una serie de operaciones de filtrado orientadas a eliminar información no relevante y a estructurar los datos de forma adecuada para su análisis posterior. En una primera etapa, se realiza un filtrado inicial con el objetivo de facilitar tanto el tratamiento como el almacenamiento de la información.

Este primer filtrado consiste en clasificar los registros AIS en función del tipo de buque a partir de los mensajes AIS tipo 5, generando conjuntos de datos independientes para cada categoría, como por ejemplo buques de tipo *Cargo*, *Fishing*, que son los dos casos que se van a estudiar en el presente trabajo. Estos dos tipos son los elegidos debido a que los primeros (*Cargo*) presentan un tipo de comportamiento estable en base a las características evolutivas de estos buques, así como siguen las derrotas típicas de aproximación a puerto. El segundo caso (*Fishing*) se caracteriza por presentar movimientos más erráticos y aleatorios en las zonas habituales de pesca en el área de estudio. De este modo, se obtiene un conjunto de archivos CSV específicos para cada tipo de embarcación, lo que permite analizar comportamientos homogéneos dentro de cada grupo.

Una vez realizada esta clasificación por tipo de buque, se introduce una segunda discriminación basada en el sentido de la maniobra, diferenciando entre buques tipo *Cargo* que entran al puerto y barcos pesqueros que faenan en la zona. Esta distinción resulta necesaria para el correcto funcionamiento de los algoritmos de aprendizaje automático y la configuración de los algoritmos basados en mapas de calor, ya que al discriminar entre tipos de buque y su actividad específica, permite a los algoritmos de aprendizaje automático aprender a replicar esos comportamientos específicos de ese tipo de buque y no distraerse con otros comportamientos asociados a esos tipos de barco como salidas de puerto, tránsitos, esperas, etc. La mezcla de estos datos podría dificultar el proceso de aprendizaje o el diseño de los algoritmos de generación de rutas. Como resultado de este segundo filtrado, el conjunto de datos original correspondiente a los buques de tipo *Cargo* y *Fishing* se estructura en: “*Cargo entrada*” y “*Fishing*”.

Los CSV de cada día de registro AIS contienen los mensajes recibidos ordenados por el tiempo de recepción (columna “*receiver timestamp*”) entrelazándose los mensajes de los diferentes barcos presentes en la zona, que pueden identificarse por su MMSI. El identificador del Servicio Móvil Marítimo (MMSI) consiste en un código numérico de nueve cifras que permite distinguir de forma única a cada estación encargada de la transmisión. En primer lugar, para poder procesar los datos, se carga cada CSV diario en un *DataFrame* de la librería Pandas. Adicionalmente se crea otro *DataFrame* auxiliar de dos columnas que relaciona cada MMSI con la clase de barco correspondiente mediante el código de tipo de barco incluido en columna “*ship_type*” en los mensajes tipo 5 y, posteriormente, se

crea una nueva columna “*ship_type_mmsi*” en todas las filas del *DataFrame* principal. Esto se realiza para que los mensajes dinámicos tengan también una columna identificativa del tipo de barco y queden relacionados. A continuación, para agrupar los tipos de barco en categorías más sencillas, se asignan, entre otros, el tipo *Cargo* a los barcos con códigos AIS 50, 52, 54, 70, 71, 72 y 79 y el tipo *Fishing* a los barcos con código 30 y se crea una columna en el *Dataframe* diario (“*ship_group*”) con el nuevo tipo de barco (*Cargo* o *Fishing*) correspondiente.

A continuación, se vuelve a agrupar el *DataFrame* principal por MMSI y se filtran los barcos que pertenezcan a los tipos “*ship_type*” *Cargo* o *Fishing*, objeto de este estudio.

3.3.3 Construcción del dataset “*Cargo entrada*”

En primer lugar, se describe el proceso de generación de los archivos CSV correspondientes a buques de tipo *Cargo* clasificados como “*Cargo entrada*”, seleccionados para este estudio debido a su comportamiento más estable durante la maniobra de aproximación al puerto. Esta mayor estabilidad se debe a las restricciones operativas existentes en la entrada, como los límites de velocidad, la presencia de canales de navegación bien definidos y la frecuente asistencia de prácticos y remolcadores, factores que favorecen trayectorias y velocidades más regulares.

Una vez realizado el filtrado de los buques de tipo *Cargo*, se lleva a cabo un bucle sobre cada uno de ellos con el objetivo de identificar aquellos que realizaron maniobras de entrada al puerto. Como resultado de este proceso, se genera un archivo CSV independiente para cada buque y para cada día registrado, que contiene exclusivamente la información correspondiente a su entrada en la zona portuaria.

Mediante los polígonos empleados en el puerto como filtro, anteriormente explicados en “Definición del área de estudio y criterios de filtrado espacial”, se continúa el cribado y adaptación de la información útil para el proceso. Gracias a estos, podemos tener una recolección de datos mucho más precisa y heterogénea, así como aliviar de carga a los mecanismos de aprendizaje automático empleados.

Asimismo, se excluyen del conjunto de datos los buques de tipo *Cargo* que permanecieron parados, así como aquellos tramos de trayectoria en los que la velocidad fue nula o muy reducida, inferior a 0,5 nudos. Del mismo modo, se descartan los buques que no disponen de trayectorias registradas, aquellos con un número insuficiente de registros (menos de diez) y las trayectorias cuya longitud total no resulta significativa para el estudio, estableciéndose un umbral mínimo de 10 km.

Adicionalmente, se eliminan los registros de posición no válidos, correspondientes a valores inexistentes o erróneos (NaN, Not a Number), con el fin de garantizar la consistencia y calidad de los datos utilizados en el análisis.

Posteriormente, se aplica un proceso de normalización espacial de las trayectorias, mediante el cual cada recorrido se transforma en un conjunto de puntos equidistantes, asociados a los valores de velocidad AIS correspondientes a cada punto. Este procedimiento permite garantizar una representación espacial homogénea de las trayectorias y asegurar que todas ellas disponen de un número mínimo de 100 registros de posición y velocidad. En aquellos casos en los que las trayectorias originales no alcanzan dicho umbral, se procede a su densificación, con el objetivo de proporcionar a las cGAN secuencias de entrada formadas por al menos 100 puntos, como se explicará en el apartado correspondiente.

Como resultado de este proceso, se obtiene para cada trayectoria una lista ordenada de *waypoints* equidistantes que describen el recorrido a seguir, manteniendo fielmente la geometría de la trayectoria original. Cada uno de estos puntos incluye la velocidad AIS asociada, preservando así la información dinámica del movimiento. La distancia entre puntos consecutivos se determina en función de la longitud total de la trayectoria y del número de registros finales establecidos:

$$\Delta d = \frac{\text{distancia total del track}}{\text{número de registros}} \quad \text{Ecuación 1}$$

Este preprocesamiento reduce las irregularidades inherentes al sistema de transmisión AIS, como el muestreo temporal no uniforme y la densidad variable de registros. Este enfoque es coherente con el empleado en otros trabajos de análisis de datos AIS, donde se aplican procedimientos similares de limpieza, interpolación y remuestreo de trayectorias con el objetivo de mejorar la consistencia espacial y temporal de los datos antes de abordar tareas de análisis, clasificación o predicción del comportamiento de los buques [66].

Asimismo, se descartaron las trayectorias de buques de tipo *Cargo* que presentaban giros excesivos o maniobras no representativas de una entrada habitual al puerto, tales como inversiones bruscas de rumbo o giros completos de 360°, con el objetivo de reducir el ruido introducido en los algoritmos de generación de trayectorias sintéticas.

De manera adicional, en aquellos casos en los que un mismo buque de tipo *Cargo* realizaba una entrada y una salida del puerto en el mismo día, se aplicó un filtro temporal para conservar exclusivamente la trayectoria correspondiente a la maniobra de entrada, descartando el resto de los registros.

Por último, todas las trayectorias se limitaron espacialmente a un radio máximo de 30 millas náuticas en torno al puerto de Barcelona, de forma que la generación de trayectorias sintéticas se centrara en las inmediaciones del entorno portuario, garantizando así la coherencia espacial del conjunto de datos empleado.

Como resultado se obtiene un CSV por cada barco tipo *Cargo* que ha realizado una trayectoria de aproximación al puerto de Barcelona. En la Figura 3-10 se muestra un ejemplo de un archivo CSV de un buque tipo “*Cargo* entrada”.

lat	lon	speed	course	receiver	timestamp	mmsi	accuracy	device	msg_type	eta	heading	alt	turn	status	second	ship_type_mmsi	ship_group	course_c	speed_c
4.092.536	1.949.437	15.9	27.7	2020-10-10 22:53:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	27.654.610	15.812.138			
4.092.591.60	19.498.226	15.891.3	27.7	2020-10-10 22:53:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	2.765.441.2	1.581.210.5			
4.092.647.21	19.502.083	15.882.6	27.7	2020-10-10 22:53:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	2.765.421.4	15.812.081			
40.927.028.2	19.505.939	15.873.5	27.7	2020-10-10 22:53:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	27.654.015	1.581.205.2			
4.092.758.43	19.509.796	15.865.3	27.7	2020-10-10 22:53:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	27.653.817	15.812.024			
4.092.814.04	19.513.653	15.856.6	27.7	2020-10-10 22:54:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	27.653.619	1.581.199.4			
4.092.869.65	19.517.509	1.584.75	27.7	2020-10-10 22:54:	209177000	TRUE	Class A	1	30.0	0.0	0.0	20.0	70.0	Cargo	27.653.421	15.811.966			

Figura 3-10 Ejemplo Dataset CSV “Cargo Entrada” (Fuente: propia)

Una vez obtenido el conjunto de datos correspondiente a “*Cargo* entrada”, se continúa con una nueva fase de filtrado orientada a refinar la información disponible. En esta etapa se seleccionan mediante órdenes en el entorno *Python* aquellos registros y variables que nos resultan relevantes para los objetivos planteados en el estudio, descartando aquellos cuya información no aporta valor al proceso de aprendizaje automático o ya han sido filtrados con anterioridad. Este filtrado adicional tiene como finalidad seleccionar las variables más adecuadas para el modelo, evitando la inclusión de datos innecesarios que podrían incrementar la complejidad del sistema y afectar negativamente al rendimiento del algoritmo durante la fase de entrenamiento, confundiéndolo. De este modo, se busca trabajar con un conjunto de datos más compacto y representativo.

En el caso concreto analizado, se han conservado aquellas columnas que permiten caracterizar el comportamiento del buque, su posición geográfica y su velocidad. Estos aspectos son fundamentales para la posterior generación de trayectorias con un comportamiento similar al observado en buques reales del mismo tipo. Los campos seleccionados se detallan a continuación:

1. Latitude (columna “lat”): Latitud en grados (WGS84) de +90(N) a -90(S)
2. Longitude (columna “lon”): Longitud en grados (WGS84) de +180(E) a -180(O)
3. Speed (columna “speed”): Velocidad del buque sobre el fondo en nudos.

Este conjunto de datos será el empleado a lo largo del trabajo para entrenar a los algoritmos, por ello, su importancia es sobresaliente.

3.3.4 Construcción del dataset “Fishing”

A continuación, se aborda el análisis de los datos correspondientes a los buques pesqueros, seleccionados por presentar un patrón de comportamiento claramente diferenciado respecto a las embarcaciones tipo de embarcaciones. A diferencia de los buques mercantes, como se explicó con anterioridad, los pesqueros muestran trayectorias más irregulares y menos estructuradas, asociadas a las labores de pesca y a desplazamientos no lineales en zonas concretas del área de estudio. Estas características hacen necesario un tratamiento específico de los datos AIS, tanto en la fase de filtrado como en el posterior análisis, con el fin de capturar de forma adecuada su comportamiento y facilitar su uso en los algoritmos propuestos en este Trabajo Fin de Grado.

En el caso de los buques pesqueros, clasificados como “Fishing”, el procedimiento seguido es análogo al aplicado previamente a los buques de tipo “Cargo”. No obstante, debido a las particularidades propias de su comportamiento operativo, tanto el proceso de filtrado como determinadas etapas del tratamiento de la información se adaptaron específicamente para este tipo de embarcaciones.

Una vez realizado el filtrado por clase de buque, tal y como se describió en el apartado de filtrado inicial, se obtiene un único conjunto de datos que agrupa todos los mensajes AIS correspondientes a buques de tipo “Fishing”, captados por la estación receptora durante el periodo de estudio comprendido entre 2020 y 2024.

Tras la realización de un primer análisis sobre los datos correspondientes a la clase “Fishing”, se observa que una parte significativa de los buques incluidos en esta categoría presenta un comportamiento similar al de buques mercantes. Este hecho se debe a que muchos de estos buques se encuentran en tránsito hacia sus zonas de pesca, lo que da lugar a trayectorias estables y predecibles, características propias de fases de navegación y no de actividad pesquera. Como consecuencia, el conjunto de datos resultante incluye tanto trayectorias asociadas a maniobras de pesca, con patrones más irregulares, como un número considerable de trayectorias de tránsito, cuyo comportamiento se asemeja al de buques de otras clases, como los de tipo “Cargo”.

Por este motivo, se introduce una capa adicional de filtrado basada en la columna “Status” de los mensajes AIS. El apartado “Status” es un identificador que señala la situación o maniobra que el buque se encuentra realizando en ese momento. Los distintos estados de maniobra en los que se puede encontrar el buque se pueden comprobar en el apartado *Status* del Anexo V: ESTRUCTURA DE UN MENSAJE AIS. En el presente caso seleccionamos el “Status” número 7, referido a aquellas embarcaciones que se encuentran faenando.

Una vez aplicado este nuevo filtrado y regeneradas las trayectorias, tal y como se muestra en la Figura 3-11, se obtiene un resultado claramente mejorado. No obstante, todavía se identifican algunas trayectorias con comportamientos estables similares a los buques tipo *Cargo* y no tan representativas del comportamiento típico del tipo de los pesqueros.

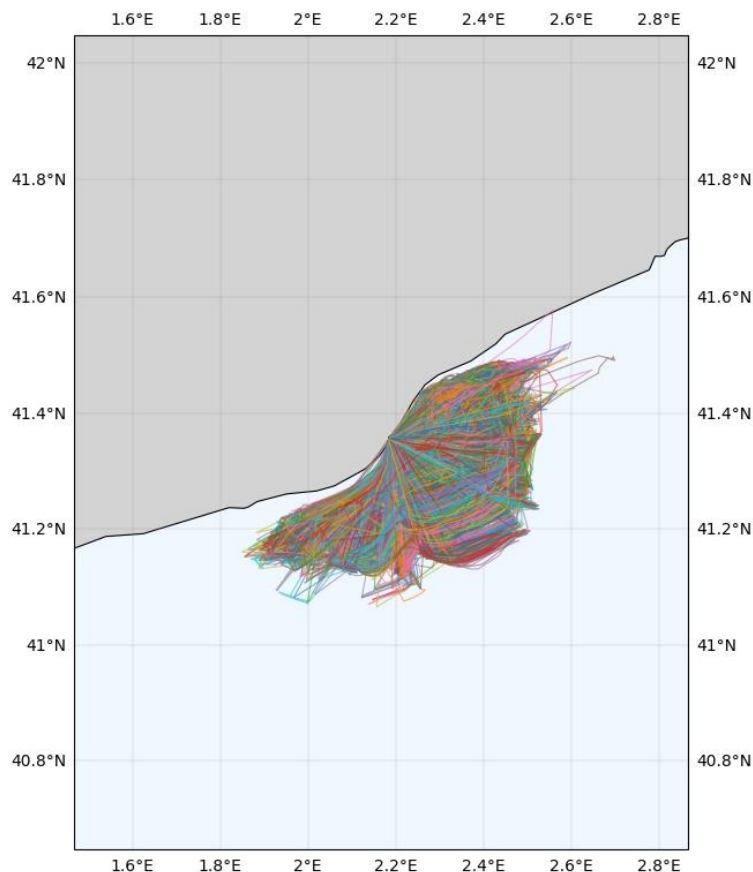


Figura 3-11 Trayectorias filtradas con status 7 (Fuente: propia)

Con el objetivo de identificar comportamientos de pesqueros claramente diferentes a otros tipos de buques, se decide considerar únicamente aquellas trayectorias que entran y salen del puerto de Barcelona. Para ello, se emplean los polígonos de delimitación definidos en apartados anteriores, que permiten identificar de forma precisa las maniobras de acceso y salida del entorno portuario.

Asimismo, con el fin de eliminar trayectorias que aportan ruido y no contribuyen de manera significativa a la calidad del análisis, se establece un umbral mínimo de distancia recorrida por el buque. En el caso de los buques de tipo “*Fishing*”, esta distancia se fija en 5 km, valor inferior al utilizado para los buques de tipo “*Cargo*”, debido a que las embarcaciones pesqueras operan habitualmente a velocidades más reducidas.

3.4 Generación de trayectorias sintéticas mediante discretización espacial H3 y mapas de calor

Los *notebooks* dedicados a la programación del presente modelo se pueden encontrar en el Anexo VI: ACCESO AL CÓDIGO DESARROLLADO.

3.4.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona

3.4.1.1 Visualización y validación de trayectorias reales

Inicialmente y con objeto de visualizar las trayectorias reales filtradas se crea una función de “ploteo” que por un lado permitirá comprobar que las trayectorias seleccionadas son adecuadas y también, representar en forma de series temporales, las variaciones de rumbo y velocidad durante las mismas.

En la Figura 3-12 se puede ver la trayectoria real de un buque del CSV generada en un mapa de proyección Mercator.

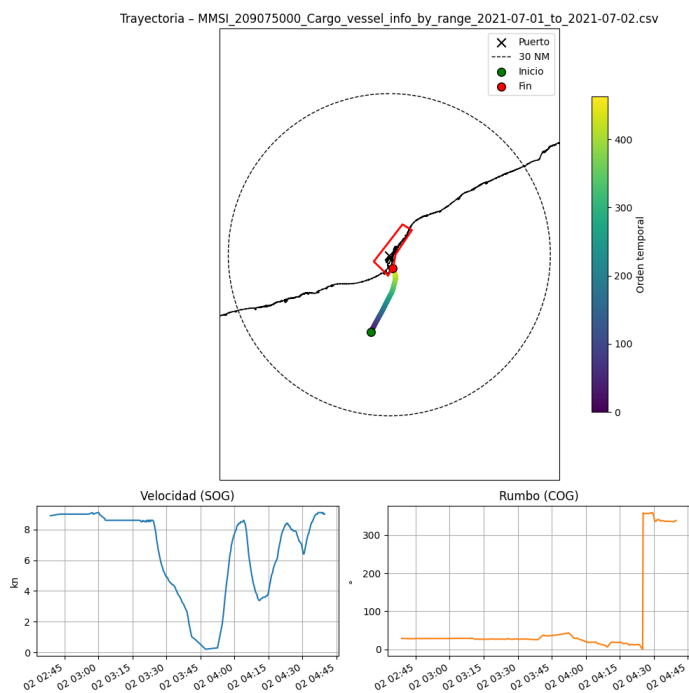


Figura 3-12 Trayectoria real según serie temporal (Fuente: propia)

Asimismo, se implementa una segunda función de visualización destinada a la representación de trayectorias reales de buques incorporando como código de colores la velocidad asociada a cada punto de la derrota. A diferencia de la visualización basada únicamente en el instante de recepción de los mensajes AIS, este enfoque considera la velocidad registrada en cada posición, permitiendo analizar la evolución dinámica del movimiento a lo largo del recorrido.

Esta representación facilita la identificación de zonas de aceleración y desaceleración, proporcionando una visión más completa del comportamiento cinemático durante la navegación. Esto nos servirá además como herramienta para posteriormente poder hacer rápida una comparación visual entre las trayectorias generadas y las reales.

En la Figura 3-13 se puede ver como el buque va desacelerando una vez se acerca al puerto.

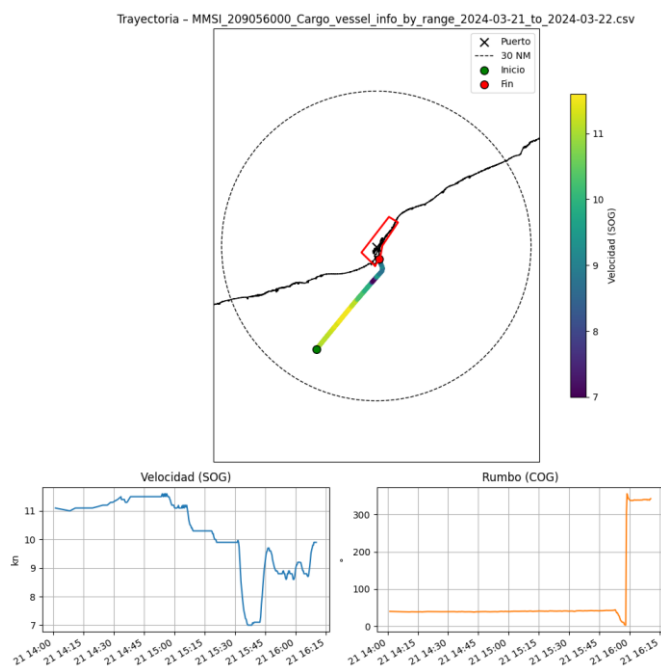


Figura 3-13 Trayectoria real según la velocidad (Fuente: propia)

3.4.1.2 Generación de cuadrículas H3 tipo UBER

El siguiente paso es el de configurar y generar las cuadrículas H3 de tipo UBER para así poder generar los mapas de calor.

En esta etapa se establecen los parámetros globales que regulan el proceso de discretización espacial, así como la delimitación del área de estudio empleada en la generación de trayectorias mediante mapas de calor.

En primer lugar, se fija la resolución del sistema de discretización espacial H3 tipo UBER como se ve en la Figura 3-14. Este parámetro determina el tamaño de las celdas hexagonales empleadas para discretizar el espacio geográfico y, por tanto, el nivel de detalle del mapa de calor generado. En el presente caso, se opta por seleccionar la resolución 9 con el objetivo de proporcionar un equilibrio adecuado entre nivel de detalle espacial y estabilidad estadística de los datos. Este nivel permite representar con suficiente precisión las maniobras en el entorno portuario sin generar una fragmentación excesiva del espacio ni incrementar innecesariamente el coste computacional. En este nivel, cada celda tiene un diámetro aproximado de 300-400 metros.

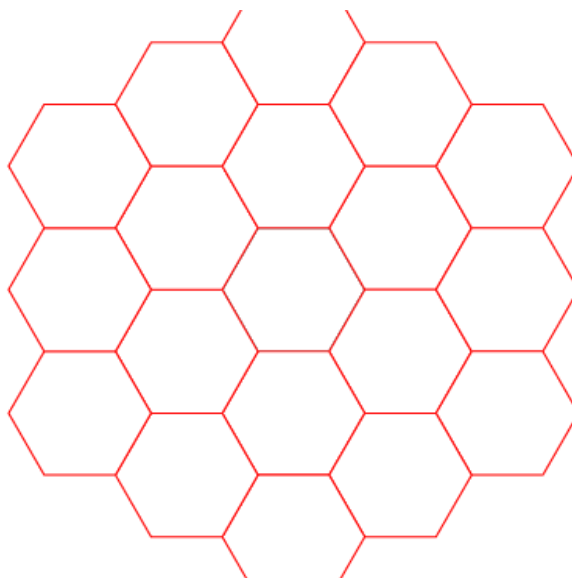


Figura 3-14 Cuadrícula H3 tipo UBER [67]

Una vez definido el tamaño de las celdas, se deben asignar cada una de las posiciones AIS de los buques a las celdas propuestas. Para ello, lo primero que se debe hacer es agrupar todos los datos de latitud y longitud de cada una de las posiciones transmitidas por los barcos recorriendo todos los archivos CSV con esta información para fusionarlos en un único conjunto de datos.

A continuación, cada punto geográfico se asocia a una celda hexagonal del sistema H3 utilizando la resolución previamente establecida. Es decir, si uno de los puntos (latitud y longitud) se encuentra dentro del perímetro de la celda, se contabilizará como que el barco que transmitió ese punto geográfico pasó por esa celda, tal y como se muestra en la Figura 3-15. Este procedimiento permite discretizar el espacio geográfico en unidades espaciales regulares, facilitando la agrupación de observaciones próximas entre sí y reduciendo la complejidad de los datos geoespaciales originales. Posteriormente, se contabiliza el número de puntos asignados a cada celda con el fin de obtener una medida cuantitativa de la densidad en cada celda, que servirá como base para la generación de los correspondientes mapas de calor.

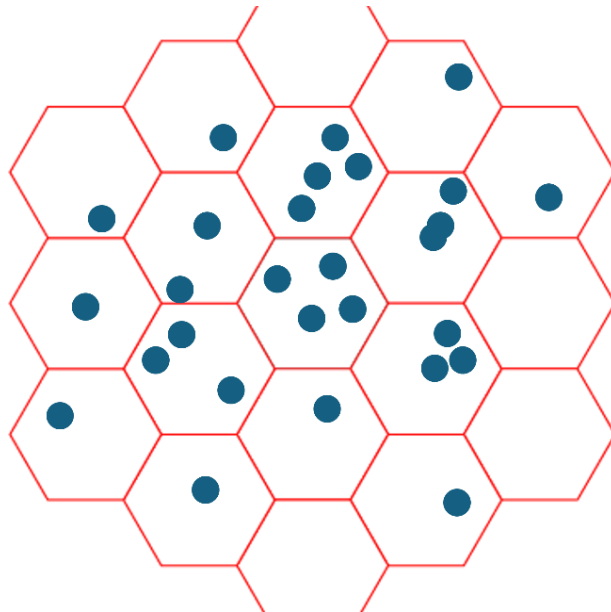


Figura 3-15 Puntos geográficos asociados a celdas (Fuente: propia con [67])

Una vez asignados todos los puntos a las celdas H3 se transforman en polígonos geoespaciales para poder representarlas gráficamente en el mapa. Para cada identificador H3 se recupera el perímetro del hexágono correspondiente, obteniendo así sus coordenadas geográficas. Dichas coordenadas se convierten posteriormente en un objeto poligonal compatible con las librerías geoespaciales empleadas en el análisis.

De forma paralela, se almacena el número de puntos AIS (latitud y longitud) asociados a cada celda, que constituye una medida de la densidad espacial de cada hexágono. Como resultado del proceso, se obtiene un conjunto estructurado de hexágonos junto con sus valores de densidad correspondientes, lo que permite construir posteriormente el mapa de calor utilizado en el análisis del tráfico marítimo.

3.4.1.3 Transiciones entre celdas

Una vez asignado cada punto geográfico a su celda H3 correspondiente y generadas las geometrías hexagonales en un formato compatible con *Python*, es necesario definir cómo se relacionan estas celdas entre sí. En particular, se debe establecer la vecindad entre celdas adyacentes, ya que esta conectividad permite modelar el movimiento entre hexágonos y resulta fundamental para los procesos posteriores de análisis y generación de trayectorias.

Como se dijo anteriormente, el objetivo de este paso es el de capturar patrones de movimiento entre regiones adyacentes, registrando los desplazamientos observados entre hexágonos consecutivos a lo largo de las trayectorias AIS. Este bloque de código tiene como finalidad analizar cómo se desplazan los buques entre las distintas celdas H3 a lo largo de sus trayectorias. En primer lugar, cada posición AIS se transforma en su celda hexagonal correspondiente, obteniendo así una secuencia ordenada de celdas que describe el recorrido del buque de forma discretizada.

A continuación, se estudian las celdas consecutivas para identificar cuándo el buque pasa de un hexágono a otro, ignorando los casos en los que permanece dentro de la misma celda. Además, únicamente se consideran válidas aquellas transiciones que se producen entre celdas vecinas, descartando posibles saltos espaciales no coherentes. Cada vez que se detecta un cambio válido entre dos celdas, se incrementa un contador que registra la frecuencia de dicho movimiento. De este modo, se construye una estructura que refleja los desplazamientos más habituales entre hexágonos, lo que puede interpretarse como un modelo de transición espacial similar a una cadena de Markov, donde la probabilidad de moverse a una celda determinada depende únicamente de la celda actual.

Esta información constituye la base para la posterior generación de trayectorias sintéticas coherentes con los patrones reales observados en los datos AIS. Este modelo de transiciones se ve más claro en la Figura 3-16, donde se resalta el lado por el cual se conecta una celda con otra.

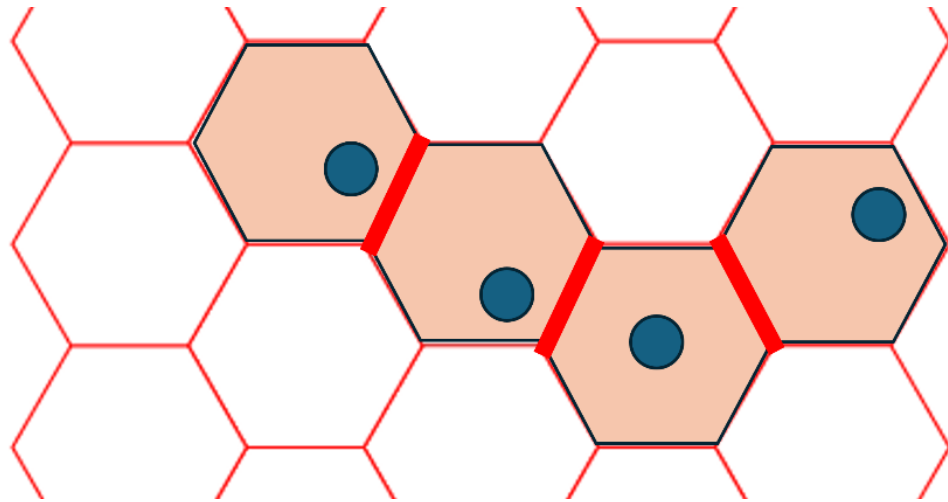


Figura 3-16 Sistema de transiciones entre celdas vecinas (Fuente: propia)

El resultado de este proceso es una estructura de datos que asocia a cada celda H3 un conjunto de celdas vecinas de destino junto con sus frecuencias de transición. Este modelo de transiciones constituye la base para la generación posterior de trayectorias sintéticas basadas en mapas de calor, ya que permite simular desplazamientos coherentes con los patrones de navegación observados en los datos reales.

3.4.1.4 Establecimiento de celdas objetivo y celdas iniciales

A la hora de comenzar a generar trayectorias sintéticas, es necesario establecer unos límites a estas rutas, para que el sistema sea capaz de saber dónde deben comenzar y acabar sus trayectorias sintéticas generadas y sean coherentes con el comportamiento observado en los datos reales.

Para que el sistema sepa dónde deben finalizar sus trayectorias, debemos definirlas. Para ello, se identifican las celdas H3 que corresponden a la zona de llegada al puerto de Barcelona, que actuará como región objetivo para la generación de trayectorias sintéticas. Para lograr esto, y dado el conjunto de celdas H3, se obtiene el punto central de cada celda y se comprueba si dicho punto se encuentra contenido dentro del polígono que define la zona de acceso al puerto. En el caso estudiado, el puerto de Barcelona cuenta con un total de 74 celdas objetivo. El conjunto resultante (74 hexágonos) define las celdas objetivo, es decir, las celdas que representan una llegada válida al puerto, tal y como se muestra en la Figura 3-17.

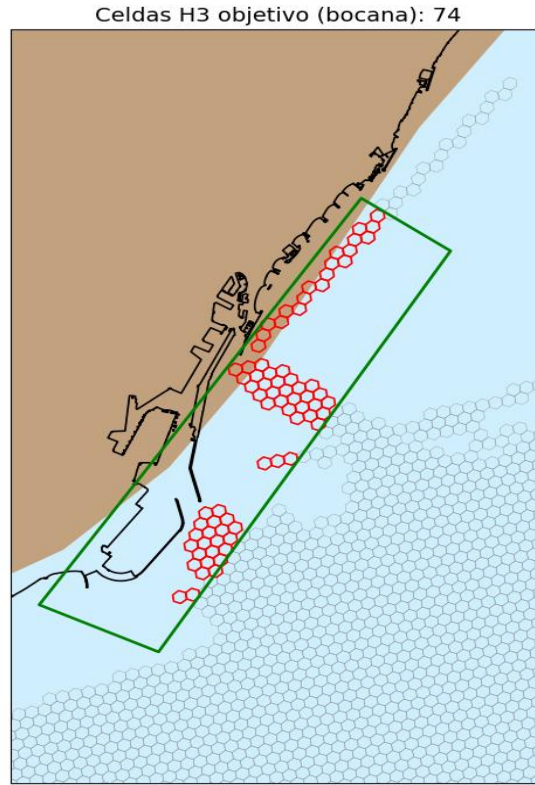


Figura 3-17 Celdas objetivo de buques tipo *Cargo* (Fuente: propia)

Este conjunto de celdas se utiliza posteriormente como condición de destino en el proceso de generación de trayectorias basadas en mapas de calor, garantizando que las trayectorias sintéticas finalicen en una región geográficamente coherente (Puerto de Barcelona) con las llegadas reales observadas.

Por otro lado, con el objetivo de saber dónde deben comenzar las trayectorias, se definen las celdas H3 que pueden actuar como puntos de inicio. Para ello, se implementa una función auxiliar que calcula la distancia entre una celda H3 y el puerto de Barcelona. La distancia se obtiene calculando la posición geográfica asociada al centro de la celda y el puerto. A la hora de calcular la distancia entre dos puntos geográficos debemos tener en cuenta que la Tierra no es plana y que, por lo tanto, debemos tener en cuenta la curvatura de esta. Para ello se emplea en la función la fórmula de Haversine:

Donde:

- φ_1, φ_2 : latitudes de los dos puntos (en radianes)
- λ_1, λ_2 : longitudes de los dos puntos (en radianes)
- R : radio medio de la Tierra ($\approx 6.371.000$ m).

Se define:

$$\Delta\varphi = \varphi_2 - \varphi_1 \quad \text{Ecuación 2}$$

$$\Delta\lambda = \lambda_2 - \lambda_1 \quad \text{Ecuación 3}$$

Siendo la distancia entre dos puntos:

$$d = 2R \arcsin \left(\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)} \right) \quad \text{Ecuación 4}$$

Este cálculo permite expresar la distancia entre dos puntos geográficos en metros, de manera consistente con el tratamiento geodésico aplicado en el resto del análisis, garantizando así la coherencia espacial de los resultados obtenidos.

A continuación, se identifican como celdas iniciales aquellas celdas H3 que actúan como origen en el modelo de transiciones espaciales de las trayectorias reales y cuya distancia al puerto es de mínimo de 15 millas náuticas. Este criterio permite evitar que las trayectorias sintéticas se inicien en zonas demasiado próximas al puerto, garantizando que las trayectorias no sean demasiado cortas. De este modo, se favorece la generación de rutas más realistas y coherentes con los patrones de navegación observados en los datos reales.

3.4.1.5 Aplicación de mejoras de la trayectoria

El siguiente paso es definir el mecanismo de selección de la siguiente celda H3 durante la generación de trayectorias sintéticas, combinando la información estadística obtenida del modelo de transiciones con restricciones geométricas y criterios direccionales orientados a que el movimiento sea lo más realista posible.

En primer lugar, se define una función auxiliar destinada a calcular el ángulo formado por tres puntos consecutivos en el plano. Esta función permite cuantificar el cambio de dirección entre dos segmentos sucesivos de una trayectoria, proporcionando una medida del giro realizado por el buque. Posteriormente, este valor se emplea como criterio de control para penalizar desviaciones excesivamente bruscas, que no resultan realistas en el contexto de la navegación marítima y podrían introducir comportamientos poco realistas en las trayectorias sintéticas generadas. En el presente caso se ha decidido descartar todos aquellos cambios de rumbo superiores a 80° . Este valor permite conservar maniobras habituales de aproximación a puerto, caracterizadas por cambios de rumbo progresivos, y descartar transiciones que impliquen variaciones excesivamente bruscas o no coherentes con el comportamiento dinámico de buques mercantes de gran tamaño.

A continuación, se implementa una función que determina la celda H3 siguiente a partir de la celda actual y del modelo de transiciones previamente aprendido. La función parte de las celdas destino observadas en los datos históricos y de sus frecuencias de transición, que actúan como peso base. Es decir, cada posible celda destino tiene inicialmente un peso proporcional al número de veces que esa transición se ha producido en los datos reales. Cuanto más frecuente ha sido ese movimiento en las trayectorias AIS observadas, mayor probabilidad tendrá esa celda de ser seleccionada.

Sin embargo, el peso final no depende únicamente de la frecuencia histórica. Este valor se ajusta mediante un factor adicional que favorece aquellas celdas que reducen la distancia al puerto, reforzando así el comportamiento de aproximación progresiva. De este modo, se combinan los patrones reales observados con un criterio físico coherente con la maniobra de entrada.

En el proceso el peso se penaliza en dos situaciones:

- Cuando la celda candidata coincide con la celda anterior (para evitar retrocesos o movimientos oscilatorios).
- Cuando el cambio de dirección respecto al movimiento previo es excesivamente brusco, como se dijo anteriormente (cambios de rumbo mayores de 80°).

Estas penalizaciones no eliminan completamente la transición, pero reducen su probabilidad, favoreciendo trayectorias más suaves y realistas.

Finalmente, los pesos resultantes se normalizan para obtener una distribución de probabilidad válida y se selecciona de forma estocástica la siguiente celda H3.

3.4.1.6 Generador de trayectorias basado en modelo Markoviano

Una vez definidos todos los elementos que condicionan el comportamiento de las trayectorias se procede a la fase de generación de trayectorias sintéticas.

En este punto, el modelo ya dispone de la información necesaria para reproducir de forma estructurada los patrones de movimiento observados en los datos AIS reales. El espacio ha sido discretizado, las transiciones entre regiones han sido cuantificadas y se han incorporado restricciones geométricas y dinámicas que garantizan un comportamiento coherente con la navegación real de los buques. Por tanto, el siguiente paso consiste en transformar esta información estadística y espacial en recorridos completos que simulen una maniobra real de entrada en puerto.

El generador de trayectorias basado en H3 funciona de forma similar a un modelo de tipo Markoviano de primer orden. Esto significa que, en cada paso, la decisión de a qué celda se mueve el buque depende únicamente de la celda en la que se encuentra en ese momento y de las transiciones que se han observado desde esa celda en los datos reales. Es decir, el modelo no analiza toda la trayectoria anterior, sino que toma la decisión basándose en la posición actual y en lo que históricamente ha ocurrido desde esa zona.

En la práctica, esto implica que, si desde una determinada celda los buques reales han tendido a moverse hacia ciertas celdas vecinas con mayor frecuencia, el generador tendrá mayor probabilidad de seleccionar esas mismas direcciones. De este modo, el comportamiento estadístico observado en los datos AIS se traslada directamente al proceso de generación. Adicionalmente, se incorporan factores correctores, como la penalización de retrocesos inmediatos y de giros excesivamente bruscos, con el objetivo de mejorar la coherencia geométrica de las trayectorias generadas.

La función para generar la trayectoria sintética de entrada en el puerto de Barcelona recibe como entrada:

- Celda inicial
- Modelo de transiciones aprendido a partir de datos reales
- Conjunto de celdas objetivo
- Coordenadas del Puerto de Barcelona

A partir de estos elementos, el proceso de generación se realiza de forma iterativa, construyendo paso a paso una secuencia de celdas H3 que representa una trayectoria sintética completa. En cada iteración se selecciona la siguiente celda en función del modelo de transiciones definido previamente, hasta alcanzar una celda objetivo, que está asociada a la llegada al puerto o cumplirse las condiciones de parada establecidas. En el presente caso se ha establecido un máximo de 1000 celdas para lograr llegar a puerto, en caso de que se sobrepase se descartará la trayectoria.

Durante cada paso del proceso, el algoritmo analiza la situación actual antes de continuar. Si la celda en la que se encuentra ya forma parte de la zona objetivo, la trayectoria se considera finalizada. Si, por el contrario, desde esa posición no existen transiciones registradas en el modelo aprendido, la generación también se detiene para evitar construir recorridos sin sentido.

Cuando ninguna de estas situaciones ocurre, el sistema calcula cuál será la siguiente celda. Esta decisión, como se dijo anteriormente, no es determinista, sino probabilística: se basa en las transiciones observadas en los datos reales, pero incorpora además una ligera tendencia a desplazarse en dirección al puerto, garantizando así que la trayectoria evolucione de manera coherente hacia su destino.

Así mismo, si desde la celda donde se encuentre existe la posibilidad de moverse directamente a una celda objetivo (área del puerto), el algoritmo incorpora una probabilidad del 40% de moverse a ella. Se ha elegido una probabilidad de 0.4 debido a que, si la probabilidad fuese demasiado alta, el generador forzaría casi siempre la llegada a puerto y, por el contrario, si fuese demasiado baja, generaría recorridos innecesariamente largos. Ambos casos favorecerían trayectorias poco realistas y naturales. Con ello, la

probabilidad del 40%, favorece que la trayectoria alcance el puerto de forma más eficiente, sin imponer un comportamiento completamente determinista. De esta manera, se acelera la llegada al destino cuando es coherente hacerlo, pero se mantiene cierta variabilidad en el recorrido generado.

Como salida del algoritmo, se obtiene una ruta compuesta por una sucesión de hexágonos H3 enlazados entre sí. Esta secuencia representa el recorrido simulado desde mar abierto hasta el entorno portuario, siguiendo la lógica espacial aprendida del tráfico real tal y como se muestra en la Figura 3-18.

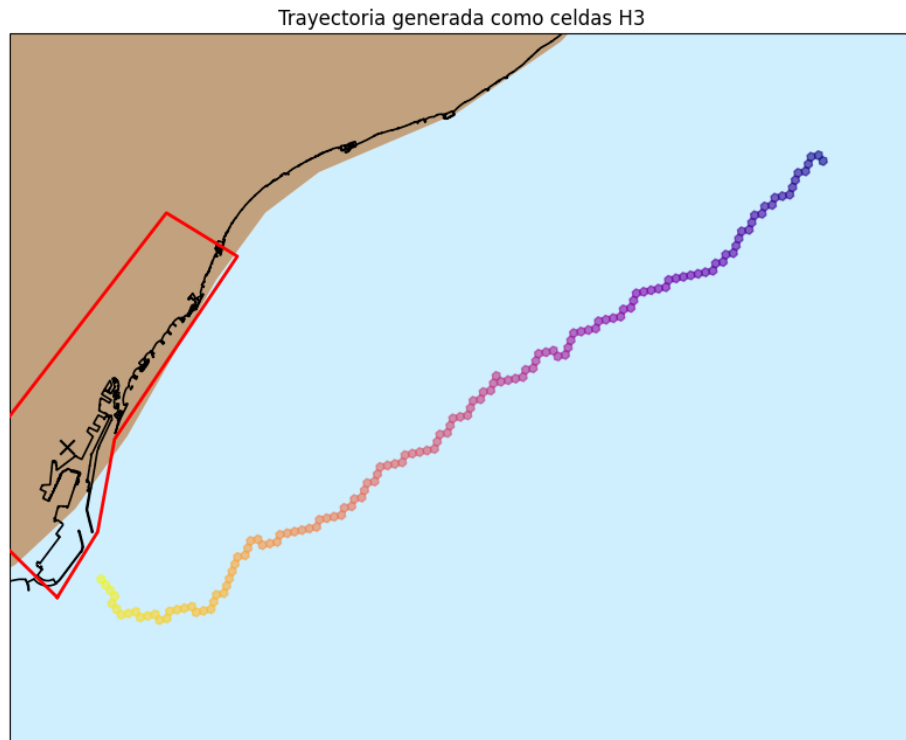


Figura 3-18 Trayectoria sintética con celdas tipo UBER (Fuente: propia)

Aunque la trayectoria está definida inicialmente en términos de celdas, se convertirá posteriormente en posiciones geográficas (latitud y longitud), lo que permite dibujarla sobre el mapa y compararla directamente con las trayectorias AIS originales.

3.4.1.7 Transformación de celdas a posiciones geográficas

Para la transformación en posiciones geográficas, se crea una función que recibe la secuencia de celdas tipo UBER y selecciona la posición geográfica basadas en el centro de estas, expresadas en latitud y longitud. Así mismo, con el objetivo de evitar trayectorias excesivamente rígidas, se le aplica a la función una pequeña perturbación, denominada “jitter”, que se encarga de desplazar ligeramente el punto seleccionado en el centro de la celda. En este caso, la perturbación es de 0,3. Cabe destacar que la perturbación aleatoria se aplica únicamente con fines de suavizado visual y no altera la estructura discreta de la trayectoria generada

El resultado final es una trayectoria en coordenadas geográficas continuas que conserva la estructura global del recorrido generado en el espacio H3, pero presenta una apariencia más natural y comparable a las trayectorias reales observadas en los datos AIS, tal y como se presenta en la Figura 3-19.

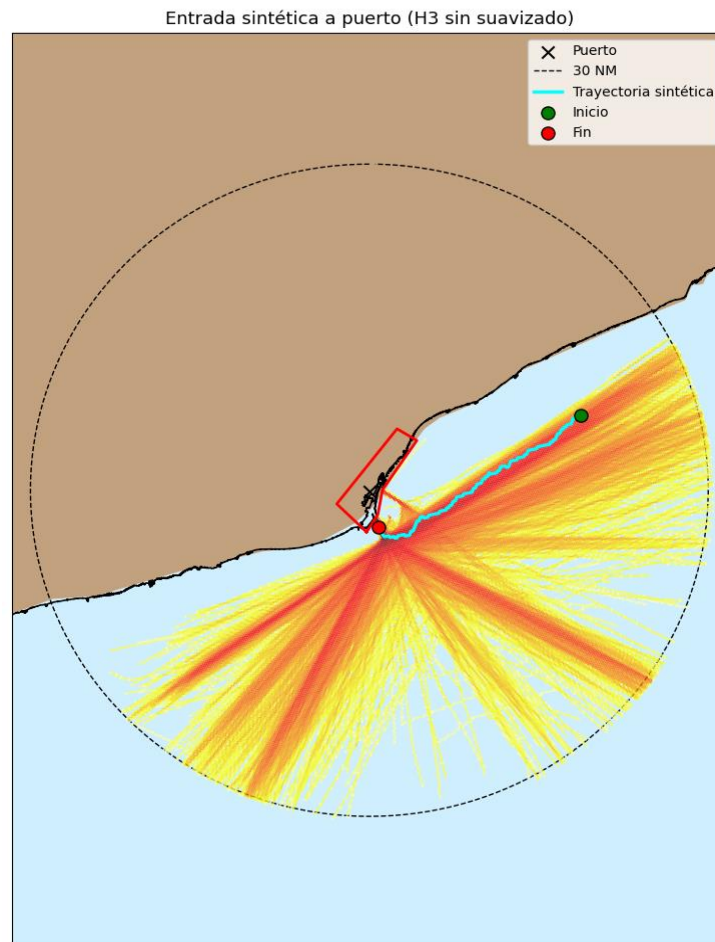


Figura 3-19 Trayectoria sintética sin suavizar (Fuente: propia)

3.4.1.8 Suavizado de la trayectoria

Aunque una vez transformada la trayectoria de celdas a puntos geográficos ya se hubiese aplicado un primer suavizado mediante el “*jitter*”, se comprobó mediante la creación de distintas trayectorias sintéticas, que estas presentaban irregularidades que las diferenciaban claramente de las reales, tal y como se muestra en la Figura 3-19.

Debido a esto, se vio necesaria la creación de un nuevo método complementario de suavizado de estas trayectorias en coordenadas geográficas, con el fin de obtener recorridos continuos que no difiriesen tanto de las reales.

El método de suavizado recibe como entrada la secuencia de la trayectoria de puntos en latitud y longitud. Si el número de puntos es menor que 5, la trayectoria se devuelve sin modificar, ya que la interpolación requiere un mínimo de 5 para garantizar estabilidad numérica.

Cuando la trayectoria contiene un número suficiente de puntos, se aplica una interpolación *spline* paramétrica para modelar conjuntamente la latitud y la longitud como funciones de un mismo parámetro continuo. De este modo, la trayectoria se interpreta como una curva en el plano y no como una sucesión independiente de puntos, lo que permite obtener un recorrido más suave y continuo. El parámetro de suavizado controla el grado de ajuste de la *spline*, permitiendo eliminar pequeñas irregularidades derivadas de la discretización espacial del modelo H3 sin modificar de forma significativa la geometría global del recorrido.

El parámetro de suavizado, fijado en $s = 0.002$, regula cuánto puede separarse la curva ajustada de los puntos originales. Valores más altos producirían una trayectoria excesivamente redondeada, alejándose de la geometría generada por el modelo H3, mientras que valores demasiado pequeños apenas

corregirían las pequeñas discontinuidades entre celdas consecutivas. El valor seleccionado representa un compromiso práctico: permite eliminar irregularidades derivadas de la discretización espacial, sin modificar de forma significativa la forma global del recorrido ni alterar el origen y destino de la trayectoria.

A partir de la *spline* ajustada se generan 200 puntos igualmente espaciados, lo que permite estandarizar la longitud de todas las trayectorias. Finalmente, se fuerzan explícitamente los puntos inicial y final de la trayectoria suavizada para que coincidan exactamente con los puntos originales, garantizando la coherencia con las condiciones de inicio y llegada. En la Figura 3-20 se puede ver una trayectoria una vez suavizada.

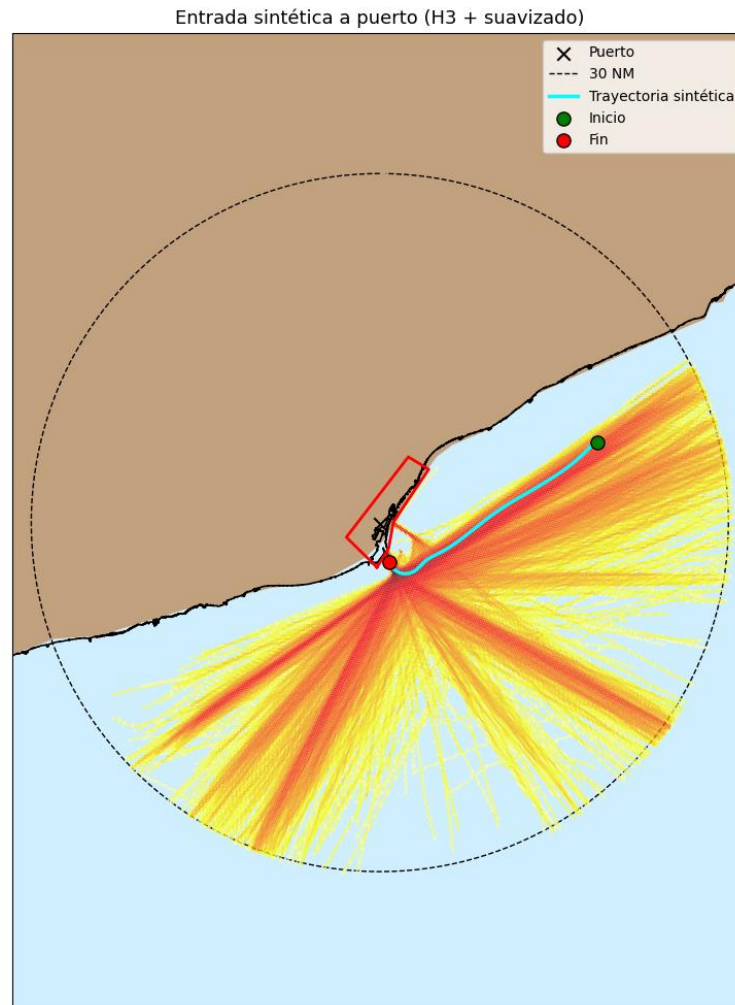


Figura 3-20 Trayectoria sintética suavizada (Fuente: propia)

3.4.1.9 Generación de trayectoria final sin velocidad

Tras aplicar el conjunto de funciones y criterios definidos, el sistema permite generar trayectorias sintéticas de buques tipo “*Cargo*” entrando en el puerto de Barcelona, manteniendo la estructura espacial y el comportamiento dinámico característico de las trayectorias reales observadas en el área de estudio. Se muestra un ejemplo en la Figura 3-21.

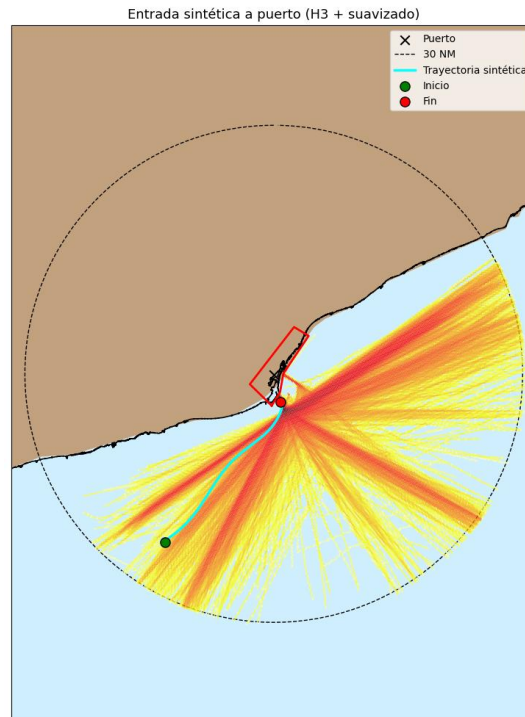


Figura 3-21 Ejemplo trayectoria sintética mediante mapas de calor y cuadrículas tipo UBER (Fuente: propia)

Con el objetivo de comprobar el correcto funcionamiento del algoritmo se le manda generar 400 trayectorias sintéticas, para analizar si son coherentes con la realidad. Esto se muestra en la Figura 3-22.

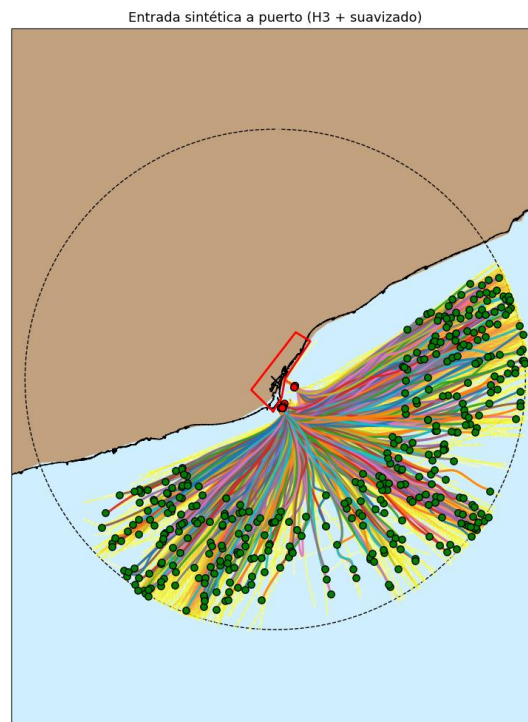


Figura 3-22 Generación de 400 trayectorias sintéticas (Fuente: propia)

3.4.1.10 Construcción de mapas de calor y asignación de velocidades

Una vez obtenidas las trayectorias que siguen los buques cargueros sintéticos generados entrando en puerto, se debe aumentar la información. Para ello, en el presente apartado se abordará cómo generar mapas de calor para un estudio más eficiente y como otorgarles velocidad a las trayectorias generadas.

Para lograr asignarle velocidad a los puntos de la ruta, se construye una estructura que asocia a cada celda H3 las velocidades reales observadas en esa zona a partir de los datos reales AIS. Para ello, recorre todos los archivos CSV quedándose con los valores de *speed* (velocidad sobre el fondo). Así mismo filtra valores de velocidad no válidos o poco representativos. Para ello se le aplica una serie de órdenes a la función para que los descarte:

- Eliminación de valores no válido (NaN)
- Eliminación de velocidades menores que 0,5 nudos, ya que menos que esa velocidad significaría que el buque está prácticamente parado o fondeado, hecho que no se busca en este estudio.
- Eliminación de velocidades mayores que 30 nudos, ya que son excesivamente altas para el modo de operación de un carguero en la zona y podrían deberse a errores de transmisión o registros anómalos.

Realizado el filtrado de velocidades, convierte cada posición geográfica en su correspondiente celda H3. De este modo, se almacena para cada hexágono la lista de velocidades registradas en él. El resultado es un listado que asocia a cada celda H3 un conjunto de valores de velocidad recibidos, lo que permite caracterizar el comportamiento cinemático típico en cada región del espacio. A continuación, se calculan estadísticas descriptivas de velocidad para cada celda H3 a partir de las velocidades previamente almacenadas.

Primeramente, se recorre el listado de velocidades previamente almacenado. Para cada celda, se convierte la lista en un array numérico y se comprueba que existan al menos 5 observaciones. Este umbral mínimo se introduce para evitar calcular estadísticas poco representativas cuando hay muy pocos datos disponibles. Con este conjunto de datos por celda, se realiza un estudio estadístico calculándose los siguientes parámetros:

Parámetro	Descripción
Mediana	Representa la velocidad central de la distribución, es decir, el valor que divide los datos en dos mitades iguales. Es menos sensible a valores extremos que la media
Percentil 25 (p25)	Indica el valor por debajo del cual se encuentra el 25 % de las velocidades observadas.
Percentil 75 (p75)	Marca el valor por debajo del cual se encuentra el 75 % de las observaciones. Junto con el percentil 25 permite estimar la dispersión central.
Media	Es el promedio aritmético de las velocidades registradas en esa celda.
Desviación estándar	Es el promedio aritmético de las velocidades registradas en esa celda.
Nº total de observaciones	Indica cuántos datos reales de velocidad se han utilizado para calcular esas estadísticas.

Tabla 3-1 Parámetros estadísticos

El listado de parámetros estadísticos calculados se almacena de nuevo en un listado por hexágono. El conjunto resultante de estadísticas por celda proporciona una descripción compacta de la distribución

de velocidades observadas en cada región, lo que permite asignar posteriormente velocidades más realistas a las trayectorias sintéticas.

A este registro de velocidades se le vuelve a aplicar un nuevo filtro mediante la aplicación del percentil 5 y percentil 95. Esto, aplicado al listado, significa que se descarta el 5 % de valores más bajos y el 5 % de valores más altos. Es una forma de eliminar valores extremos que podrían distorsionar el análisis o la visualización.

Una vez hecho esto, y con todo el registro de velocidades por celdas, se calcula la velocidad máxima, mínima y media en el total de las celdas. Esto será útil a la hora de representar el mapa de calor de velocidades.

Velocidad	Resultado (en nudos)
Velocidad mínima	0,7
Velocidad media	13,945390791744781
Velocidad máxima	23,92679060420812

Tabla 3-2 Velocidades calculadas

Fijados estos valores, el siguiente paso es representar el mapa de calor espacial de la velocidad media asociada a las celdas H3 en el área de estudio. Para ello cada celda H3 se representa mediante un polígono hexagonal coloreado en función de la velocidad característica observada históricamente en dicha celda. El valor utilizado para la codificación cromática es la mediana de la velocidad, lo que permite una representación robusta frente a valores extremos.

Finalmente, se incluye una barra de color que indica la correspondencia entre los colores y los valores de velocidad, permitiendo interpretar de forma cuantitativa los patrones espaciales observados. El resultado permite identificar de forma clara las zonas donde los buques tienden a navegar a mayor o menor velocidad en la aproximación al puerto. El resultado se puede observar en la Figura 3-23.

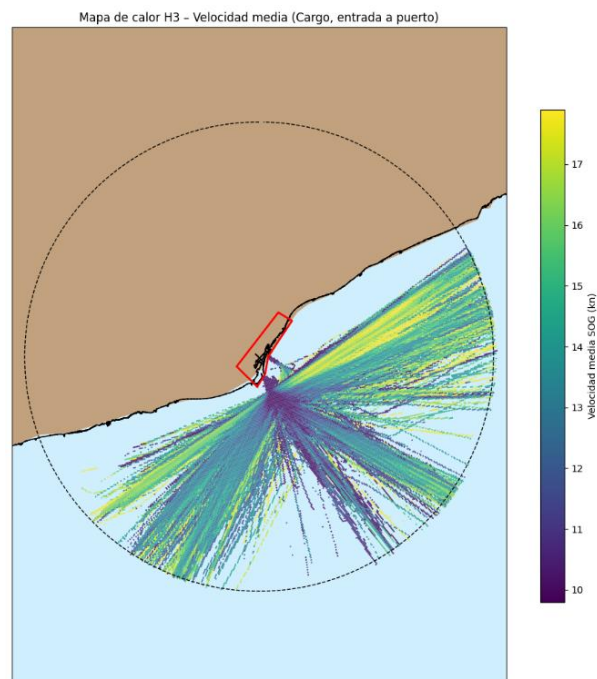


Figura 3-23 Mapa de calor según velocidad (Fuente: propia)

Una vez realizados los correspondientes estudios sobre la velocidad, es el momento de generar velocidades sintéticas que asignar a cada tramo de las derrotas simuladas.

Para ello, se comprueba si existen datos estadísticos asociados en el listado de velocidades estadísticas filtrado de cada celda. En caso afirmativo, se genera una velocidad aleatoria siguiendo una distribución normal con media y desviación estándar iguales a las observadas en esa celda. De este modo, la velocidad asignada no es fija, sino que reproduce la variabilidad real registrada en esa zona.

En caso de que una celda no disponga de estadísticas suficientes, se le asigna una velocidad por defecto, marcada en 10 nudos. Posteriormente, se aplican tres controles:

- Velocidad mínima: Marcada en 0,7 nudos, basándonos en las velocidades calculadas en el apartado anterior donde la velocidad mínima estadística de todas las celdas era de 0,7 nudos.
- Velocidad máxima: Marcada en 24 nudos, basándose en las velocidades calculadas en el apartado anterior donde la velocidad máxima estadística es de 23,96 nudos.
- Evitar que la velocidad sea negativa forzando que esta sea mayor que 0.

En la generación de estas trayectorias con velocidad, se observa que el cambio de velocidades sintéticas entre celdas es demasiado abrupto, diferenciándose considerablemente al comportamiento de un carguero real.

Para evitar esto, se le aplica al resultado un suavizado al perfil de velocidades asociado a la trayectoria sintética. Para ello, se aplica un filtro de Savitzky-Golay con el objetivo de eliminar fluctuaciones abruptas no realistas, manteniendo al mismo tiempo la tendencia global del perfil y que es adecuado cuando los datos varían rápidamente [85]. En las Figura 3-24 y Figura 3-25, se puede observar una comparación entre las velocidades durante la trayectoria obtenida sin suavizar y suavizadas.

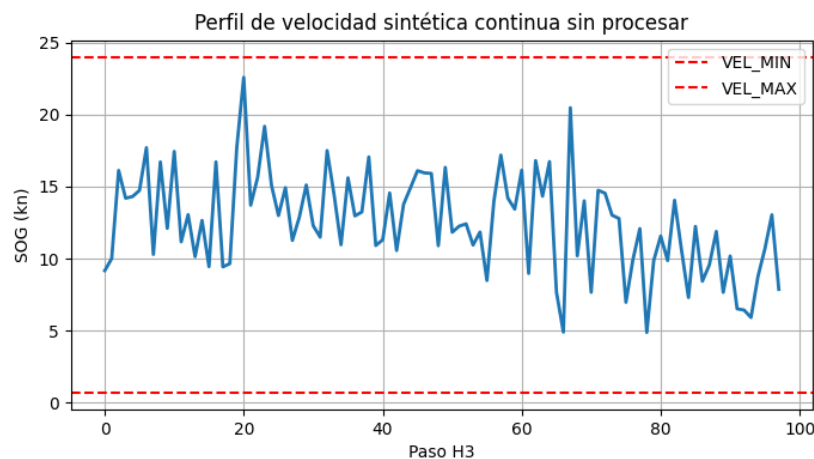


Figura 3-24 Trayectoria con velocidad sin suavizar (Fuente: propia)

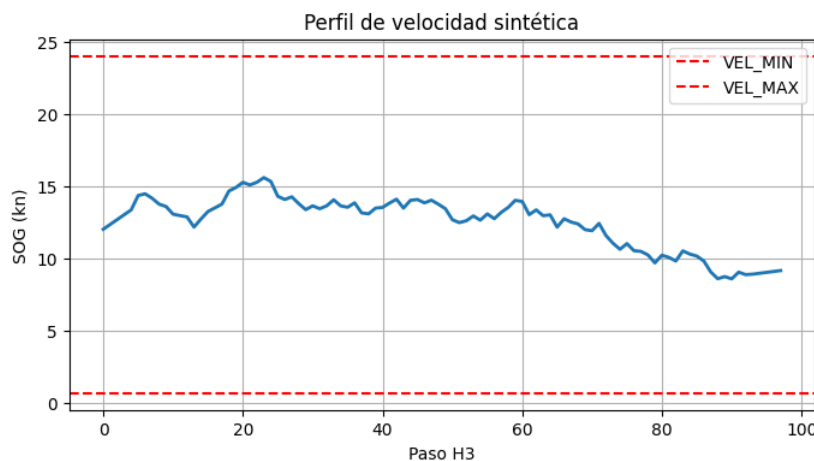


Figura 3-25 Trayectoria con velocidad suavizada (Fuente: propia)

En la segunda tabla (Figura 3-25) se observa como el suavizado ha funcionado, pero, como se puede observar la velocidad va disminuyendo de forma gradual y estable una vez se va acercando al puerto. Este comportamiento es irreal en comparación a un buque real, donde no están continuamente variando la velocidad, si no que la mantienen durante determinados periodos de tiempo.

Como resultado, se obtiene una secuencia de velocidades asociada a cada punto de la trayectoria, coherente con los patrones dinámicos observados en los datos reales. Este perfil permite dotar a la trayectoria sintética de información cinemática, facilitando su análisis desde el punto de vista del comportamiento del buque y posibilitando comparaciones directas tanto con trayectorias reales como con aquellas generadas mediante otros métodos.

Para ello, se define una función para discretizar el perfil de velocidades generado. Lo que hace es redondear cada valor de velocidad al múltiplo más cercano. Por ejemplo, una velocidad de 12,3 nudos pasaría a ser una velocidad de 12. En caso de que la velocidad sea finalizada en .5 se mantendrá la misma velocidad, por ejemplo 13,5 seguiría valiendo 13,5. De esta manera, se obtiene una velocidad discretizada con la resolución deseada (cambios de 0,5 nudos), haciendo que los cambios de velocidad no sean tan continuos y, por lo tanto, acercándose más al comportamiento de un buque carguero real. Esta función se ve reflejada en la Figura 3-26, donde se observa un perfil de velocidades sintético mucho más parecido al comportamiento real.

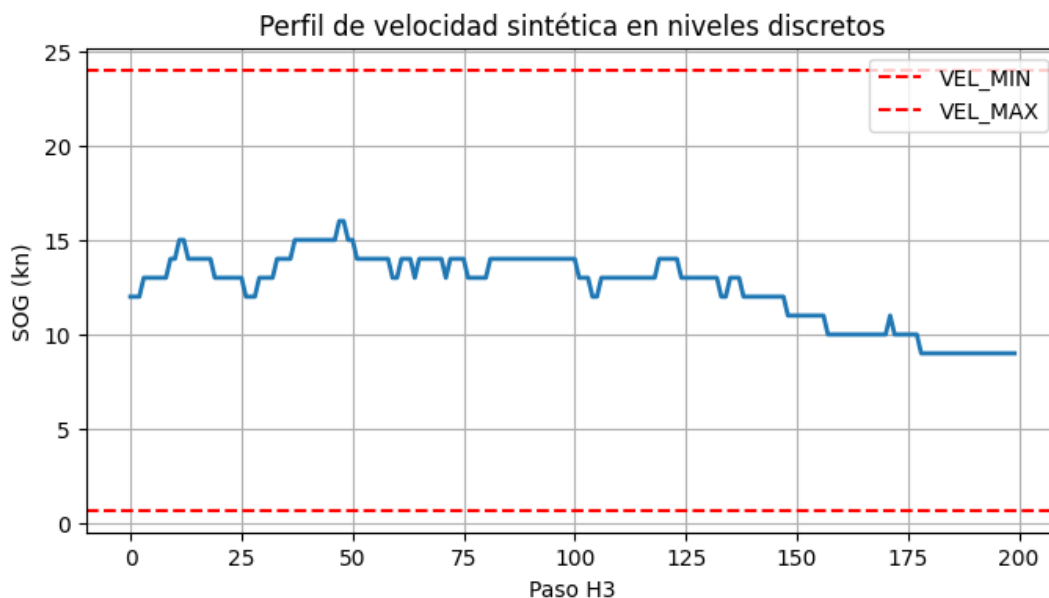


Figura 3-26 Perfil de velocidad sintético discretizado (Fuente: propia)

Una vez aplicado este último discretizado, ya podemos obtener las trayectorias sintéticas a partir de mapas de calor y celdas tipo UBER. Obteniéndose con cada una de las trayectorias sintéticas generadas una tabla con el perfil de velocidades y perfil de rumbos asociados a la trayectoria para facilitar la comparación posterior. En la Figura 3-27 se muestra el resultado de una trayectoria sintética generada mediante este método.

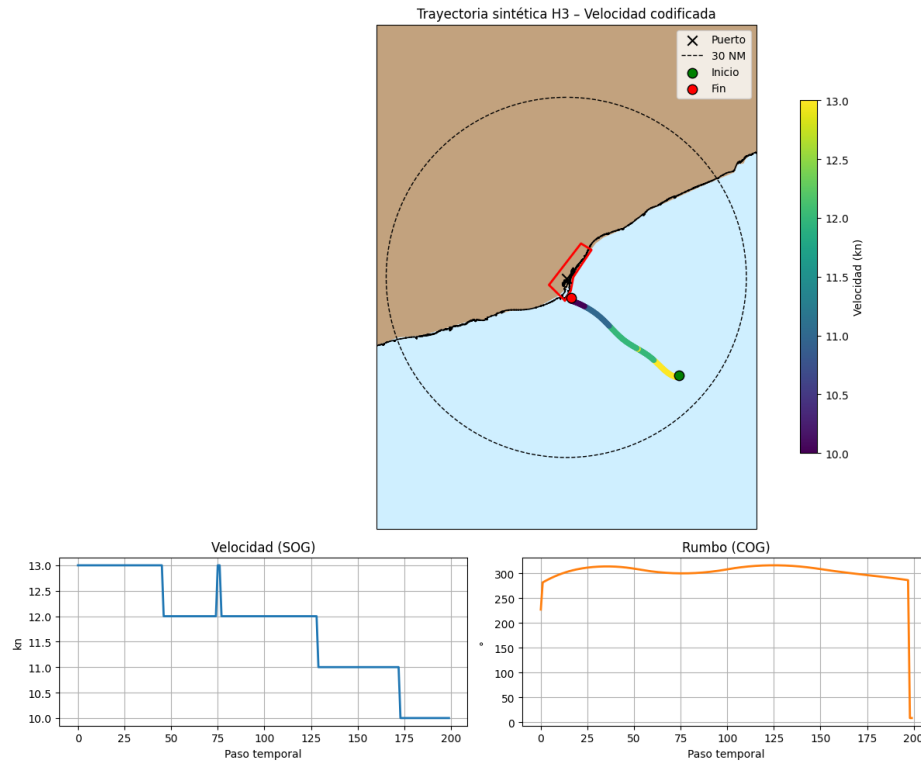


Figura 3-27 Resultado final método mapas de calor y celdas tipo H3 (Fuente: propia)

3.4.2 Buques tipo Fishing que salen, faenan y entran en las inmediaciones del Puerto de Barcelona

3.4.2.1 Discretización espacial del tráfico pesquero mediante cuadrículas tipo UBER(H3)

Al igual que en el caso anterior se emplearán las cuadrículas H3 tipo UBER. En primer lugar, se fija la resolución del sistema de discretización espacial. Este parámetro determina el tamaño de las celdas hexagonales empleadas para discretizar el espacio geográfico y, por tanto, el nivel de detalle del mapa de calor generado. En el presente caso, se opta por seleccionar la resolución 8. A diferencia de los buques tipo *Cargo*, los pesqueros operan de forma más errática y compleja, por ello, en este caso no empleamos una resolución mayor como la empleada en los cargueros (9), con objeto de simplificar el problema al reducir el número de transiciones posibles entre celdas. Una vez probada esta resolución, se apreció que este nivel permite representar con suficiente precisión las maniobras de los pesqueros sin generar una fragmentación excesiva del espacio ni incrementar innecesariamente el coste computacional.

Así mismo, se escoge esta menor resolución para hacer posible la comparación con el otro método de generación de pesqueros, que será desarrollado en apartados posteriores, debido a limitaciones de memoria del hardware empleado (NVIDIA RTX-3060).

Una vez definido el tamaño de las celdas para los pesqueros, se procede a determinar las celdas ocupadas en cada una de las posiciones de las trayectorias AIS de los buques. Para ello, como se hizo en los cargueros, se agrupan todos los datos de latitud y longitud de cada una de las posiciones transmitidas por los barcos recorriendo todos los archivos CSV para fusionarlos en un único conjunto de datos.

A continuación, cada posición AIS registrada para los buques pesqueros, es asignada a una celda hexagonal del sistema H3 empleando la resolución previamente definida. De esta forma, cuando una coordenada cae dentro de los límites de una celda concreta, se considera que el pesquero ha operado o transitado por dicha celda.

Una vez realizada la asignación, se contabiliza el número de registros asociados a cada celda. Este recuento permite estimar la intensidad de presencia de los pesqueros en cada zona del área de estudio, sirviendo como base para la construcción de mapas de calor que reflejan las áreas donde la actividad pesquera es más frecuente.

En el caso de los pesqueros, y a diferencia del caso de los cargueros, a la hora de generar las derrotas no se tendrán solo en cuenta las transiciones entre celdas, sino que también se tendrán en cuenta la tasa de ocupación y como se explicará más adelante, las celdas clasificadas como pertenecientes a zonas de pesca.

3.4.2.2 Definición de vecindad y transiciones entre celdas H3

Como en el caso de los cargueros, es necesario definir cómo se relacionan estas celdas entre sí. En particular, se debe establecer la vecindad entre celdas adyacentes, ya que esta conectividad permite modelar el movimiento entre hexágonos, esto es, las transiciones entre celdas.

Como se dijo anteriormente, el objetivo de este paso es el de capturar patrones de movimiento entre regiones adyacentes, registrando los desplazamientos observados entre hexágonos consecutivos a lo largo de las trayectorias AIS. De esta manera se contabilizará las veces que una traza transita de una celda a otra

Este procedimiento ya ha sido descrito con mayor detalle en el apartado correspondiente a los buques tipo *Cargo*, por lo que no se entrará en mayor detalle.

3.4.2.3 Cálculo y normalización de la tasa de ocupación de celdas H3

Gracias a la contabilización del número de registros AIS asociados a cada celda H3, es decir, al número de veces que los buques transitan por cada una de las celdas, es posible definir una tasa de ocupación. Este indicador refleja la frecuencia de uso de cada zona dentro del área de estudio y permite identificar aquellas celdas que han sido más recurrentes en las trayectorias reales.

Dicha tasa de ocupación se incorpora posteriormente al modelo generativo como un factor adicional en el cálculo de probabilidades de transición, contribuyendo a que las trayectorias sintéticas reproduzcan de forma más realista los patrones espaciales observados en los datos AIS originales.

Una vez definida la tasa de ocupación a partir del número de registros asociados a cada celda, es necesario transformar estos valores absolutos en un parámetro normalizado que pueda incorporarse de manera estable al modelo probabilístico. Para ello, se obtiene en primer lugar el valor máximo de ocupación registrado entre todas las celdas, que actúa como referencia de escala.

Posteriormente, a cada celda se le aplica una transformación logarítmica del tipo:

$$\log(1 + v) \qquad \text{Ecuación 5}$$

Donde v representa el número de registros AIS contabilizados en dicha celda. Esta transformación permite reducir la influencia de valores extremadamente altos y suavizar las diferencias entre zonas muy frecuentadas y zonas de ocupación media, evitando que unas pocas celdas dominen excesivamente el proceso generativo.

Finalmente, los valores transformados se normalizan dividiéndolos por el valor máximo también en escala logarítmica, obteniendo así un coeficiente de ocupación comprendido entre 0 y 1. De esta forma, la celda con mayor número de registros toma el valor unitario, mientras que el resto quedan ponderadas de forma proporcional según su frecuencia relativa de uso.

Este coeficiente normalizado es el que se incorpora posteriormente en el cálculo de los pesos de transición, permitiendo que las trayectorias sintéticas tiendan con mayor probabilidad hacia aquellas zonas que, en los datos reales, presentan una mayor recurrencia espacial.

3.4.2.4 Identificación de zonas de salida, faena y entrada a puerto

3.4.2.4.1 Generación de mapa de calor para análisis de zonas de pesca

Con el objetivo de identificar las zonas de pesca habituales de los buques pesqueros en las inmediaciones del puerto de Barcelona, se plantea un análisis basado en mapas de calor según la velocidad registrada en los datos AIS. El comportamiento de un buque pesquero varía de forma clara en función de la fase operativa en la que se encuentre. Durante el tránsito entre puerto y zona de faena, la navegación suele realizarse a velocidades moderadas y con trayectorias relativamente definidas. Sin embargo, cuando el buque se encuentra desarrollando la actividad pesquera, su velocidad se reduce y tiende a mantenerse en valores bajos durante periodos prolongados.

Por este motivo, se construyen mapas de calor a partir de la velocidad media registrada en cada celda H3 del área de estudio. De esta forma, aquellas zonas donde se observe una concentración significativa de velocidades bajas y relativamente estables pueden asociarse con áreas donde los pesqueros desarrollan su actividad de pesca de manera recurrente.

Para ello, se recorren todos los archivos CSV correspondientes a trayectorias AIS de pesqueros. De cada registro se extraen la latitud, longitud y velocidad sobre el fondo (SOG). Así mismo, se descartan aquellos valores de velocidad nulos o superiores a 30 nudos, ya que estos pueden corresponder a errores de medición o valores no coherentes con la operativa habitual de este tipo de embarcaciones.

Cada posición válida se convierte posteriormente en su celda H3 correspondiente mediante la resolución espacial previamente definida. Finalmente, la velocidad asociada a ese punto se añade a la lista de velocidades de la celda correspondiente.

Como resultado, se obtiene un diccionario en el que cada celda H3 contiene el conjunto completo de velocidades observadas en ella. Esta estructura permite, en fases posteriores, calcular métricas como la velocidad media por celda y construir mapas de calor que faciliten la identificación de zonas de pesca habituales.

Una vez recopiladas las velocidades registradas en cada celda H3, se procede al cálculo de estadísticas que permitan caracterizar el comportamiento cinemático de los pesqueros en cada zona del área de estudio.

Para cada celda se requiere un número mínimo de cinco observaciones, con el fin de evitar estimaciones poco representativas derivadas de muestras demasiado reducidas. En aquellas celdas que cumplen este criterio, se calculan distintas métricas estadísticas explicadas con anterioridad en este trabajo: media, mediana, percentiles 25 y 75, desviación estándar y número total de registros.

Aunque se calculan varias magnitudes, la variable empleada para la representación espacial es la mediana de la velocidad. Esta elección se debe a que la mediana es menos sensible a valores extremos o mediciones puntuales anómalas que pueden aparecer en los datos AIS.

Finalmente, para evitar que valores extremos distorsionen la escala de representación, se establece un rango dinámico basado en los percentiles 5 y 95 de las medianas calculadas. Este procedimiento mejora la interpretación visual del mapa, resaltando de forma más clara las zonas donde la velocidad característica de los pesqueros es significativamente inferior al resto del área.

Se genera así, un mapa de calor según las velocidades, tal y como se ve en la Figura 3-28. Gracias a este, podemos analizar cuáles son las zonas más proclives a la pesca.

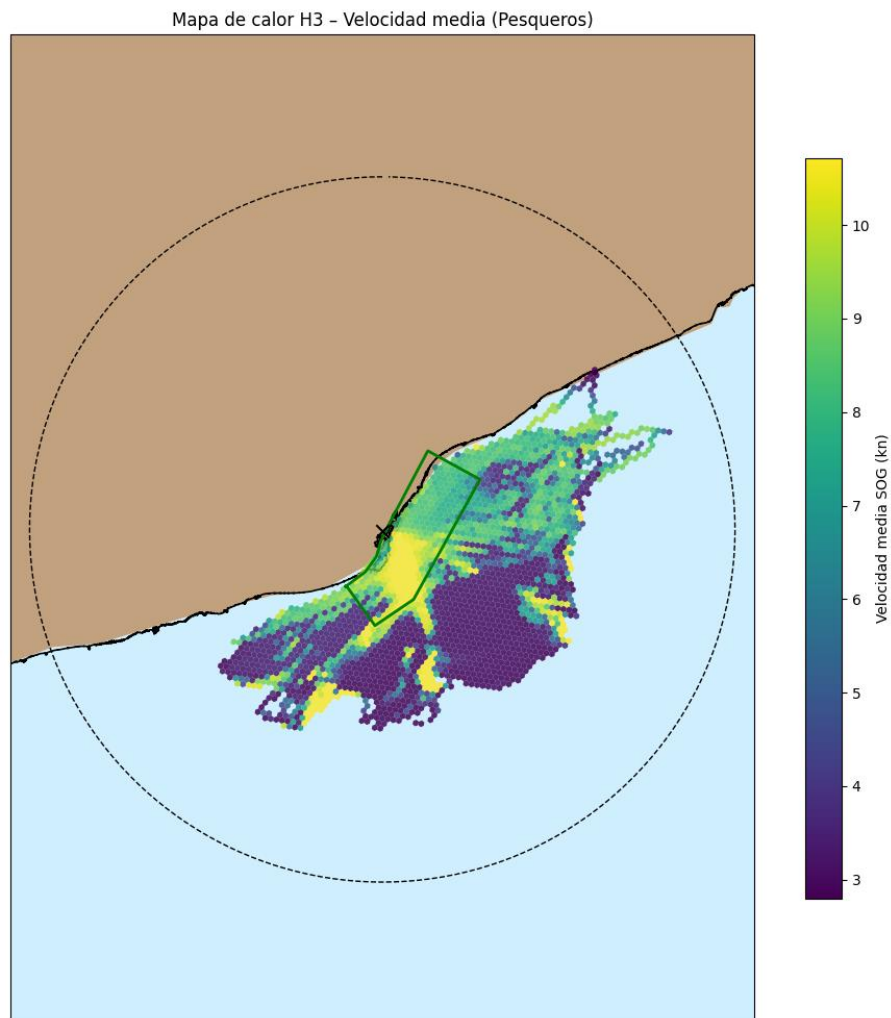


Figura 3-28 Mapa de calor según velocidad media de pesqueros (Fuente: propia)

Así mismo se calculan los valores de velocidad máxima, mínima y media que siguen los buques, valores que serán útiles para cálculos y análisis posteriores:

Velocidad	Resultado (en nudos)
Velocidad mínima	0,1037
Velocidad máxima	12,1
Velocidad media	5.971565700742

Tabla 3-3 Velocidades calculadas

3.4.2.4.2 Identificación de las zonas de pesca

Con el fin de identificar de forma cuantitativa aquellas celdas que presentan un comportamiento compatible con actividad pesquera, se define un criterio basado en la velocidad característica de cada celda.

Para cada celda H3 se toma como referencia la mediana de la velocidad previamente calculada. Se establece un intervalo de velocidades comprendido entre 1.5 y 5.0 nudos como rango potencialmente asociado a la fase de faena. Las celdas cuya velocidad representativa se sitúa fuera de dicho intervalo se consideran no compatibles con actividad de pesca, asignándoles una probabilidad nula.

Dentro del intervalo definido, se aplica una función gaussiana centrada en 3.5 nudos, de forma que la probabilidad de pesca es máxima en torno a ese valor y disminuye progresivamente al alejarse del mismo. Este enfoque evita establecer un umbral rígido y permite modelar de manera más realista la variabilidad natural de la velocidad durante la actividad pesquera.

El intervalo de velocidades (1,5 y 5 nudos), así como la velocidad de 3,5 nudos que representa el centro de la función gaussiana, se escogieron en función a las velocidades que suelen seguir los pesqueros mientras faenan.

Como resultado, cada celda queda asociada a un valor continuo entre 0 y 1 que representa su grado de compatibilidad con comportamiento de pesca. Este parámetro servirá posteriormente como componente del modelo semi-Markov para diferenciar entre fases de tránsito y fases de actividad pesquera. Esto se aprecia en la Figura 3-29.

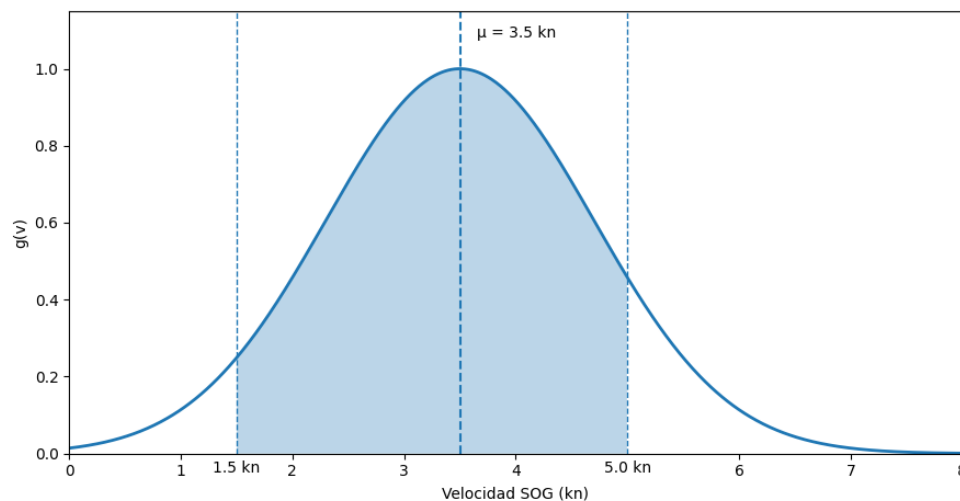


Figura 3-29 Función gaussiana de velocidades (Fuente: propia)

Una vez calculada la velocidad característica por celda H3 y definida la probabilidad de pesca asociada a dicha velocidad, se genera una visualización conjunta para interpretar de forma directa las posibles zonas de actividad pesquera.

En la representación gráfica, las celdas identificadas como compatibles con actividad pesquera se muestran mediante marcadores circulares situados en el centro geométrico de cada celda H3. El tamaño de cada círculo no es constante, sino que se define de forma proporcional al valor de la probabilidad de pesca previamente calculada para dicha celda. Esta probabilidad depende de la proximidad de la velocidad mediana al valor central de la función gaussiana establecida en 3.5 nudos, disminuyendo progresivamente al alejarse de dicho valor dentro del intervalo considerado.

De este modo:

- **Círculos grandes:** celdas cuya velocidad se aproxima más al patrón cinemático esperado durante la fase de faena.
- **Círculos pequeños:** representan zonas donde, aun estando dentro del rango de velocidades definido, la compatibilidad con el modelo es menor.

Esta codificación visual permite introducir una dimensión cuantitativa adicional en la figura, ya que no solo se indica la localización espacial de las posibles zonas de pesca, sino también la intensidad relativa asociada a cada una de ellas, facilitando así su interpretación comparativa dentro del área de estudio. Esta representación se puede apreciar en la Figura 3-30.

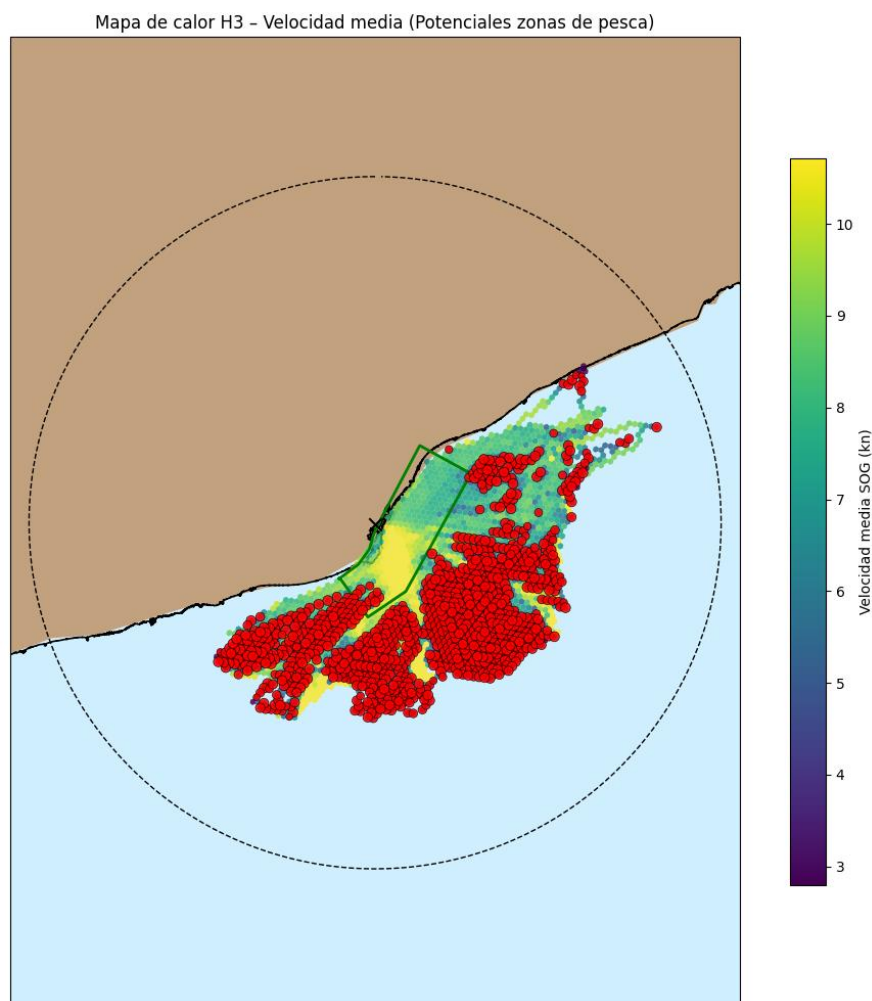


Figura 3-30 Celdas más probables según su probabilidad de ser zona de pesca (Fuente: propia)

3.4.2.5 Establecimiento de celdas objetivo y celdas iniciales

Para definir las posibles celdas iniciales de las trayectorias sintéticas de los buques pesqueros, se parte de las trayectorias reales previamente discretizadas en H3. En este caso, no interesa considerar cualquier punto del área de estudio como origen, sino únicamente aquellas celdas que representen una salida real desde la zona portuaria asociada a los pesqueros.

Con este objetivo, se define una función que permite comprobar si una celda H3 se encuentra dentro del polígono específico de acceso portuario de los buques pesqueros. Para ello, se convierte el identificador de la celda en sus coordenadas centrales (latitud y longitud), y posteriormente se verifica si ese punto está contenido o en el borde del polígono.

A continuación, se recorre el conjunto de trayectorias reales discretizadas y se selecciona la primera celda de cada trayectoria, ya que esta representa el punto inicial del movimiento registrado. Si dicha celda pertenece al polígono de acceso portuario de los pesqueros, se añade al conjunto de celdas iniciales.

De esta manera, se obtiene un conjunto de celdas de inicio basadas exclusivamente en comportamientos reales observados en los datos AIS. Este criterio permite garantizar que las trayectorias sintéticas generadas posteriormente comiencen en ubicaciones coherentes con la operativa real de los buques pesqueros, obteniendo un total de 169 celdas de inicio posibles.

Por otro lado, como en el caso de los buques tipo *Cargo*, se definen celdas objetivo, en las que deben acabar las trayectorias generadas. Mientras que para las celdas de salida se consideraban únicamente aquellas registradas como primera celda real de cada trayectoria, en el caso de las celdas objetivo se

seleccionan todas las celdas H3 que se encuentran dentro del polígono de acceso portuario de los pesqueros, independientemente de si han sido registradas explícitamente como punto final de una trayectoria.

El resultado es el conjunto que contiene todas las celdas UBER situadas en la zona de acceso portuario de los pesqueros considerada como destino válido para las trayectorias sintéticas.

Este enfoque permite modelar el retorno a puerto de los pesqueros de forma más flexible, ya que no se limita exclusivamente a los puntos finales observados en los datos reales, sino que admite como destino cualquier celda coherente con la zona operativa de entrada al puerto. De esta manera se obtienen un total de 298 celdas objetivo.

En la Figura 3-31 se puede observar el conjunto de celdas iniciales, representadas en amarillo, y celdas objetivo, representadas en rojo y amarillo. Hay que destacar que las celdas iniciales comparten función de celda de salida y llegada.

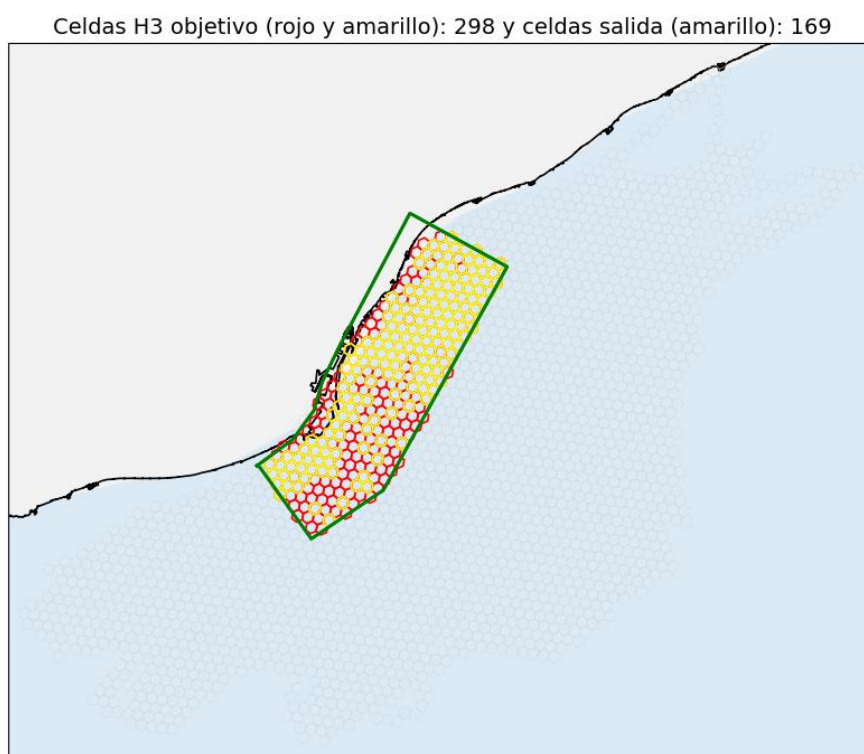


Figura 3-31 Conjunto de celdas iniciales y objetivo de los buques de pesca (Fuente: propia)

3.4.2.6 Generación de trayectorias sintéticas

El generador de trayectorias sintéticas se basa en un modelo probabilístico que, en cada paso, determina la siguiente celda H3 a visitar a partir de la celda actual. La decisión no es determinista, sino que se construye a partir de una combinación de distintos factores extraídos del análisis previo de los datos AIS reales: Transiciones, tasa de ocupación y probabilidad de actividad pesquera estimada para cada celda.

En primer lugar, para cada celda actual se consideran únicamente aquellas celdas vecinas para las que existe información de transición y que cumplen las restricciones espaciales definidas (que no se encuentren en el polígono interior del puerto, para evitar trayectorias sobre tierra). Además, solo se consideran celdas para las que existe información sobre probabilidad de pesca previamente calculada. Sobre este conjunto de celdas candidatas se calcula un peso asociado a cada una de ellas.

El peso base de cada posible transición se obtiene combinando la probabilidad histórica de transición entre celdas (extraída de las trayectorias reales) y el coeficiente de ocupación normalizado de la celda

de destino. De este modo, no solo se tiene en cuenta la frecuencia con la que se ha realizado una transición concreta en los datos reales, sino también la intensidad de uso histórico de la zona de destino.

Adicionalmente, durante determinadas fases del movimiento se introduce un factor asociado a la probabilidad de pesca previamente calculada a partir de la velocidad característica de cada celda. Este factor se activa mientras el tiempo transcurrido fuera de la zona objetivo es menor o igual que la duración asignada a la fase de pesca. Cuando el modelo se encuentra en una fase en la que se favorece la actividad pesquera, las celdas con mayor compatibilidad con el patrón de velocidad de faena reciben un incremento en su peso, aumentando así la probabilidad de que la trayectoria sintética permanezca en zonas coherentes con la actividad de pesca. Finalmente, la siguiente celda se selecciona mediante muestreo aleatorio ponderado según la distribución de probabilidad obtenida.

Para evitar comportamientos poco realistas, se incorporan dos mecanismos adicionales. Por un lado, se penaliza la vuelta inmediata a la celda anterior, reduciendo la probabilidad de oscilaciones repetitivas entre dos estados consecutivos. Por otro lado, se introduce un refuerzo de continuidad espacial que favorece aquellas transiciones que mantienen una dirección coherente con el movimiento previo, aportando inercia al desplazamiento.

Una vez calculados los pesos para todas las celdas candidatas, estos se normalizan para obtener una distribución de probabilidad. Además, se aplica un parámetro de temperatura que controla el grado de aleatoriedad del proceso: valores menores intensifican la diferencia entre opciones más y menos probables, haciendo el movimiento más determinista. En el presente trabajo se elige una temperatura de 0,5. Este valor se escogió mediante la experimentación con los diversos parámetros del generador.

Durante las fases de salida y retorno se incrementa el peso del término de ocupación con el objetivo de favorecer transiciones hacia celdas históricamente más utilizadas. Este mecanismo permite reproducir corredores de navegación observados en los datos AIS reales, facilitando los movimientos de tránsito entre el puerto y las zonas de operación. Durante la fase de pesca, en cambio, la ocupación pierde relevancia relativa frente al término asociado a la probabilidad de actividad pesquera. Además, cuando el modelo se encuentra en celdas con baja probabilidad de pesca, se vuelve a priorizar la ocupación para reforzar los desplazamientos de tránsito.

La trayectoria generada se puede dividir en tres fases diferenciadas:

➤ **Fase de salida:**

La trayectoria sintética comienza en una celda inicial y se genera de forma iterativa, con un máximo de 2000 pasos. Para evitar recorridos excesivamente cortos, se establece un número mínimo de pasos aleatorios comprendidos entre 10 y 50.

Inicialmente, el buque se encuentra dentro del conjunto de celdas objetivo (zona asociada al puerto). En el momento en que la celda actual deja de pertenecer a dicho conjunto, se activa un indicador interno y comienza a incrementarse un contador de tiempo en zona que registra el número de pasos realizados fuera de esta zona inicial (entre 10 y 50). Este mecanismo garantiza que la trayectoria incluya una fase real de salida antes de poder considerar un posible retorno.

➤ **Fase de faena:**

Durante la generación de la trayectoria se define una duración asociada a la fase de pesca, que toma un valor aleatorio entre 2 y 6 pasos. Mientras el número de pasos fuera del conjunto objetivo (tiempo en zona) sea menor o igual a este valor, el modelo favorece transiciones hacia celdas con mayor probabilidad de pesca, previamente calculada a partir de la velocidad característica. Este mecanismo no impone una duración fija de permanencia en zona de faena, sino que incrementa temporalmente la probabilidad de seleccionar celdas compatibles con actividad pesquera.

Finalizado dicho intervalo, en el momento en que el tiempo en zona supera el tiempo en pesca, la trayectoria continúa guiada únicamente por las probabilidades de transición y ocupación, pudiendo

permanecer o abandonar la zona de pesca según la dinámica resultante del modelo, pero transicionando hacia la fase de retorno (regreso a la zona próxima al puerto).

➤ **Fase de retorno:**

La trayectoria no finaliza simplemente cuando vuelve a entrar en una celda objetivo. Para evitar terminaciones prematuras justo en el borde de la zona o que finalicen sin salir de la zona inicial, se exige que deba haber salido previamente de esa zona y que el buque permanezca dentro del conjunto objetivo durante al menos 10 pasos consecutivos.

Para ello, se utiliza un contador que solo se incrementa cuando la celda actual pertenece al conjunto objetivo y ya se ha producido la fase de salida. Si este contador alcanza el valor 10, la trayectoria se considera completada y se detiene la simulación.

Una vez ajustados estos comportamientos se genera una trayectoria en celdas H3 empleando el generador definido previamente, utilizando como conjunto de destino las celdas objetivo y seleccionando aleatoriamente una celda inicial. A continuación, la trayectoria en celdas H3 se convierte a coordenadas geográficas (latitud y longitud) y se aplica un proceso de suavizado sobre la trayectoria resultante para evitar saltos irreales en la trayectoria.

Tras esto se generan en el mapa las trayectorias sintéticas. En el primer caso se generan 3 trayectorias para su verificación, tal y como se muestra en la Figura 3-32.

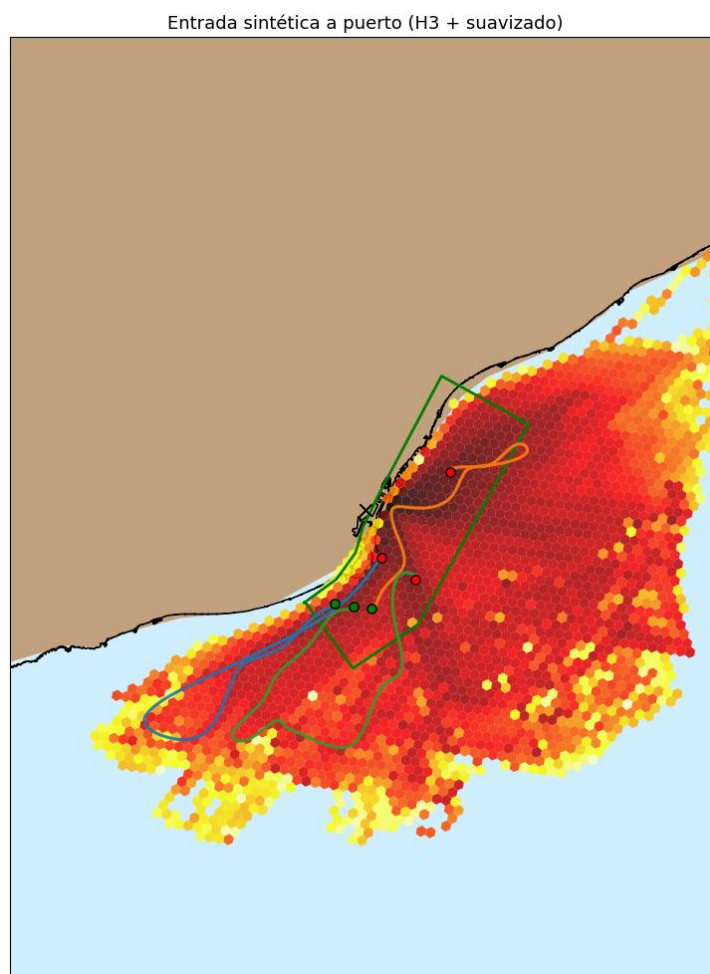


Figura 3-32 Trayectorias sintéticas de tres pesqueros empleando mapas de calor (Fuente: propia)

A continuación, se generan 200 trayectorias para comprobar si el modelo es capaz de generar trayectorias que lleguen a todo el espectro del mapa de calor. Esto se muestra en la Figura 3-33, donde se comprueba que las trayectorias se distribuyen por todo el dominio disponible de manera correcta.

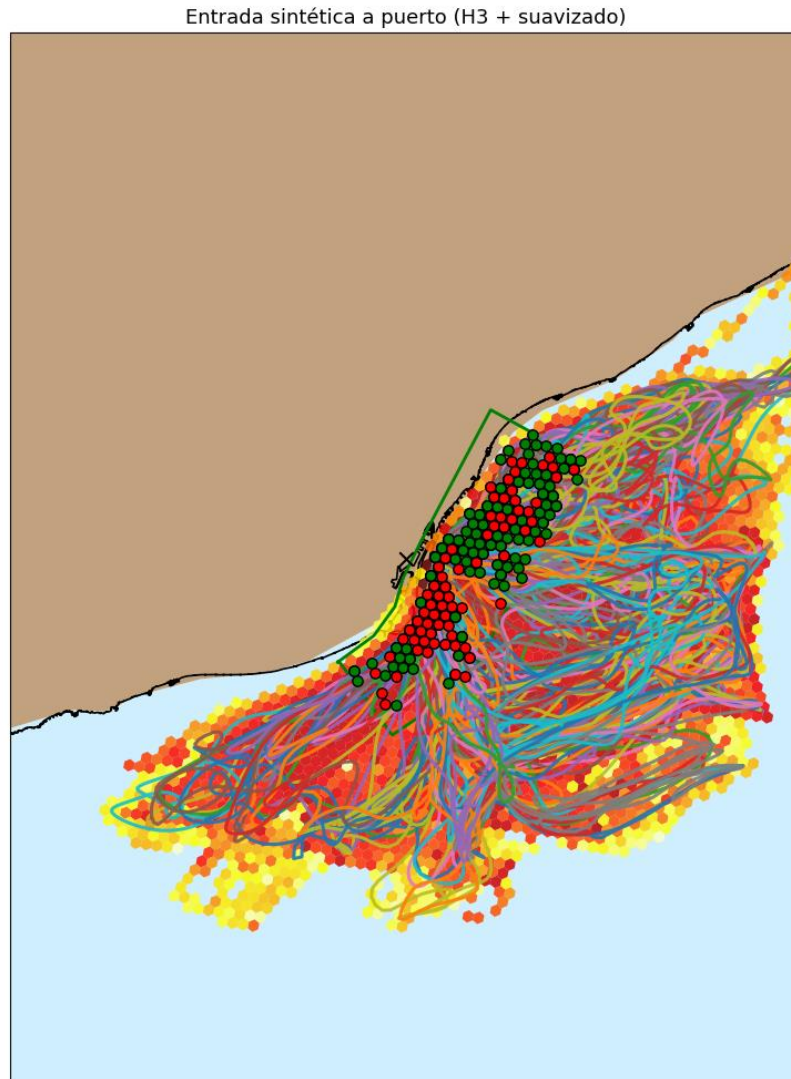


Figura 3-33 200 trayectorias sintéticas mediante mapas de calor y celdas H3 (Fuente: propia)

Con el objetivo de identificar si las trayectorias generadas sintéticamente se corresponden con las trayectorias reales AIS, se generan estas 200 trayectorias sobre el conjunto de datos de trayectorias reales. Esto se muestra en la Figura 3-34, donde se puede apreciar, que las trayectorias generadas mediante mapas de calor y celdas H3 se cubren de manera parecida el dominio de las trayectorias reales, comportándose de similar manera en la mayoría de los casos. En la figura las líneas grises representan las trayectorias reales y las generadas aparecen en diversos colores.

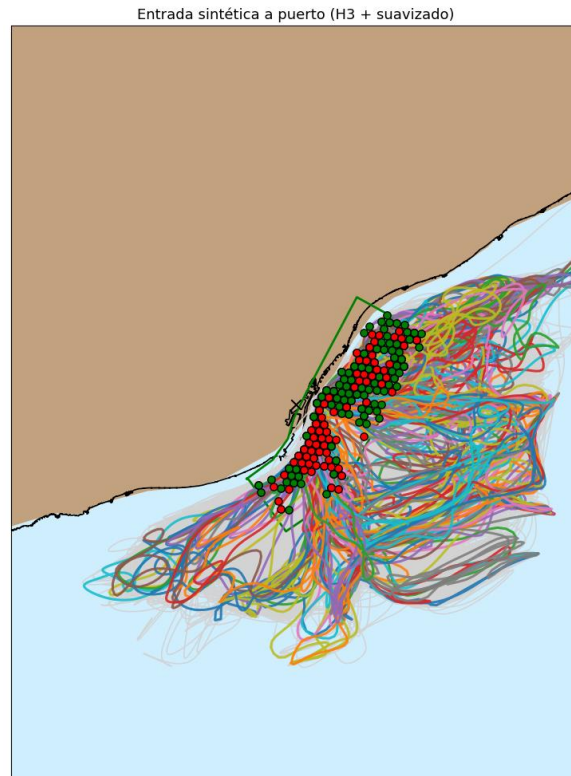


Figura 3-34 Comparación de trayectorias generadas mediante mapas de calor y trayectorias reales (Fuente: propia)

3.4.2.7 Asignación de velocidades a las trayectorias

La asignación de velocidades a las trayectorias se realiza de igual manera que en el caso de los buques tipo *Cargo*, pudiéndose ver el proceso en el apartado “Construcción de mapas de calor y asignación de velocidades”.

Aplicando estas velocidades a las trayectorias se representarían tal y como aparece en la Figura 3-35.

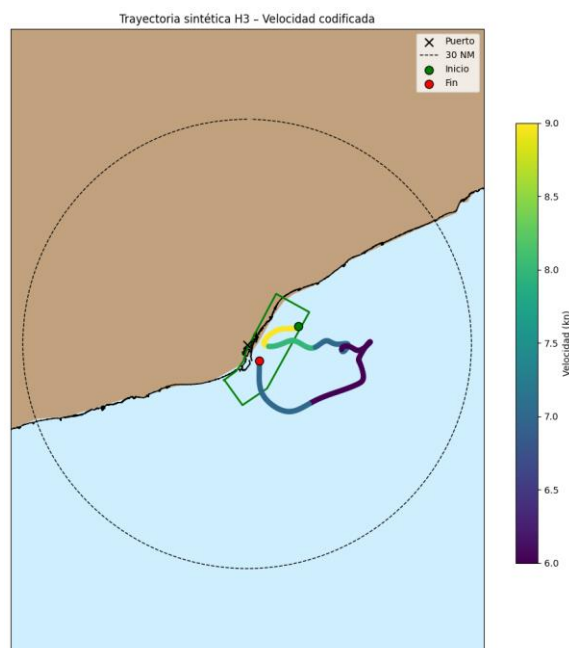


Figura 3-35 Representación de trayectoria con velocidad asignada de los pesqueros (Fuente: propia)

Así mismo, como se hizo en los buques cargueros, y con el objetivo de facilitar la posterior comparación de resultados, se le añade al resultado dos gráficas de rumbo y velocidad. Esto se muestra en la Figura 3-36.

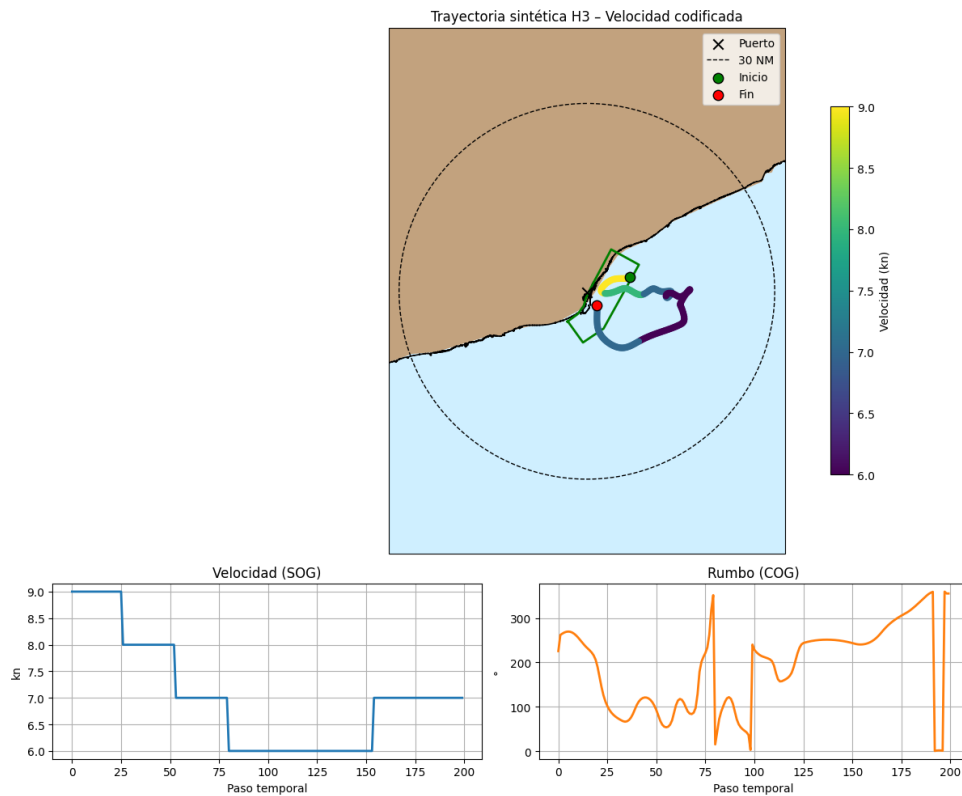


Figura 3-36 Trayectoria completa de buque pesquero con gráficas de rumbo y velocidad (Fuente: propia)

3.5 Generación de trayectorias sintéticas mediante redes generativas antagónicas condicionales (cGAN) y estimación cinemática con KNN

En este apartado se desarrolla un segundo enfoque para la generación de trayectorias sintéticas basado en redes generativas antagónicas condicionales (cGAN) combinadas con un mecanismo de selección mediante *K-Nearest Neighbors* (KNN). A diferencia del método basado en cuadrículas H3, este enfoque se apoya en técnicas de aprendizaje automático para modelar directamente la estructura de las trayectorias a partir de los datos AIS reales.

Este enfoque híbrido combina la capacidad generativa de las cGAN con la simplicidad de los métodos basados en vecindad (KNN), buscando obtener trayectorias sintéticas que mantengan tanto coherencia estadística como realismo.

3.5.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona

3.5.1.1 Preparación del conjunto de datos para el entrenamiento de la cGAN

Como en el caso de los mapas de calor, primeramente, debemos definir de dónde vamos a obtener los datos AIS correspondientes a los buques tipo *Cargo* que entran en el puerto de Barcelona. Al igual que anteriormente, estos se obtendrán del CSV procesado y filtrado previamente, obteniendo un total de 3786 trayectorias útiles para el proceso.

Una vez tenemos el conjunto de datos AIS, el primer paso es definir el número de puntos que componen cada secuencia de posiciones utilizada como entrada y salida del modelo Conditional Generative Adversarial Network (cGAN).

Este valor determina la longitud temporal de las trayectorias utilizadas en el entrenamiento, es decir, el número de observaciones consecutivas de posición que describen el movimiento de cada buque. En el presente estudio se fijará en 100, lo que significa que cada trayectoria se representa como una secuencia compuesta por cien puntos ordenados en el tiempo, garantizando que todas las muestras tengan la misma dimensión y facilitando así su procesamiento por parte de la red neuronal.

A continuación, se realiza un remuestreo para que todas las secuencias tengan una longitud temporal fija. En concreto, transforma una trayectoria original de longitud variable en una nueva trayectoria con un número fijo de puntos, que en este trabajo se ha establecido en 100. Esta operación es necesaria para poder utilizar las trayectorias como entrada de la cGAN, ya que este tipo de modelos requiere que todas las muestras del conjunto de datos tengan la misma dimensión. Una vez se recibe la trayectoria con longitud variable existen dos opciones:

- Si la trayectoria ya tiene la longitud deseada: se devuelve sin modificaciones.
- Si la trayectoria no tiene la longitud deseada: se genera un nuevo conjunto de puntos mediante interpolación lineal obteniendo una nueva secuencia de posiciones igualmente distribuidas a lo largo del recorrido original.

Como resultado se obtendrá un nuevo conjunto de datos organizado en tres dimensiones que en el presente caso es (3785, 100, 2):

Número	Descripción
3785	Número total de trayectorias
100	Número de puntos que tiene cada trayectoria
2	Las dos coordenadas de cada punto: latitud y longitud

Tabla 3-4 Estructura de 3 dimensiones

Una vez realizado el paso de homogeneizar las muestras, aún debemos realizar tres pasos más para que los datos se puedan emplear como entrada del modelo cGAN, partiendo del conjunto de datos remuestreado en el paso anterior.

Partiendo de las trayectorias remuestreadas que tiene tres dimensiones (3785, 100, 2), se llevan a cabo tres pasos fundamentales para la preparación de los datos para entrenar la cGAN:

3.5.1.1.1 Definición de las condiciones

Se definen las variables condicionantes que se utilizarán en la cGAN. En este caso, se ha decidido emplear como condición el punto inicial y el punto final de cada trayectoria, ya que ambos determinan el origen y el destino del movimiento y aportan información clave sobre la maniobra realizada por el buque. Para construir esta condición, se extrae el primer y el último punto de cada secuencia y se agrupan en un único vector. De este modo, se obtiene una matriz denominada condiciones de dimensiones (3785, 4) donde:

Número	Descripción
3785	Número total de trayectorias
4	Latitud inicial, latitud final, longitud inicial y longitud final

Tabla 3-5 Representación en dos dimensiones

3.5.1.1.2 Cálculo de incrementos temporales

Se calculan las diferencias entre puntos consecutivos de cada trayectoria, conocidas como deltas. En este caso, cada trayectoria está formada por 100 posiciones sucesivas en latitud y longitud. Al aplicar esta función, lo que se obtiene no son las posiciones absolutas, sino el cambio entre un punto y el siguiente. Es decir, para cada trayectoria se calculan 99 incrementos, donde cada uno indica cuánto se ha desplazado el buque en latitud y longitud entre dos puntos consecutivos. En la Figura 3-37 se puede apreciar el incremento de latitud y longitud entre dos puntos.

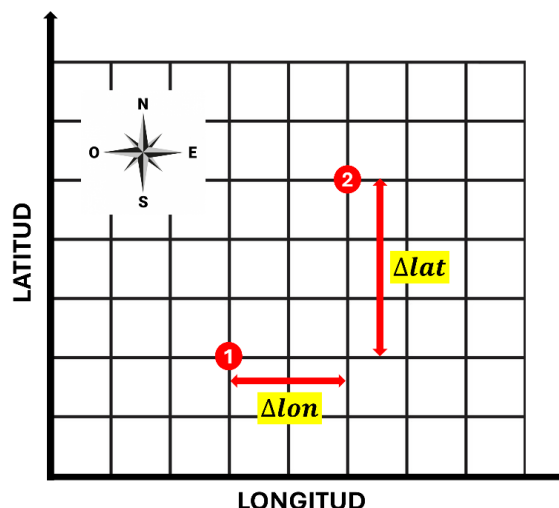


Figura 3-37 Incremento de latitud y longitud (Fuente: propia)

Por lo tanto, el nuevo tensor tendrá las siguientes dimensiones: (3785, 99, 2) donde:

Número	Descripción
3785	Número total de trayectorias
99	Número de incrementos de cada trayectoria
2	Incremento de latitud y longitud

Tabla 3-6 Estructura de tres dimensiones

Trabajar con deltas en lugar de posiciones absolutas permite que el modelo aprenda patrones de movimiento relativos, lo que suele facilitar el entrenamiento y mejorar la estabilidad del modelo generativo.

3.5.1.1.3 Normalización de datos

Esta etapa es necesaria porque las magnitudes de las variables son muy diferentes:

- Deltas: incrementos pequeños entre puntos consecutivos (cambios en latitud y longitud).
- Condiciones: coordenadas absolutas de inicio y fin (valores en torno a 41 o 2 grados)

Si ambos tipos de datos se introdujeran directamente en la red, las diferencias de escala podrían afectar negativamente al proceso de entrenamiento.

Por este motivo, se emplean dos escaladores independientes, uno para deltas y otro para condiciones, ambos con rango de salida comprendido entre -1 y 1. Este procedimiento permite que todos los valores que entran en la red estén en una escala comparable, lo que favorece la estabilidad del entrenamiento y mejora el comportamiento del modelo generativo.

3.5.1.2 Arquitectura del generador del cGAN

En una GAN (Generative Adversarial Network), el generador es la red neuronal encargada de crear datos sintéticos. Su objetivo es producir muestras artificiales que se parezcan lo máximo posible a los datos reales.

Para comenzar, hay que definir la arquitectura del generador de la cGAN. Para ello se empleará como base la estructura estándar de *PyTorch* para redes neuronales. Este modelo tiene como objetivo generar trayectorias sintéticas, pero no de forma completamente aleatoria, sino condicionadas a un punto inicial y un punto final previamente definidos.

El funcionamiento es el siguiente: el generador recibe dos entradas. Por un lado, un vector de ruido aleatorio, que introduce variabilidad y permite que no todas las trayectorias generadas sean iguales. Por otro lado, un vector de condición que contiene la latitud y longitud iniciales y finales. Al combinar ambas entradas, el modelo aprende a generar una secuencia de desplazamientos espaciales (deltas) que conecte coherentemente el punto de inicio con el punto de destino. En lugar de generar directamente coordenadas absolutas, el modelo produce incrementos sucesivos de latitud y longitud. Esto facilita el aprendizaje del patrón de movimiento paso a paso.

El generador se inicializa indicando una serie de parámetros que definen su estructura y el tipo de datos que va a manejar. En primer lugar, se especifica la dimensión del espacio latente, que corresponde al tamaño del vector de ruido aleatorio que introduce variabilidad en las trayectorias generadas. También se define la dimensión del vector de condición, que en este trabajo está formada por cuatro valores: latitud y longitud iniciales y finales. Asimismo, se establece el tamaño del estado oculto, que determina la capacidad interna de la red para modelar patrones complejos, y la longitud temporal de la secuencia, que fija el número de pasos que tendrá cada trayectoria generada que en este caso es 99. Finalmente, la dimensión de salida se fija en 2, ya que en cada paso se generan dos valores correspondientes a los desplazamientos en latitud y longitud.

En primer lugar, el generador combina el vector de ruido aleatorio y el vector de condición (inicio y fin de la trayectoria) y los pasa por una red totalmente conectada. Esta parte simplemente transforma esa información inicial en un conjunto de valores internos con el tamaño adecuado para poder trabajar después como si fuese una secuencia temporal. Es decir, aquí no se genera todavía la trayectoria, sino que se prepara una representación inicial a partir del ruido y de la condición.

Después, esa representación se reorganiza para que tenga forma de secuencia y se introduce en una red LSTM de dos capas. La LSTM es la parte encargada de dar coherencia temporal al movimiento. Como una trayectoria es una sucesión ordenada de desplazamientos, esta red aprende cómo evoluciona el movimiento paso a paso, capturando la dependencia entre posiciones consecutivas.

Finalmente, la salida de la LSTM pasa por una capa lineal que convierte cada estado interno en dos valores: el desplazamiento en latitud y el desplazamiento en longitud. Estos valores se limitan al rango $[-1, 1]$, que es el mismo rango utilizado en la normalización previa de los datos, lo que facilita que el modelo genere resultados compatibles con las trayectorias reales.

En la Figura 3-38 se puede ver un esquema del funcionamiento del generador del cGAN.

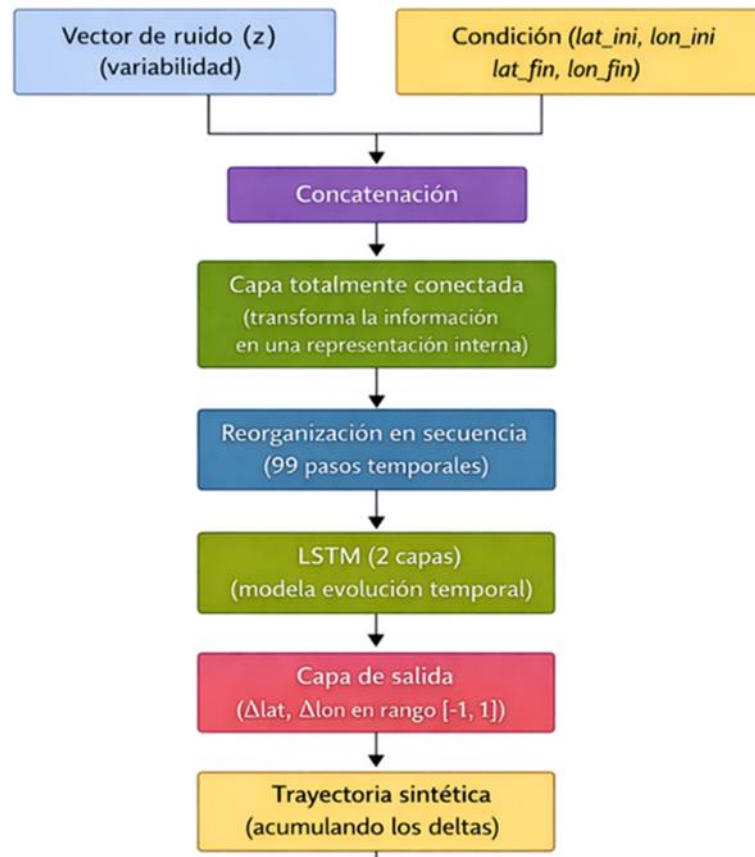


Figura 3-38 Esquema del generador cGAN (Fuente: propia)

En resumen, el generador recibe un vector de ruido y el vector que contiene el punto inicial y final de la trayectoria, y los une en una única entrada para que la red disponga de toda la información necesaria. Esa entrada se transforma primero en una representación interna con forma de secuencia y posteriormente se procesa mediante una LSTM, que modela la evolución temporal del movimiento. Finalmente, la red produce una serie de desplazamientos en latitud y longitud que, al acumularse desde el punto inicial, dan lugar a la trayectoria sintética generada.

3.5.1.3 Arquitectura del discriminador del cGAN

En una red generativa antagónica (GAN), el discriminador es la red encargada de evaluar si una muestra es real o generada artificialmente. En el caso de este trabajo, el discriminador debe decidir si la trayectoria de buque (expresada como secuencia de deltas en latitud y longitud) pertenece al conjunto de datos AIS reales o si ha sido generada por el generador.

Al tratarse, en este caso, de una cGAN, esta decisión no se toma únicamente a partir de la trayectoria, sino también teniendo en cuenta una condición adicional: el punto inicial y el punto final de la ruta. Por tanto, el discriminador no solo analiza cómo se mueve el buque, sino si ese movimiento es coherente con el origen y destino indicados.

Para comenzar, hay que definir la arquitectura del discriminador de la cGAN, y para ello, lo primero es saber que este se divide en tres bloques claramente diferenciados:

1. Rama de la trayectoria (análisis de deltas):

Esta parte del discriminador, como su propio nombre indica, se encarga de analizar la trayectoria propiamente dicha, que en este trabajo no se introduce como posiciones absolutas, sino como una secuencia de desplazamientos consecutivos en latitud y longitud (deltas). Al tratarse de una secuencia temporal, se utiliza una red LSTM de dos capas para poder capturar la dependencia entre movimientos

sucesivos. En concreto, se emplea una LSTM bidireccional, lo que significa que la red analiza la trayectoria tanto en el sentido temporal directo como en el inverso. Esto facilita la detección de inconsistencias geométricas o patrones de movimiento poco realistas que podrían no ser evidentes al analizar la secuencia únicamente hacia adelante, permitiendo obtener una representación más completa del recorrido, ya que el modelo no solo tiene en cuenta cómo evoluciona el movimiento hacia adelante, sino también la coherencia global del conjunto de desplazamientos. El resultado de esta parte es un vector interno que resume el comportamiento completo de la trayectoria y que servirá posteriormente para tomar la decisión final.

2. Rama de la condición (inicio y fin):

La segunda parte del discriminador analiza la información estática asociada a la trayectoria, es decir, el punto inicial y el punto final definidos por la condición de la cGAN. Esta información se introduce como un vector formado por cuatro valores: latitud y longitud iniciales, y latitud y longitud finales. A diferencia de la trayectoria, esta información no tiene carácter temporal, por lo que se procesa mediante una red totalmente conectada sencilla. El objetivo de este bloque no es generar una decisión, sino transformar las coordenadas de inicio y destino en una representación interna compatible con la obtenida en la rama de la trayectoria. De este modo, el modelo podrá evaluar no solo cómo se mueve el buque, sino si ese movimiento es coherente con el origen y destino indicados.

3. Fusión y clasificación final:

Una vez obtenidas las representaciones internas de la trayectoria y de la condición, ambas se concatenan en un único vector que contiene toda la información relevante: cómo se ha movido el buque y cuál era su punto de partida y llegada. Este vector combinado se introduce en una serie de capas totalmente conectadas que actúan como clasificador final. En esta etapa se aplican funciones de activación no lineales y una capa de *dropout*. El *dropout* lo que hace es “apagar” de forma aleatoria algunas neuronas durante el entrenamiento. Esto evita que la red aprenda de memoria las trayectorias reales del conjunto de datos.

Finalmente, una neurona de salida con activación sigmoide genera un valor comprendido entre 0 y 1, que representa la probabilidad de que la trayectoria analizada sea real. De esta forma, el discriminador emite su veredicto, permitiendo entrenar al generador de manera adversarial. Este valor (entre 0 y 1) se utiliza durante el entrenamiento adversarial para guiar tanto la actualización del discriminador como la del generador.

3.5.1.4 Proceso de entrenamiento adversarial

El entrenamiento de una red cGAN se basa en un proceso adversarial en el que intervienen dos redes neuronales que se entrenan de manera simultánea, pero con objetivos opuestos. Por un lado, el discriminador aprende a diferenciar entre trayectorias reales procedentes de los datos AIS y trayectorias generadas artificialmente; por otro, el generador intenta producir trayectorias sintéticas que sean lo suficientemente realistas como para engañar al discriminador. En cada iteración, primero se ajustan los pesos del discriminador utilizando ejemplos reales y generados, y después se actualizan los del generador para mejorar su capacidad de imitación. En el caso de una cGAN, este proceso se realiza siempre bajo una condición adicional, en este trabajo, el punto inicial y final de la trayectoria. De esta forma, el modelo no solo aprende a generar trayectorias realistas, sino también coherentes con el origen y destino establecidos.

Con el fin de obtener trayectorias más realistas, se añade al entrenamiento del generador un término adicional en la función de pérdida llamado *spatial smoothness loss*. Este término penaliza los cambios bruscos entre desplazamientos consecutivos en latitud y longitud. Como el generador produce la trayectoria en forma de deltas, se calcula la diferencia entre cada delta y el siguiente; si la trayectoria es suave, esas diferencias serán pequeñas, mientras que si existen giros bruscos serán mayores. Este valor se suma a la pérdida total del generador, obligándolo no solo a engañar al discriminador, sino también a

generar trayectorias más suaves y coherentes desde el punto de vista físico, y respecto a las características de las derrotas de los barcos tipo *Cargo*.

Antes de comenzar el entrenamiento se deben ajustar una serie de datos para el buen funcionamiento del entrenamiento:

Parámetro	Traducción	Número
EPOCHS	Épocas	150
BATCH SIZE	Tamaño de Batch	64
LATENT DIMENSION	Dimensión del espacio latente	100
CONDITION DIMENSION	Dimensión de la condición	4
INPUT DIMENSION	Dimensión de entrada	2
HIDDEN DIMENSION	Dimensión del estado oculto	64
SEQUENCE LENGHT	Longitud de la secuencia	99
LEARNING RATE	Tasa de aprendizaje	0,0001
B1	B1	0,5
LAMBDA SPATIAL	Peso de penalización espacial	0,05

Tabla 3-7 Parámetros elegidos para el entrenamiento

En primer lugar, el **número de épocas**, que en nuestro caso es 150, indica cuántas veces el modelo recorre completamente el conjunto de datos de entrenamiento. Una época equivale a una pasada completa por todas las trayectorias disponibles. Aumentar el número de épocas permite que el modelo aprenda con mayor profundidad, aunque un número excesivo podría provocar sobreajuste.

El **tamaño de batch**, que en nuestro caso es 64, representa el número de trayectorias que se procesan simultáneamente antes de actualizar los pesos del modelo. En lugar de utilizar todo el conjunto de datos en cada actualización, se divide en pequeños grupos o lotes. Esto reduce el consumo de memoria y permite un entrenamiento más estable y eficiente.

El parámetro **dimensión del espacio latente**, que en nuestro caso es 100, define la dimensión del vector de ruido aleatorio que se introduce en el generador. Este vector es el que aporta variabilidad a las trayectorias generadas, permitiendo que el modelo no produzca siempre la misma ruta para una misma condición.

El valor 4 de **dimensión de la condición**, corresponde al tamaño del vector de condición, formado por las coordenadas de latitud y longitud iniciales y finales. Esta información guía la generación para que las trayectorias sean coherentes con el origen y destino establecidos.

El parámetro de **dimensión de entrada** es 2 e indica que cada paso de la secuencia está compuesto por dos valores: el incremento en latitud y el incremento en longitud. Es decir, cada delta contiene dos componentes espaciales.

El valor 64 de **dimensión del estado oculto** define el tamaño del estado interno de las capas LSTM tanto en el generador como en el discriminador. Este parámetro determina la capacidad del modelo para capturar patrones complejos en las trayectorias. Valores mayores permiten representar relaciones más complejas, aunque incrementan el coste computacional.

La variable **longitud de secuencia** representa la longitud temporal de cada trayectoria, es decir, el número de deltas que componen cada secuencia, que en este caso es 99. En este trabajo coincide con el número de puntos fijado tras el remuestreo previo.

La **tasa de aprendizaje** ($LR = 0.0001$) controla la magnitud de los ajustes que se realizan en los pesos del modelo durante cada actualización. Un valor pequeño hace que el aprendizaje sea más estable, aunque más lento, sin embargo valores más grandes permiten una mayor variabilidad en el aprendizaje.

El parámetro $B1 = 0.5$ corresponde al primer coeficiente del optimizador Adam y regula el promedio móvil de los gradientes. Este valor es habitual en el entrenamiento de GAN, ya que contribuye a una mayor estabilidad[79].

Finalmente, el **coeficiente peso de penalización espacial** 0.05 determina el peso que tiene la penalización de suavidad espacial dentro de la pérdida total del generador. Este valor controla el equilibrio entre la pérdida adversarial y la penalización geométrica. Si fuese demasiado alto, el modelo priorizaría en exceso la suavidad; si fuese demasiado bajo, la penalización apenas tendría efecto.

Por lo tanto, el entrenamiento de la cGAN se realiza durante 150 épocas, lo que significa que el modelo recorre completamente el conjunto de trayectorias 150 veces. En cada iteración se procesan grupos de 64 trayectorias (batch), formadas por secuencias de deltas normalizados junto con su condición asociada, compuesta por las coordenadas iniciales y finales. Para el aprendizaje adversarial se utiliza la función de pérdida *Binary Cross Entropy* (BCE), ya que el discriminador debe clasificar cada trayectoria como real o generada. Tanto el generador como el discriminador se optimizan mediante el algoritmo Adam con una tasa de aprendizaje inicial de 0,0001.

La BCE es una métrica diseñada para problemas de clasificación binaria que cuantifica la distancia entre la probabilidad predicha por el modelo (un valor continuo entre 0 y 1) y la etiqueta real (0 para trayectorias generadas, 1 para reales). Funciona penalizando de forma logarítmica las predicciones erróneas, castigando severamente al modelo cuando está muy seguro de una clasificación incorrecta[15].

El optimizador Adam es ampliamente empleado porque ajusta la velocidad de aprendizaje de forma automática e independiente para cada peso de la red neuronal, basándose en la inercia (hacia dónde iba) y en la fricción (cuánto está variando). Esto hace que el entrenamiento sea mucho más rápido y estable sin necesidad de realizar ajustes manualmente [80].

Durante la primera mitad del entrenamiento, el modelo utiliza un *learning rate* de 0,0001. Esto permite que los pesos se ajusten con cambios relativamente amplios, lo que facilita que el modelo aprenda rápidamente los patrones generales de las trayectorias. Sin embargo, cuando el entrenamiento avanza y el modelo ya se encuentra cerca de una solución estable, mantener ese mismo ritmo de actualización puede provocar pequeñas oscilaciones alrededor del resultado buscado.

Por este motivo, en la época 75 se reduce el *learning rate* a la mitad (0,00005). Al disminuir el tamaño de los ajustes, los cambios en los pesos pasan a ser más pequeños y controlados. Esto permite que el modelo refine mejor la solución alcanzada, evitando saltos bruscos y mejorando la estabilidad del entrenamiento en las últimas épocas.

En cada *batch* se actualiza primero el discriminador. Para ello, se le presentan trayectorias reales etiquetadas como 1 y trayectorias generadas etiquetadas como 0. A partir de estas predicciones se calcula la pérdida media y se actualizan sus pesos para mejorar su capacidad de distinguir entre ambas. Posteriormente se entrena el generador. En esta fase se generan nuevas trayectorias sintéticas a partir de un vector de ruido aleatorio y la condición correspondiente, y se introducen en el discriminador. Ahora el objetivo del generador es que dichas trayectorias sean clasificadas como reales, es decir, engañar al discriminador.

Además de la pérdida adversarial, se incorpora la función *spatial smoothness loss*, previamente explicada, que penaliza cambios bruscos entre deltas consecutivos. Este término se multiplica por un coeficiente $\lambda = 0,05$ y se suma a la pérdida total del generador. De este modo, el modelo no solo aprende a producir trayectorias que el discriminador no pueda distinguir, sino también a generar movimientos más suaves y coherentes desde el punto de vista geométrico. Durante el entrenamiento se calcula la pérdida media por época y se guarda el estado del modelo correspondiente a la menor pérdida del generador, permitiendo conservar la configuración más estable obtenida.

Por otro lado, el modelo permite guardar los resultados con el objetivo de preservar los resultados obtenidos y permitir su reutilización posterior sin necesidad de repetir el proceso completo de entrenamiento.

3.5.1.5 Reconstrucción y postprocesado geométrico de trayectorias

Una vez entrenado el modelo, el generador produce trayectorias sintéticas en forma de incrementos sucesivos de latitud y longitud (deltas) normalizados. Sin embargo, estos valores no pueden representarse directamente sobre un mapa, ya que no corresponden a coordenadas geográficas reales, sino a desplazamientos relativos en un rango escalado. Por este motivo, es necesario realizar un proceso de reconstrucción que transforme dichos deltas en posiciones absolutas expresadas en latitud y longitud.

En primer lugar, se procede a deshacer la normalización aplicada tanto a los deltas generados como a las condiciones asociadas, recuperando así los valores en su escala original. Posteriormente, para cada trayectoria generada, se calcula la suma acumulativa de los deltas desescalados, lo que permite obtener la evolución espacial relativa del movimiento. A continuación, se suma el punto inicial real a dicha trayectoria relativa, obteniéndose así una secuencia de coordenadas absolutas que describe el recorrido completo del buque.

No obstante, debido a la naturaleza probabilística del modelo generativo, puede producirse una pequeña desviación acumulada que provoque que el punto final reconstruido no coincida exactamente con el punto de destino establecido en la condición. Para garantizar la coherencia global de la trayectoria, se implementa una corrección geométrica progresiva denominada anclaje. Esta corrección consiste en calcular el error total existente entre el punto final generado y el punto final objetivo, y distribuir dicho error de forma lineal a lo largo de la trayectoria. De este modo, la corrección aplicada es mínima en los primeros instantes y aumenta progresivamente hasta compensar completamente la desviación en el último punto.

Como resultado de este procedimiento, se obtienen trayectorias sintéticas de 100 puntos expresados en coordenadas geográficas reales (latitud y longitud) que respetan tanto la dinámica local aprendida por el modelo como la coherencia global impuesta por la condición de inicio y destino.

Sin embargo, con el objetivo de poder comparar las trayectorias generadas tanto con otros métodos como con los datos reales AIS, se había acordado que todas las trayectorias debían tener 200 puntos.

Para ello, se genera una nueva función donde se aplica un proceso adicional de suavizado con el objetivo de mejorar su continuidad y apariencia geométrica. Este procedimiento cumple dos funciones principales: por un lado, aumenta la densidad de puntos de la trayectoria de 100 a 200, y por otro, reduce pequeñas irregularidades derivadas del proceso generativo o de la discretización previa.

En primer lugar, la trayectoria original, compuesta por 100 puntos, se remuestrea para obtener una nueva secuencia de 200 puntos distribuidos uniformemente a lo largo del recorrido. Este aumento de resolución mediante la función *interpolar_trayectoria*, no introduce información adicional, sino que interpola entre los puntos existentes, permitiendo representar la trayectoria con mayor continuidad.

Adicionalmente, se aplica un suavizado mediante una interpolación *spline* paramétrica, explicada con anterioridad, que modela la latitud y la longitud como funciones suaves. A diferencia de una interpolación lineal simple, este método permite obtener una curva más continua, reduciendo cambios bruscos de dirección y pequeñas oscilaciones no realistas. El grado de suavizado está controlado por el parámetro s , que regula cuánto puede separarse la curva ajustada de los puntos originales. En este trabajo se ha adoptado un valor $s=0,003$, seleccionado de forma experimental atendiendo a la escala geográfica de las coordenadas empleadas. Valores inferiores apenas corregían las irregularidades, mientras que valores superiores producían un sobre suavizado que alteraba la geometría global de la trayectoria. El valor adoptado permite eliminar oscilaciones puntuales manteniendo la forma general del recorrido.

Finalmente, se garantiza que el primer y último punto de la trayectoria suavizada coincidan exactamente con los puntos originales. De este modo, el proceso de suavizado no altera ni el origen ni el destino del recorrido, sino únicamente su evolución intermedia. El resultado es una trayectoria más densa, visualmente más continua y geoméricamente coherente, adecuada para su representación cartográfica y análisis posterior. En la Figura 3-39 se puede ver el resultado al generar 10 trayectorias sintéticas.

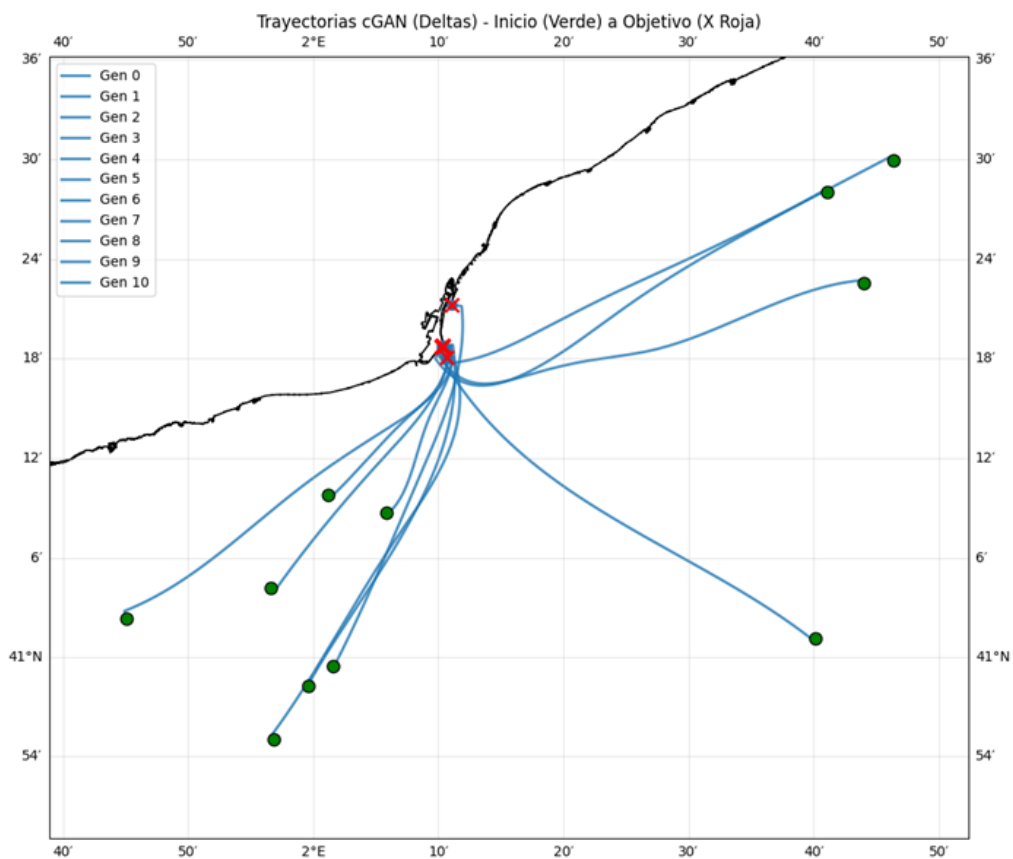


Figura 3-39 Generación de trayectorias sintéticas con cGAN (Fuente: propia)

3.5.1.6 Estimación de velocidades mediante *K-Nearest Neighbors* (KNN)

El algoritmo *K-Nearest Neighbors* (KNN) es un método de aprendizaje supervisado basado en la proximidad entre datos. Su funcionamiento consiste en identificar, para una muestra determinada, los K elementos más cercanos dentro del conjunto de datos según una métrica de distancia previamente

definida. A partir de estos vecinos más próximos, se pueden estimar características asociadas a la muestra, como una clase o un valor numérico. En el presente trabajo este algoritmo se va a emplear para la simulación de velocidades empleadas en las trayectorias generadas por cGAN.

Al igual que en los casos anteriores, lo primero que se debe hacer es el filtrado de información de los CSV. Para ello hace un barrido de la base de datos, extrayendo los datos que interesan para el estudio: latitud, longitud y velocidad. Estos datos se unifican bajo una única denominación, generando un nuevo conjunto de datos. De otra manera, cada punto geográfico real (longitud y latitud) tendrá asignada una velocidad real.

Además, se realiza un filtrado de aquellas velocidades que no son realistas, eliminando aquellas de menos de 0,5 nudos y más de 30 nudos. De esta manera de las 3785 trayectorias disponibles se obtienen un total de 2452479 puntos válidos de los 2454266 procesados.

Posteriormente, se construye y entrena un modelo de tipo *k-nearest neighbors* (KNN) que actúa como estimador de velocidad en función de la posición geográfica. Este modelo se utilizará más tarde como uno capaz de asignar valores de velocidad plausibles a puntos de trayectoria en función de su localización. Dicho de otra manera, dada una latitud y longitud, el modelo buscará puntos reales cercanos y estimará cuál debería ser la velocidad en esa zona.

En este caso el algoritmo empleará para su estimación los 5 puntos más cercanos, influyendo en la velocidad elegida, como es lógico, los vecinos más cercanos. Cuanto más cerca esté un punto AIS, mayor peso tendrá en la predicción, siendo la velocidad estimada una media ponderada por distancia. Se emplea para el cálculo 5 vecinos ($K=5$) con el objetivo de mantener el comportamiento local y no suavizar demasiado, así como para mantener diferencias entre zonas.

Con el objetivo de mejorar la coherencia temporal del perfil de velocidades asociado a cada trayectoria sintética, se implementa una función de suavizado basada en el filtro de Savitzky-Golay. El filtro de Savitzky-Golay es un algoritmo de suavizado digital basado en el ajuste polinómico local mediante mínimos cuadrados. Su objetivo es reducir el ruido de alta frecuencia en una señal sin distorsionar sus tendencias originales (como la amplitud y la posición de los picos).

El filtro recibe un vector con las velocidades a lo largo de la trayectoria, comprobando primeramente que la trayectoria tenga 100 puntos, en caso de tener menos, no aplica ningún suavizado y devuelve la velocidad original; aunque realmente se trata de una seguridad adicional, ya que las trayectorias previas generadas cuentan con 200 puntos. Posteriormente aplica el filtro obteniendo una tendencia general de perfil de velocidad sin eliminar la variabilidad de la velocidad.

Esto se aplica debido a que, en las trayectorias generadas, la velocidad puede cambiar de forma demasiado abrupta entre celdas consecutivas. Este filtro elimina esos saltos poco realistas, generando un perfil de velocidad más continuo y físicamente coherente.

A continuación del proceso de suavizado, se realiza una discretización de la velocidad continua generada, limitando sus incrementos a intervalos de 1 nudo. Con este proceso de discretización se busca mejorar la estabilidad y el realismo de las velocidades generadas, evitando pequeñas oscilaciones que carecen de significado desde el punto de vista físico y que podrían introducir ruido innecesario en análisis posteriores. Asimismo, la discretización facilita la interpretación y comparación de los resultados, especialmente cuando se contrastan las trayectorias sintéticas con los datos históricos reales.

3.5.1.7 Generación final de trayectorias cGAN + KNN

Con todo esto, y aplicando límites de velocidad máximos (24 nudos) y mínimos (0,7 nudos) se podrá asignar a la trayectoria sintéticas velocidades adecuadas de la siguiente manera:

- 1) Para cada trayectoria generada, se realiza inicialmente una predicción directa de la velocidad en cada punto mediante el modelo KNN, utilizando exclusivamente la posición geográfica como variable de entrada.

- 2) Se aplica un proceso de recorte que limita las velocidades al rango físicamente admisible definido entre 0,7 y 24 nudos, evitando la aparición de valores negativos o excesivamente reducidos o elevados.
- 3) Se suaviza y discretiza la velocidad según la trayectoria.
- 4) Cada trayectoria final se almacena como una secuencia de triples (latitud, longitud, velocidad), obteniendo así un conjunto completo de trayectorias sintéticas con geometría y dinámica coherentes, listas para su análisis, visualización o exportación.

Posteriormente, se realiza una función encargada de calcular el rumbo entre puntos consecutivos de cada trayectoria y añade esa información como una nueva variable. Obteniendo una trayectoria con 4 dimensiones: Latitud, Longitud, Velocidad y Rumbo.

En la Figura 3-40 se puede ver la generación de 400 trayectorias generadas con cGAN junto con KNN.

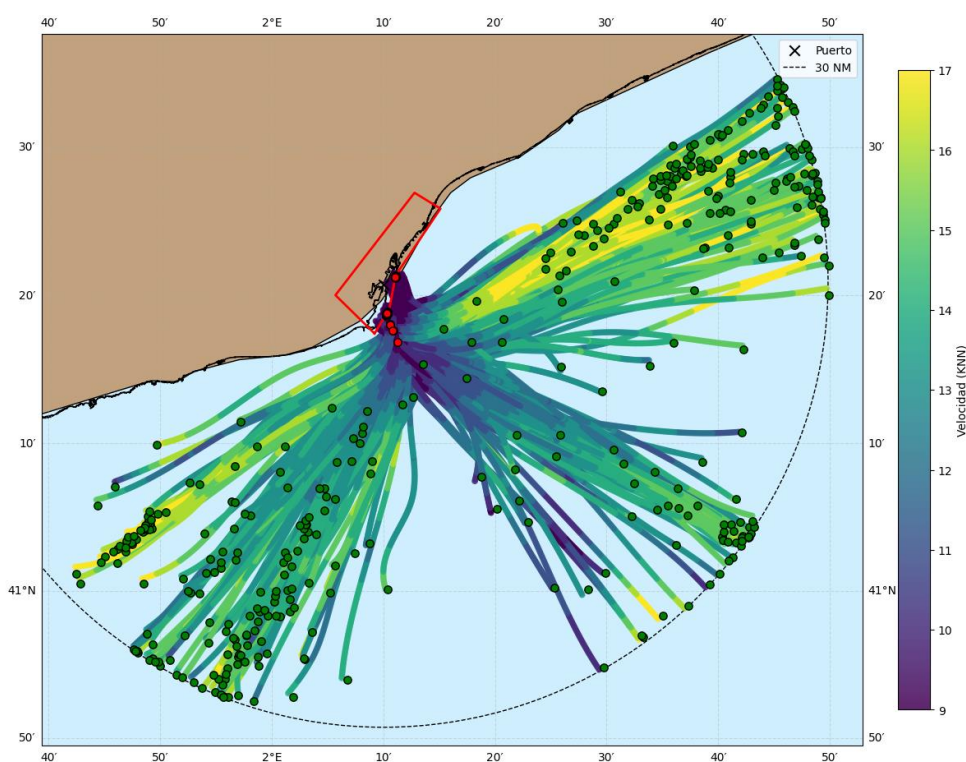


Figura 3-40 Trayectorias generadas con cGAN junto con KNN (Fuente: propia)

De manera análoga se generan las trayectorias por separado, mostrando, además, una gráfica que muestra la variación de su velocidad y de su rumbo. De esta manera, se podrá comparar el resultado de manera más precisa con las trayectorias reales y con las generadas con otros métodos. Además, se aplica de nuevo la conversión a una trayectoria de 200 puntos, como se hizo con anterioridad, para facilitar la comparación con otros métodos. Véase un ejemplo del resultado final en la Figura 3-41.

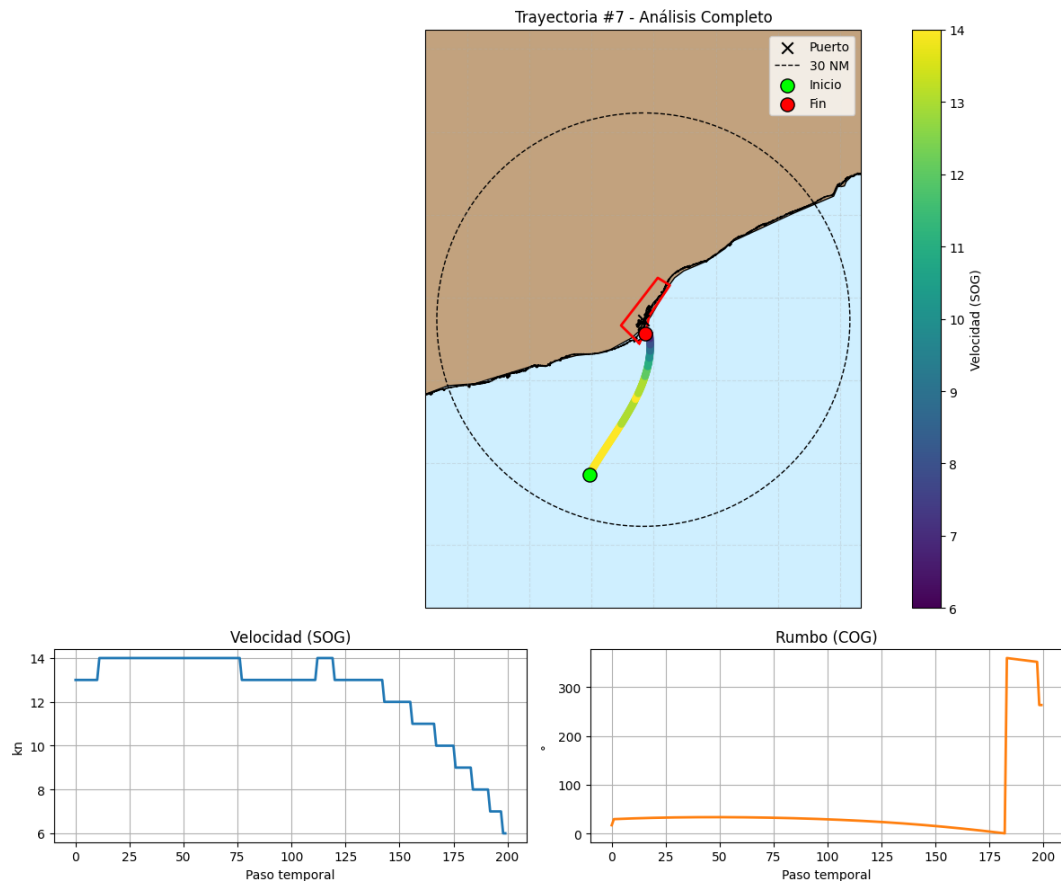


Figura 3-41 Trayectoria generada con cGAN y KNN junto con sus gráficas de velocidad y rumbo (Fuente: propia)

3.6 Generación conjunta de trayectoria y velocidad mediante redes generativas antagónicas condicionales (cGAN)

En este apartado se desarrolla un tercer enfoque para la generación de trayectorias sintéticas basado en redes generativas antagónicas condicionales (cGAN), que en este caso se emplearán tanto para la generación de trayectorias como la asignación de velocidades.

3.6.1 Buques tipo Cargo que entran en las inmediaciones del Puerto de Barcelona

3.6.1.1 Preparación del conjunto de datos para la cGAN espacio-dinámica

Al igual que en los anteriores métodos, lo primero que se debe realizar es la lectura de los datos de buques mercantes entrando en el puerto de Barcelona en los CSV correspondientes a los 5 años de estudio. Primeramente, se localizarán y contabilizarán los archivos de trayectorias que se utilizarán como entrada para la GAN que, como en los casos anteriores son 3785 derrotas de entrada o aproximación a puerto.

A continuación, se somete cada trayectoria a un proceso de limpieza y filtrado de calidad. En concreto, se eliminan aquellos puntos de la trayectoria donde la velocidad registrada sea anómala o indique que el buque está detenido (fuera del rango de 0.5 a 30.0 nudos). Tras aplicar este filtro, se verifica la viabilidad de la trayectoria: si la secuencia resultante conserva menos de 5 puntos válidos, se descarta por completo, ya que se considera que carece de información suficiente para ser reconstruida.

De esta manera, como en los métodos anteriores, se establecen tres parámetros de control:

Parámetro	Número	Descripción
Velocidad mínima	0,5 nudos	Velocidad mínima admisible
Velocidad máxima	30 nudos	Velocidad máxima admisible
Mínimo número de puntos	5 puntos	Número mínimo de puntos que deben tener las trayectorias.

Tabla 3-8 Parámetros de control

Para las trayectorias que superan este filtro de calidad, se procede a estandarizar su longitud para que todas posean exactamente el mismo número de puntos (definido por el parámetro de la red), garantizando que cada punto conserve sus tres variables: latitud, longitud y velocidad. Si la trayectoria ya tiene la longitud deseada, se mantiene sin modificaciones. En caso contrario, se construye un eje temporal normalizado entre 0 y 1 para la secuencia original y la nueva, aplicando una interpolación lineal independiente para cada una de las tres variables.

Una vez procesados todos los CSV de forma sistemática, se apilan en una única estructura de datos multidimensional (tensor). Esta matriz final constituye el conjunto de datos de entrenamiento directo para la red neuronal generativa.

Las trayectorias válidas se transforman en matrices con estructura $(T_0, 3)$, donde T_0 es la longitud original de la trayectoria y las tres columnas corresponden a la latitud, longitud y velocidad.

Con el fin de garantizar una longitud uniforme en todo el conjunto de datos, se aplica la función de remuestreo, que interpola cada trayectoria hasta una longitud fija de 100 puntos. Realizado esto, todas las trayectorias se agrupan en un tensor con forma:

$$(N, T_{GAN}, 3)$$

Donde:

Parámetro	Número	Descripción
N	3784	Número de trayectorias válidas tras filtrado
T_{GAN}	100	Longitud fija de la trayectoria
3	3	Latitud, longitud y velocidad

Tabla 3-9 Parámetros tras filtrado

Tras este filtrado para almacenamiento de datos, se realiza otro control, que tiene como objetivo el verificar la coherencia dinámica al final de cada trayectoria, comprobando si alguna termina con una velocidad excesivamente alta, en este caso marcada en 20 nudos. Esto se realiza ya que en un escenario de aproximación a puerto no es habitual que un buque de carga finalice su trayectoria con velocidades elevadas, resultando que sólo una de las trayectorias contenía datos superiores a este valor.

Tras estas comprobaciones de control, finalmente el presente estudio cuenta con un total de 3783 trayectorias disponibles para entrenar la cGAN.

Una vez llegado a este punto, la metodología será igual a la aplicada en el apartado anterior, sin embargo, en vez de solo tener en cuenta los datos de latitud y longitud a la hora de realizar la cGAN, también se tendrá en cuenta la velocidad:

3.6.1.1.1 Definición de las condiciones

Al igual que en el método anterior, se definen las variables condicionantes que se utilizarán en la cGAN. En este caso, se ha decidido emplear como condición a parte del punto inicial y el punto final de cada trayectoria, la velocidad inicial y final. Para construir esta condición, se extrae el primer y el último punto de cada secuencia y se agrupan en un único vector. De este modo, se obtiene una matriz denominada condiciones de dimensiones (3783, 6) donde:

Número	Descripción
3783	Número total de trayectorias
6	latitud inicial, latitud final, longitud inicial, longitud final, velocidad inicial y velocidad final

Tabla 3-10 Estructura de dos dimensiones

3.6.1.1.2 Cálculo de incrementos temporales

Como ya se hizo en el apartado anterior, se calculan las diferencias entre puntos consecutivos de cada trayectoria, conocidas como deltas. En este caso, cada trayectoria está formada por 100 posiciones sucesivas en latitud, longitud y velocidad. Al aplicar esta función, lo que se obtiene no son las posiciones absolutas, sino el cambio entre un punto y el siguiente. Es decir, para cada trayectoria se calculan 99 incrementos, donde cada uno indica cuánto se ha desplazado el buque en latitud, longitud y velocidad entre dos puntos consecutivos.

En la Figura 3-42 se puede apreciar el incremento de latitud y longitud entre dos puntos.

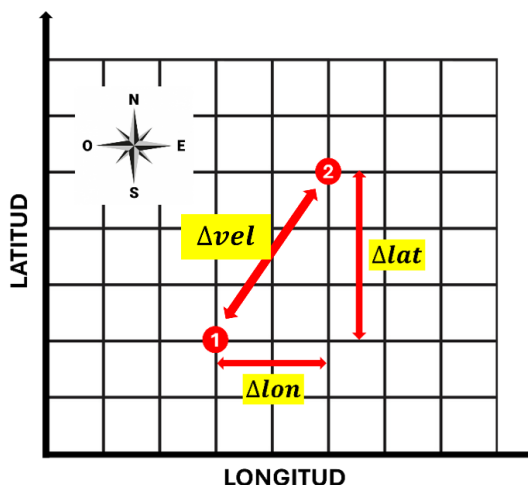


Figura 3-42 Incremento de latitud, longitud y velocidad (Fuente: propia)

Trabajar con deltas permite que el modelo aprenda patrones de movimiento relativos, lo que suele facilitar el entrenamiento y mejorar la estabilidad del modelo generativo.

3.6.1.1.3 Normalización de los datos

Con el objetivo de garantizar la estabilidad y la eficiencia del proceso de entrenamiento, se aplica una normalización Min-Max tanto a los deltas como a las condiciones asociadas a cada trayectoria. En concreto, se emplea, al igual que en el método anterior, el escalado al rango $[-1, 1]$, de forma que todas las variables queden acotadas dentro de un intervalo común.

La normalización es un paso fundamental en el entrenamiento de redes neuronales, ya que las variables originales presentan escalas significativamente distintas. Por ejemplo, la latitud y la longitud

se expresan en grados geográficos, mientras que la velocidad se mide en nudos y los deltas espaciales pueden tener magnitudes muy reducidas. Si estas variables se introdujeran directamente en el modelo sin escalado previo, aquellas con mayor rango numérico podrían dominar el proceso de aprendizaje, generando inestabilidad en los gradientes y dificultando la convergencia.

El escalado Min-Max transforma cada variable de acuerdo con sus valores mínimo y máximo observados en el conjunto de datos, mapeándolos al intervalo $[-1,1]$. De este modo, tanto los deltas $[\Delta Lat, \Delta Lon, \Delta Vel]$ como las condiciones $[Lat_{inicio}, Lon_{inicio}, Lat_{fin}, Lon_{fin}, Vel_{inicial}, Vel_{fin}]$ quedan expresados en una escala homogénea.

3.6.1.2 Generador condicional espacio-dinámico

Previo al comienzo del generador, se establece una función que define un término adicional de pérdida que penaliza cambios bruscos en la trayectoria generada, con el objetivo de favorecer movimientos más suaves y físicamente coherentes. Esta función se centra en los incrementos de latitud y longitud de los pasos temporales, dejando de lado a la velocidad. Esta operación mide cuánto cambia el desplazamiento entre un instante y el siguiente. Si la trayectoria es suave, estos cambios serán pequeños; si existen giros bruscos o movimientos erráticos, las diferencias serán mayores. Como resultado, esta función obtiene un escalar que cuantifica el grado de irregularidad geométrica de la trayectoria. Cuanto mayor sea este valor, mayor será la penalización aplicada al generador durante el entrenamiento. De este modo, el modelo no solo aprende a engañar al discriminador, sino también a generar trayectorias con una evolución espacial más continua y realista.

Antes de comenzar el entrenamiento, como se hizo en el método anterior, se deben ajustar una serie de datos para el buen funcionamiento del entrenamiento:

Parámetro	Traducción	Número
EPOCHS	Épocas	150
BATCH SIZE	Tamaño de Batch	64
LATENT DIMENSION	Dimensión del espacio latente	100
CONDITION DIMENSION	Dimensión de la condición	6
INPUT DIMENSION	Dimensión de entrada	3
HIDDEN DIMENSION	Dimensión del estado oculto	64
SEQUENCE LENGHT	Longitud de la secuencia	99
LEARNING RATE	Tasa de aprendizaje	0,0002
B1	B1	0,5
LAMBDA SPATIAL	Peso de penalización espacial	0,05

Tabla 3-11 Parámetros elegidos entrenamiento

Estos parámetros empleados ya fueron explicados en el anterior método de cGAN junto con KNN, por lo que no se volverá a profundizar en ellos. Sin embargo, cabe resaltar aquellos valores que si han cambiado respecto al apartado anterior:

- **CONDITION DIMENSION:** Se refiere al tamaño de la condición que, en este caso, es 6 $[Lat_{inicio}, Lon_{inicio}, Lat_{fin}, Lon_{fin}, Vel_{inicial}, Vel_{fin}]$. Esto significa que el modelo está “guiado” tanto por posición como por velocidad al inicio y al final.
- **INPUT DIMENSION:** Cada paso de la secuencia temporal, que en este caso es 3 $[\Delta Lat, \Delta Lon, \Delta Vel]$
- **LEARNING RATE:** Mediante la experimentación en este caso se ha decidido emplear una tasa de aprendizaje de 0,0002. Un valor excesivamente alto podría provocar oscilaciones fuertes y comportamientos inestables, mientras que un valor demasiado bajo ralentizaría el proceso de aprendizaje y dificultaría la convergencia. El valor adoptado representa, por tanto, un compromiso adecuado entre velocidad de entrenamiento y estabilidad numérica, coherente con la escala normalizada de los datos y con prácticas habituales en arquitecturas adversariales.

3.6.1.3 Arquitectura del discriminador

El discriminador es la red encargada de evaluar si una trayectoria es real o generada por el modelo. Para ello, recibe como entrada una secuencia de desplazamientos espaciales (deltas de latitud y longitud) y un vector de condición que contiene las coordenadas inicial y final de la trayectoria. La parte principal del modelo es una LSTM bidireccional de dos capas, que analiza la secuencia temporal en ambos sentidos (desde el inicio hasta el final y viceversa). Esto permite capturar mejor la coherencia global del movimiento. De la salida de la LSTM se extrae un vector que actúa como resumen completo de la trayectoria, es decir, una representación compacta de cómo se ha movido el buque.

En paralelo, el vector de condición (inicio y fin) se procesa mediante una pequeña red completamente conectada que lo transforma en una representación interna más rica. Posteriormente, se concatenan el resumen dinámico de la trayectoria y la representación de la condición, combinando así “lo que el barco hizo” con “lo que debía hacer”. Esta información conjunta pasa por un bloque final de capas lineales con activación no lineal y regularización mediante *dropout*, que reduce el riesgo de sobreajuste. Finalmente, una función sigmoide genera un valor entre 0 y 1 que representa la probabilidad de que la trayectoria analizada sea real.

3.6.1.4 Configuración y estrategia de entrenamiento

Con el objetivo de generar trayectorias sintéticas parecidas a las reales es esencial el entrenamiento completo de una red generativa adversaria condicional (cGAN) para la generación de trayectorias geoespaciales que, en este caso, incluyen información espacial y dinámica, concretamente desplazamientos en latitud, longitud y velocidad.

El entrenamiento de la cGAN se realiza durante 150 épocas, lo que significa que el modelo recorre completamente el conjunto de trayectorias 150 veces. En cada iteración se procesan grupos de 64 trayectorias (*batch*), formadas por secuencias de deltas normalizados junto con su condición asociada, compuesta por las coordenadas iniciales y finales. Para el aprendizaje adversarial se utiliza la función de pérdida *Binary Cross Entropy*, ya que el discriminador debe clasificar cada trayectoria como real o generada. Tanto el generador como el discriminador se optimizan mediante el algoritmo Adam con una tasa de aprendizaje inicial de 0,0002.

Como en el otro método, el modelo está compuesto por un generador y un discriminador entrenados, como se dijo anteriormente, de forma adversarial. El generador recibe como entrada un vector de ruido latente junto con una condición formada por las coordenadas y velocidades iniciales y finales de la trayectoria. A partir de esta información, el generador produce una secuencia temporal de incrementos espaciales y de velocidad utilizando una arquitectura basada en capas densas y redes LSTM, lo que le

permite capturar dependencias temporales de largo alcance. El discriminador, también basado en LSTM bidireccionales, evalúa la coherencia global de la trayectoria generada en función de la misma condición.

El entrenamiento adversario se complementa con términos adicionales de regularización incorporados en la función de pérdida del generador. Por un lado, se introduce un término de suavidad dinámica que penaliza variaciones bruscas de velocidad entre pasos temporales consecutivos. Por otro lado, se añade un término de suavidad espacial, indicado en la función explicada anteriormente, que penaliza cambios bruscos de dirección en los desplazamientos de latitud y longitud. Este último término actúa como una regularización geométrica, reduciendo oscilaciones locales (*zig-zag*) observadas en las trayectorias generadas y promoviendo trayectorias espacialmente más suaves, sin modificar explícitamente la velocidad.

Adicionalmente, con el objetivo de mejorar la estabilidad del entrenamiento y evitar que el discriminador domine al generador en fases avanzadas, se implementa un mecanismo de reducción progresiva del *learning rate*, reduciendo dicho valor a la mitad (0,0001) tras alcanzar la mitad del número total de épocas de entrenamiento (75). Este enfoque permite un aprendizaje más rápido de la estructura global de las trayectorias en las fases iniciales y un refinamiento más controlado en las fases finales.

Además, se añade la posibilidad de guardar el entrenamiento para recuperar lo aprendido posteriormente sin tener que repetir todo el proceso.

3.6.1.5 Reconstrucción de trayectorias absolutas

Una vez entrenado el modelo, el generador produce trayectorias sintéticas en forma de incrementos sucesivos normalizados de latitud, longitud y velocidad $[\Delta Lat, \Delta Lon, \Delta Vel]$. Estos valores representan variaciones relativas dentro de un rango escalado y no corresponden directamente a coordenadas geográficas ni a velocidades reales, por lo que no pueden utilizarse ni representarse de forma inmediata. Por este motivo, es necesario aplicar un proceso de reconstrucción que deshaga la normalización y acumule dichos incrementos a partir de las condiciones iniciales, obteniendo así trayectorias absolutas expresadas en latitud, longitud y velocidad en unidades físicas reales.

En primer lugar, se revierte la normalización aplicada durante el preprocesado tanto a los incrementos generados por la red como a las condiciones asociadas a cada trayectoria. De este modo, los valores dejan de estar expresados en el rango normalizado $[-1,1]$ y se recuperan en sus unidades físicas originales. Las condiciones desescaladas proporcionan el estado inicial y final completo de cada trayectoria, es decir, latitud, longitud y velocidad tanto en el punto de partida como en el punto de destino. Paralelamente, los deltas generados se transforman nuevamente a su escala real, permitiendo trabajar con incrementos físicos coherentes.

A continuación, se procede a la reconstrucción propiamente dicha. En primer lugar, se separan los incrementos espaciales (latitud y longitud) de los incrementos de velocidad. Mediante una suma acumulativa de los deltas espaciales se obtiene la evolución relativa del movimiento a lo largo del tiempo. Esta trayectoria relativa se traslada al sistema absoluto sumando el punto inicial real (latitud y longitud iniciales) lo que permite obtener la secuencia completa de coordenadas geográficas. De forma análoga, la velocidad absoluta se reconstruye acumulando los incrementos de velocidad a partir de la velocidad inicial, obteniéndose así un perfil dinámico coherente con el patrón aprendido por el generador.

Sin embargo, debido a la naturaleza probabilística del modelo generativo y al carácter acumulativo del proceso, puede aparecer una pequeña deriva espacial que provoque que el último punto reconstruido no coincida exactamente con el destino que se estableció en la condición. Para corregir esta desviación y garantizar la coherencia global del recorrido, se aplica un ajuste geométrico progresivo sobre las componentes de latitud y longitud. Este ajuste se basa en el cálculo del error total entre el punto final obtenido y el punto objetivo, distribuyéndose linealmente a lo largo de todos los instantes de la trayectoria. De esta manera, la corrección aplicada es prácticamente nula en los primeros puntos y

aumenta gradualmente hasta compensar completamente la desviación en el último, asegurando una llegada exacta al destino sin introducir discontinuidades artificiales. En cambio, la velocidad no se somete a un anclaje forzado, preservándose la evolución dinámica generada por el modelo.

Como resultado de este procedimiento, se obtienen trayectorias sintéticas compuestas por 100 puntos descritos mediante latitud, longitud y velocidad en unidades reales. Estas trayectorias mantienen la dinámica local aprendida por la red en términos de variaciones espaciales y de velocidad, al tiempo que respetan estrictamente las condiciones globales de inicio y fin impuestas durante la generación.

Al igual que en el método cGAN-KNN, se remuestran y suavizan las trayectorias para pasar de 100 a 200 puntos, con el objetivo de poder comparar las trayectorias generadas con otros métodos y con los datos reales AIS.

Además, como en el método anterior, se calcula el rumbo de cada trayectoria y se añade como nueva variable dinámica para, de este modo, poder realizar un análisis más exhaustivo de las trayectorias.

Por otro lado, con el objetivo de regular y estabilizar la velocidad final de las trayectorias generadas se limita la velocidad máxima a 24 nudos, actuando como umbral máximo admisible y la velocidad mínima a 0,7 nudos, evitando velocidades demasiado altas o bajas que no resultarían realistas en un contexto operativo. Las velocidades, al igual que con el método KNN, son sometidas a un proceso de suavizado y discretización.

Una vez realizados los últimos pasos del postprocesado, ya se pueden generar las trayectorias sintéticas que se emplearán para compararlas con otros métodos y las trayectorias reales AIS. De la misma manera que se hizo en casos anteriores, se visualizan las trayectorias en mapas tipo Mercator donde se marca la línea de costa de Barcelona. En la Figura 3-43 se pueden observar 400 trayectorias generadas mediante cGAN tanto para la trayectoria como la velocidad.

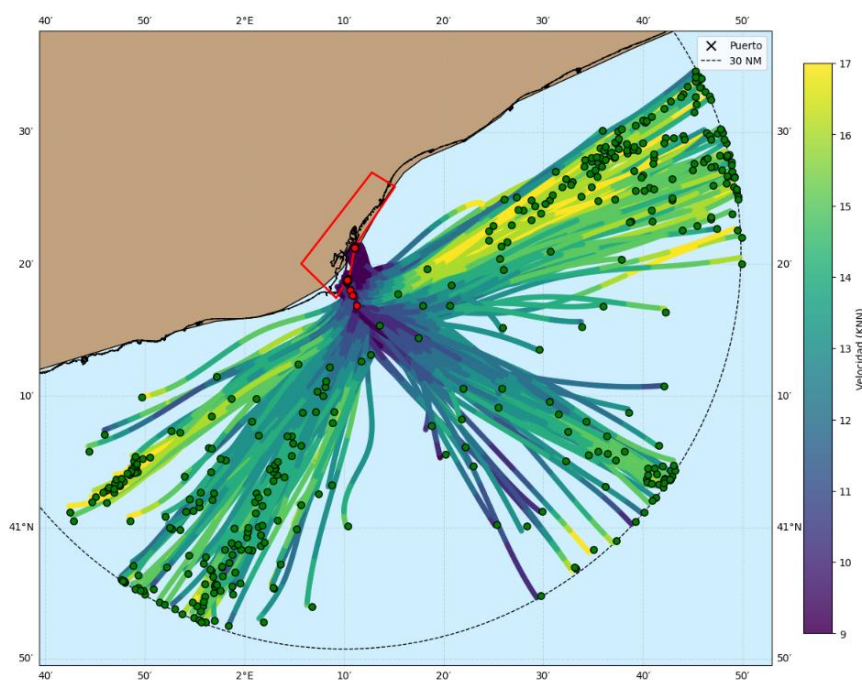


Figura 3-43 400 trayectorias generadas con cGAN (Fuente: propia)

Así mismo, y como se realizó en métodos anteriores, se añaden a las trayectorias generadas, gráficas con la variación de velocidad y rumbo para facilitar el posterior análisis y comparación. Esto se muestra en la Figura 3-44.

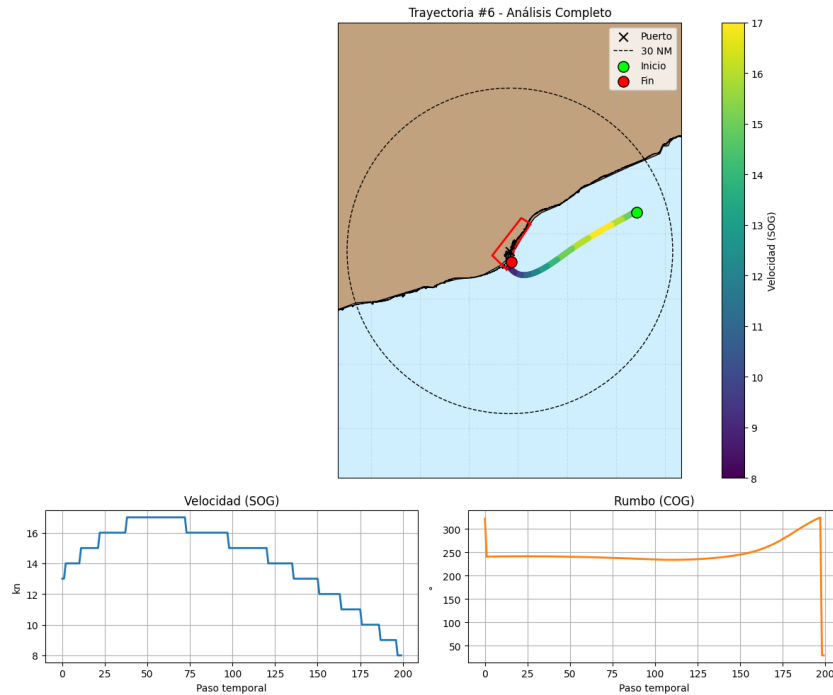


Figura 3-44 Trayectoria sintética generada con cGAN con gráficos de rumbos y velocidades (Fuente: propia)

3.7 Modelado de trayectorias de buques tipo Fishing mediante LSTM

3.7.1 Limitaciones del enfoque cGAN en buques pesqueros

Una vez implementado el modelo cGAN para la generación de trayectorias de buques tipo *Cargo* y analizados sus resultados, se planteó la posibilidad de aplicar el mismo enfoque a los buques pesqueros. Sin embargo, tras diversos intentos de adaptación y entrenamiento del modelo, los resultados obtenidos no alcanzaron el nivel de realismo esperado.

La principal dificultad radica en la naturaleza del comportamiento de los buques pesqueros. A diferencia de los cargueros, cuyas derrotas suelen ser más estructuradas y repetitivas, los pesqueros presentan trayectorias notablemente más irregulares y cambiantes, especialmente durante la fase de faena. Esta elevada variabilidad espacial y cinemática dificultó que la red neuronal aprendiera patrones suficientemente estables como para generar trayectorias coherentes.

Por este motivo, se opta por replantear el enfoque de generación, aplicando una red neuronal LSTM autorregresiva y apoyándose en una discretización espacial mediante cuadrículas H3. Este nuevo planteamiento permite estructurar mejor la información espacial y facilita la generación de trayectorias sintéticas que reproduzcan de manera más fiel el comportamiento observado en los datos AIS reales.

3.7.2 Justificación del modelo LSTM autorregresivo

Se emplea una red LSTM (Long Short-Term Memory), un tipo de red neuronal recurrente diseñada específicamente para el tratamiento de secuencias. La principal característica de la LSTM es su capacidad para mantener una memoria interna que se actualiza en cada paso temporal, permitiendo que el modelo conserve información relevante del pasado y la utilice para predecir el siguiente estado. En este contexto, la LSTM aprende patrones de transición espacial condicionados por el historial completo de la trayectoria, capturando dependencias temporales de medio y largo alcance. Esta capacidad resulta especialmente adecuada para modelar trayectorias de buques pesqueros, cuya dinámica espacial no depende únicamente de la posición actual, sino del comportamiento acumulado a lo largo del tiempo.

Como en todos los casos anteriores, los datos reales que se van a emplear para el entrenamiento del modelo se obtienen de los respectivos CSV de los pesqueros. En el caso de los pesqueros, los datos AIS

originales presentan una frecuencia de transmisión variable, dependiente de la velocidad y estado de navegación del buque. Esta irregularidad temporal puede dificultar el aprendizaje secuencial del modelo, ya que la red neuronal no recibe intervalos de tiempo homogéneos entre observaciones consecutivas. Para evitar este problema, cada trayectoria se ordena cronológicamente y se remuestrea en ventanas temporales de 2 minutos.

3.7.3 Discretización espacial mediante cuadrículas H3

Para la discretización espacial de las trayectorias AIS, al igual que en los apartados anteriores, se emplea el sistema jerárquico de cuadrículas hexagonales H3 tipo UBER. En el presente caso se fija una resolución espacial igual a 8, lo que determina el tamaño de las celdas hexagonales utilizadas para representar el espacio geográfico.

Inicialmente se realizaron pruebas con una resolución superior (9), al igual que en los cargueros, que proporciona un mayor nivel de detalle. Sin embargo, esta opción generaba un número muy elevado de celdas distintas y, en consecuencia, un aumento considerable en la dimensionalidad del problema. Esto incrementaba significativamente el coste computacional durante el entrenamiento del modelo LSTM, superando las capacidades prácticas del equipo empleado para el trabajo (GPU RTX-3060 con 12 Gb de RAM) e imposibilitando el correcto desarrollo de este.

Desde el punto de vista metodológico, la resolución 8 representa un compromiso adecuado entre precisión espacial y estabilidad del aprendizaje. En el caso de los buques pesqueros, cuyas trayectorias presentan una elevada variabilidad y cambios frecuentes de rumbo durante la fase de faena, una resolución demasiado alta provoca una fragmentación excesiva del espacio, generando secuencias muy erráticas y difíciles de generalizar por parte del modelo secuencial.

Adicionalmente, el uso de la misma resolución que en el método basado en mapas de calor permite realizar una comparación directa entre las trayectorias generadas mediante técnicas de Machine Learning y las obtenidas mediante el modelo probabilístico descrito anteriormente.

Como ya se hizo con anterioridad en otros modelos, se realiza una discretización espacial con el objetivo de transformar las coordenadas continuas de latitud y longitud en el sistema de cuadrículas hexagonales jerárquicas H3 en resolución 8.

En la práctica, cada punto AIS (latitud, longitud) se transforma en el identificador de la celda H3 a la que pertenece. De este modo, cada trayectoria real pasa a representarse como una secuencia ordenada de celdas hexagonales.

Con el objetivo de evitar incongruencias en el modelo, se aplica un proceso adicional de depuración con el objetivo de eliminar repeticiones consecutivas de una misma celda dentro de cada secuencia. Es decir, si dos puntos geográficos en una misma trayectoria se encuentran en la misma celda, el sistema se quedará con solo uno.

Así mismo, con el fin de evitar secuencias con muy poca información, se descartan aquellas trayectorias cuya longitud sea inferior a 10 celdas.

Con el fin de caracterizar las secuencias obtenidas tras la discretización espacial en celdas H3, se analiza la longitud de cada trayectoria, entendida como el número de celdas que componen la secuencia temporal de cada buque. Para ello se calculan los siguientes parámetros:

Número de celdas	Resultado (Nº de celdas)
Mínimo	5
Máximo	273
Medio	101.187386

Tabla 3-12 Número de celdas

3.7.4 Construcción del vocabulario espacial

Con el objetivo de preparar los datos para su utilización en el modelo LSTM, se identifica el conjunto total de celdas H3 distintas presentes en todas las trayectorias. Este conjunto constituye el vocabulario espacial del problema. En este caso, debido a que existen 2253 celdas distintas, el vocabulario tendrá 2253 “palabras”.

Dado que en el conjunto de datos aparecen 2253 celdas diferentes, el modelo debe elegir, en cada paso temporal, una de entre 2253 posibles opciones.

Esta elevada cantidad de clases o celdas tiene varias implicaciones desde el punto de vista del aprendizaje. En primer lugar, aumenta la complejidad del problema de clasificación, ya que el modelo debe discriminar entre un número elevado de estados espaciales posibles.

Además, la distribución espacial de las trayectorias no es uniforme. Algunas celdas aparecen con mucha frecuencia (por ejemplo, zonas próximas al puerto o áreas habituales de pesca), mientras que otras presentan una menor representación en los datos. Esta desigualdad puede dificultar el aprendizaje equilibrado del modelo, ya que las clases menos frecuentes disponen de menos ejemplos para su entrenamiento.

Por último, un mayor número de clases implica un incremento del coste computacional y de la cantidad de parámetros del modelo, especialmente en la capa de salida. No obstante, esta discretización detallada permite mantener un nivel de precisión espacial adecuado en la generación de trayectorias sintéticas.

De esta forma, el modelo no trabaja directamente con coordenadas continuas de latitud y longitud, sino con un conjunto limitado de celdas H3 que representan las diferentes zonas del área de estudio. Cada celda actúa como un estado dentro del espacio, lo que permite simplificar el problema y facilitar el aprendizaje de patrones de movimiento entre zonas.

Con el objetivo de preparar las secuencias para su utilización en el modelo LSTM, se construye un vocabulario espacial que incluye todas las celdas H3 distintas presentes en el conjunto de trayectorias.

Para ello, además de las palabras provenientes de celdas reales, se incorporan tres palabras especiales:

Palabra	Función
PAD	Se emplea para rellenar secuencias hasta que todas tengan la misma longitud.
BOS	Marca el inicio de una trayectoria
EOS	Marca el final de una trayectoria

Tabla 3-13 Palabras especiales

Estas palabras no son celdas reales, pero ayudan al modelo a entender la estructura de las secuencias.

El tamaño final del vocabulario corresponde al número total de celdas distintas más las tres palabras especiales, determinando así la dimensión del espacio de salida del modelo en un total de 2256 “palabras”.

Con el objetivo de adaptar las secuencias al formato requerido por la red LSTM, se procede a convertir cada trayectoria en una representación numérica de longitud fija. Esta longitud fija se estableció en la longitud máxima: 273 celdas.

Además, a cada celda se le otorga la palabra especial de inicio y final (BOS y EOS) para identificar los límites de la trayectoria. Por otro lado, las secuencias cuya longitud es inferior a la máxima se completan mediante la palabra de relleno (PAD), garantizando que todas las trayectorias tengan la misma dimensión.

3.7.5 Formulación del modelo LSTM autorregresivo

3.7.5.1 Planteamiento secuencial del problema y construcción del dataset de entrenamiento

Con el objetivo de entrenar el modelo en modo autorregresivo, cada trayectoria codificada se transforma en un par de secuencias desplazadas una posición. Concretamente, la secuencia de entrada se define como todos los elementos excepto el último, mientras que la secuencia objetivo corresponde a la misma trayectoria desplazada una posición hacia adelante.

De este modo, el modelo recibe como entrada una sucesión de celdas H3 y aprende a predecir, en cada paso, cuál es la siguiente celda real de la trayectoria. Este planteamiento permite modelar la probabilidad condicional de cada estado espacial en función del historial previo, capturando así la naturaleza secuencial del movimiento del buque.

De este modo, el modelo recibe como entrada una subsecuencia $(c_0, c_1, \dots, c_{t-1})$ y aprende a predecir el siguiente estado c_t .

Las trayectorias se agrupan en lotes de 64 secuencias durante el entrenamiento y se barajan en cada época, lo que mejora la estabilidad del aprendizaje y favorece la generalización del modelo.

3.7.5.2 Arquitectura del modelo

El modelo empleado para la generación de trayectorias sintéticas se basa en una red LSTM autorregresiva compuesta por tres bloques principales: una capa de *embedding*, una red LSTM y una capa lineal de salida.

3.7.5.2.1 Capa de *embedding*

En primer lugar, cada celda H3 representada como un índice entero se transforma mediante una capa de *embedding* en un vector de dimensión 128. Esta transformación es necesaria porque los índices numéricos utilizados para identificar las celdas no contienen información estructural por sí mismos, sino que actúan únicamente como etiquetas.

La capa de *embedding* permite asociar a cada celda una representación vectorial que el modelo aprende automáticamente durante el entrenamiento. De este modo, celdas que aparecen en contextos espaciales similares (por ejemplo, zonas con patrones de movimiento parecidos) pueden acabar teniendo representaciones similares dentro del modelo. De este modo, el *embedding* puede aprender que, aunque se trate de celdas diferentes desde el punto de vista geográfico, ambas desempeñan un papel similar dentro de la trayectoria del buque, al estar asociadas a patrones de movimiento comparables.

A modo de ejemplo, pueden considerarse dos zonas de pesca distintas dentro del área de estudio, una situada al norte y otra al sur. Desde el punto de vista geográfico se trata de áreas diferentes; sin embargo, el comportamiento de los buques en ambas puede ser muy similar. En las dos zonas los barcos tienden a reducir su velocidad, presentan movimientos oscilatorios característicos de la fase de faena y muestran patrones de transición espacial parecidos. Aunque las celdas H3 que representan estas zonas sean distintas, el modelo puede aprender que desempeñan un papel similar dentro de la dinámica de la trayectoria. Como consecuencia, la capa de *embedding* tenderá a asignarles representaciones vectoriales similares, situándolas próximas entre sí en el espacio de 128 dimensiones utilizado por la red.

3.7.5.2.2 Red LSTM (Long Short-Term Memory)

Una vez obtenidas las representaciones vectoriales de cada celda mediante la capa de *embedding*, la secuencia resultante se introduce en una red LSTM compuesta por dos capas y con una dimensión interna de 256 unidades ocultas.

La LSTM procesa la trayectoria de forma secuencial, es decir, analiza cada paso temporal teniendo en cuenta no solo el estado actual, sino también la información acumulada de los pasos anteriores. Para ello, mantiene un estado interno que actúa como memoria dinámica. En cada instante, esta memoria se actualiza incorporando la información relevante del nuevo vector de entrada y reteniendo aquellos aspectos del historial que el modelo considera importantes.

La dimensión interna de 256 unidades ocultas, determina la capacidad del modelo para representar el contexto acumulado de la trayectoria. En otras palabras, en cada paso temporal la red genera una representación interna de 256 valores que resume el comportamiento previo del buque hasta ese momento.

3.7.5.2.3 Capa lineal de salida

En cada instante temporal, la red LSTM genera un vector interno de dimensión 256 que resume el contexto acumulado de la trayectoria hasta ese momento. Esta representación no constituye todavía una predicción directa de la siguiente celda, sino una codificación del estado dinámico del buque basada en su historial de movimiento.

Dicho vector se proyecta posteriormente mediante una capa lineal sobre el espacio completo del vocabulario, cuya dimensión coincide con el número total de celdas posibles. El resultado es un vector de puntuaciones, en el que cada componente se asocia a una celda H3 concreta.

Estas puntuaciones se transforman en probabilidades mediante una función de normalización durante el proceso de entrenamiento, obteniéndose así una distribución de probabilidad sobre todas las posibles celdas siguientes. De este modo, el modelo estima la probabilidad condicional de cada estado espacial dado el historial previo de la trayectoria.

Este procedimiento implementa un enfoque claramente autorregresivo, ya que la generación de la trayectoria se realiza paso a paso: en cada instante se predice la siguiente celda en función de las anteriores, incorporándola posteriormente a la secuencia para continuar el proceso.

3.7.6 Proceso de entrenamiento

El entrenamiento del modelo se realiza durante 60 épocas, entendiendo por época una pasada completa por el conjunto de datos de entrenamiento. En cada iteración, las trayectorias se procesan en lotes de 64 secuencias, lo que permite un aprendizaje más eficiente y estable.

Para cada lote, el modelo genera una predicción sobre la siguiente celda en cada paso temporal de la secuencia. Estas predicciones se comparan con los valores reales mediante una función de pérdida, que cuantifica el error cometido por el modelo y guía la actualización de sus parámetros.

En este caso se emplea la entropía cruzada (*CrossEntropyLoss*), adecuada para problemas de clasificación multiclase como el presente, en el que el modelo debe seleccionar una celda entre un conjunto amplio de posibles estados espaciales. En cada paso temporal, el modelo produce una distribución de probabilidad sobre todas las celdas del vocabulario. La celda real observada en los datos puede interpretarse como una distribución ideal en la que la probabilidad es 1 para la celda correcta y 0 para el resto. La entropía cruzada mide la discrepancia entre ambas distribuciones: si el modelo asigna alta probabilidad a la celda correcta, la pérdida será baja; en cambio, si concentra la probabilidad en celdas incorrectas, la pérdida aumenta. De este modo, la función de pérdida incentiva que el modelo aprenda a asignar la mayor probabilidad posible al estado real siguiente.

Dado que la salida del modelo tiene dimensión tridimensional (tamaño del lote, longitud de la secuencia y tamaño del vocabulario), se reorganizan los tensores para adaptarlos al formato requerido por la función de pérdida, de manera que cada paso temporal contribuya individualmente al cálculo del error global.

Posteriormente, se aplica el algoritmo de optimización Adam para actualizar los parámetros del modelo en función del gradiente calculado. Este proceso se repite en cada época, permitiendo que la red reduzca progresivamente la pérdida media y mejore su capacidad para predecir las transiciones espaciales entre celdas.

3.7.7 Procedimiento autorregresivo de generación

Una vez entrenado el modelo LSTM, la generación de trayectorias sintéticas se realiza de forma autorregresiva, es decir, construyendo la secuencia paso a paso a partir de un estado inicial. El proceso comienza con el token de inicio “BOS”, que indica el comienzo de la trayectoria.

En cada paso, el modelo recibe la secuencia generada hasta ese momento y produce una distribución de probabilidad sobre todas las posibles celdas H3 del vocabulario. A partir de esta distribución se selecciona la siguiente celda, introduciendo un cierto grado de aleatoriedad controlado mediante el parámetro *temperature*. Este parámetro permite ajustar el equilibrio entre trayectorias más deterministas (si se favorecen las opciones más probables) y trayectorias más variadas (si se permite mayor dispersión en la selección). En el presente caso, mediante experimentación, se estableció este parámetro en 0,8.

La generación continúa de forma iterativa, incorporando cada nueva celda a la secuencia y utilizándola como contexto para la siguiente predicción. De este modo, la trayectoria se construye progresivamente en función del historial acumulado, manteniendo la coherencia espacial aprendida durante el entrenamiento.

Además, se incorporan restricciones geométricas relacionadas con el polígono que delimita el puerto. El modelo solo permite finalizar la trayectoria cuando se cumplen ciertas condiciones:

1. Que el buque haya abandonado el polígono del puerto.
2. Que haya alcanzado una distancia mínima respecto al puerto (0,003 en grados geográficas, aproximadamente 330 metros).
3. Regresar al polígono y recorrerlo durante mínimo 5 celdas.

Esta lógica adicional evita terminaciones prematuras y garantiza que las trayectorias generadas reproduzcan un patrón completo de salida, actividad en mar abierto y retorno al puerto.

De esta forma, la generación combina el comportamiento aprendido por la red neuronal con reglas espaciales que aseguran la coherencia operativa de las trayectorias sintéticas.

La longitud máxima de las trayectorias generadas se fija en 273 celdas, valor correspondiente a la trayectoria más larga del conjunto de entrenamiento. No obstante, la generación puede finalizar antes si se cumplen las condiciones de retorno al puerto establecidas en el algoritmo. En consecuencia, las trayectorias sintéticas presentan una longitud variable, acotada superiormente por dicho valor máximo.

Posteriormente, se transforma la trayectoria de celdas H3 a puntos geográficos (latitud y longitud) para que puedan ser representados en un mapa, excluyendo las palabras especiales (BOS, EOS y PAD), ya que no representan posiciones geográficas reales.

El resultado es un conjunto ordenado de coordenadas que permite visualizar la trayectoria sintética sobre un mapa, manteniendo la coherencia con la discretización espacial utilizada durante el entrenamiento del modelo.

3.7.8 Comparación visual entre trayectorias reales y sintéticas

Con el fin de evaluar cualitativamente el comportamiento del modelo, se realiza una comparación visual entre trayectorias reales y trayectorias sintéticas generadas. Para ello, se representan sobre el mismo mapa 50 trayectorias reales (en color gris) y 50 trayectorias generadas por el modelo (en color rojo). Tal y como se muestra en la Figura 3-45.

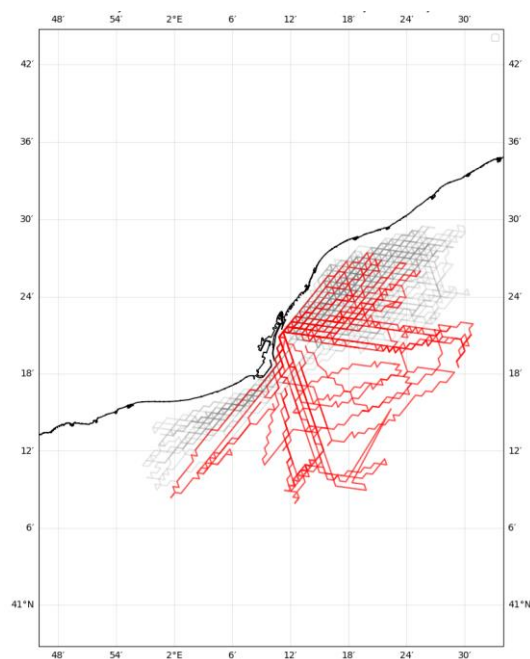


Figura 3-45 comparación de 50 trayectorias sintéticas con 50 trayectorias reales (Fuente: propia)

Esta visualización permite analizar si el modelo es capaz de reproducir los patrones espaciales característicos observados en los datos AIS reales, tales como los corredores habituales de salida y retorno al puerto, la dispersión en las zonas de faena y la coherencia geométrica del movimiento.

La superposición de ambas tipologías de trayectorias permite verificar de forma cualitativa que las trayectorias sintéticas mantienen una estructura espacial consistente con la dinámica real del tráfico pesquero en el área de estudio.

3.7.9 Suavizado y remuestreo espacial de trayectorias

Como en casos anteriores, se realiza un suavizado de la trayectoria generada, con el objetivo de que sea lo más real posible, así como para estandarizar el número de puntos que debe tener cada trayectoria para así facilitar su comparación con otros métodos.

En primer lugar, se calcula la distancia acumulada entre puntos consecutivos empleando la fórmula de *Haversine*, anteriormente empleada, que permite estimar distancias sobre la superficie terrestre considerando un radio medio de la Tierra igual a 6.371.000 metros. A partir de las coordenadas en latitud y longitud, se obtienen las diferencias angulares entre puntos sucesivos y se transforma dicha variación en distancia lineal en metros. El resultado es un vector de distancia acumulada que describe el recorrido total desde el punto inicial hasta cada posición intermedia de la trayectoria.

A continuación, la trayectoria se vuelve a definir utilizando la distancia acumulada como referencia, de forma que el recorrido se analice en función de los metros realmente avanzados y no únicamente según el orden de los puntos. Sobre esta nueva parametrización se ajusta una *spline* bidimensional que describe de manera conjunta la evolución de la latitud y la longitud, obteniendo así una versión suavizada del trayecto.

El parámetro de suavizado empleado ($s=0,0005$) permite reducir pequeñas irregularidades o “picos” en la trayectoria sin modificar de forma significativa su forma general. Antes de aplicar la *spline*, se eliminan los puntos consecutivos repetidos, ya que pueden generar problemas durante el ajuste y afectar a la estabilidad del cálculo.

Una vez ajustada la *spline*, se generan primero 2000 puntos a lo largo de la curva suavizada. Esto se hace para poder calcular con mayor precisión la distancia real que recorre la trayectoria una vez suavizada. Sobre estos puntos se vuelve a aplicar la fórmula de *Haversine* para obtener la distancia acumulada en metros.

Después, en lugar de quedarnos con esos 2000 puntos, se seleccionan únicamente 200 puntos repartidos de forma uniforme según la distancia recorrida. Es decir, los nuevos puntos no están separados por el mismo índice, sino por la misma distancia en metros a lo largo del trayecto.

Por último, se reconstruye la trayectoria final utilizando las coordenadas correspondientes a esas distancias seleccionadas y se asegura que el primer y el último punto coincidan exactamente con los originales. De esta forma, se obtiene una trayectoria suavizada formada por 200 puntos aproximadamente equiespaciados en metros, lo que facilita su comparación con otras trayectorias y mejora su visualización en el mapa.

Con el objetivo de realizar una comparación geométrica homogénea entre trayectorias reales y generadas, las trayectorias reales discretizadas en celdas H3 se transforman primero a coordenadas geográficas (latitud y longitud). Posteriormente, se aplica el mismo procedimiento de suavizado mediante *spline* empleado en las trayectorias generadas.

El resultado se muestra en la Figura 3-46, donde se aprecian las trayectorias sintéticas suavizadas comparadas con las reales. Se observa que las trayectorias generadas y suavizadas reproducen de forma coherente el patrón general de salida desde el puerto, navegación en mar abierto y posterior retorno. Los puntos iniciales y finales de las trayectorias sintéticas se concentran en el interior del polígono portuario, lo que confirma el correcto funcionamiento de las condiciones de finalización implementadas en el algoritmo de generación.

La superposición con las trayectorias reales permite comprobar que los corredores espaciales y la dispersión en la zona de faena son consistentes con los datos AIS originales, evidenciando la capacidad del modelo para capturar patrones espaciales relevantes.

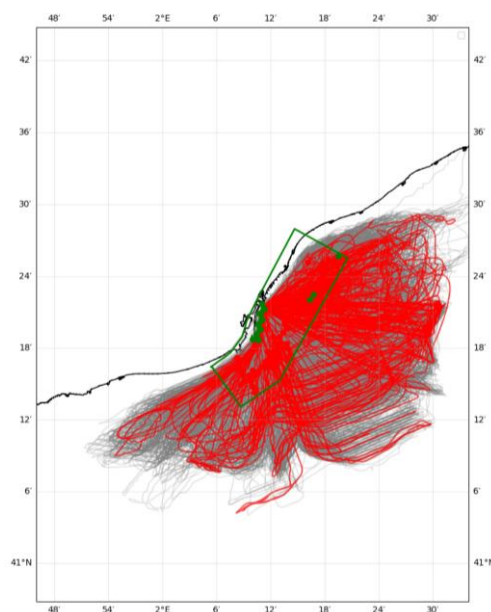


Figura 3-46 trayectorias sintéticas suavizadas comparadas con trayectorias reales (Fuente: propia)

En la Figura 3-47 se representan todas las trayectorias reales suavizadas (en gris) y un subconjunto de 50 trayectorias generadas (en rojo) seleccionadas del total de 200 generadas por el modelo. Esta selección permite una visualización más clara sin saturar excesivamente la figura.

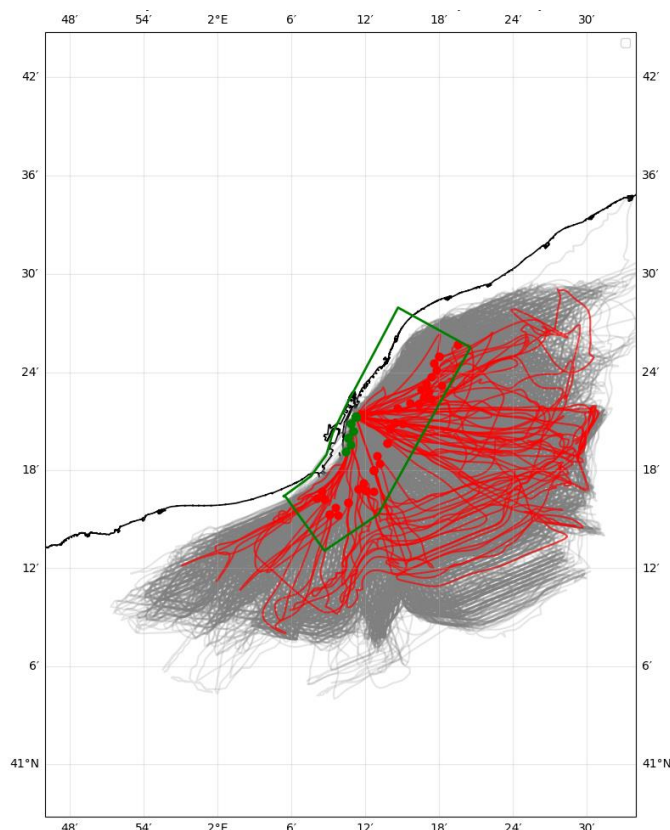


Figura 3-47 50 trayectorias sintéticas en comparación con todas las reales (Fuente: propia)

Se observa que las trayectorias sintéticas parten del interior del polígono portuario, se alejan hacia mar abierto y regresan al puerto, reproduciendo el patrón operativo característico de los buques pesqueros. La superposición con las trayectorias reales permite verificar la coherencia espacial del modelo y la similitud en los corredores de navegación.

Para la asignación de velocidades, como se hizo y se explicó con anterioridad, se emplea el modelo de tipo *K-Nearest Neighbors* (KNN), a partir de trayectorias de buques previamente procesadas y almacenadas en archivos CSV.

Así mismo se incorpora el cálculo del rumbo a las trayectorias generadas, con el objetivo de enriquecer la información dinámica asociada a cada punto de la trayectoria y así poder añadir gráficos al resultado.

En la Figura 3-48 se muestra la generación de trayectorias con sus respectivas velocidades asignadas, integrando de forma conjunta la información geométrica, dinámica y contextual del entorno marítimo.

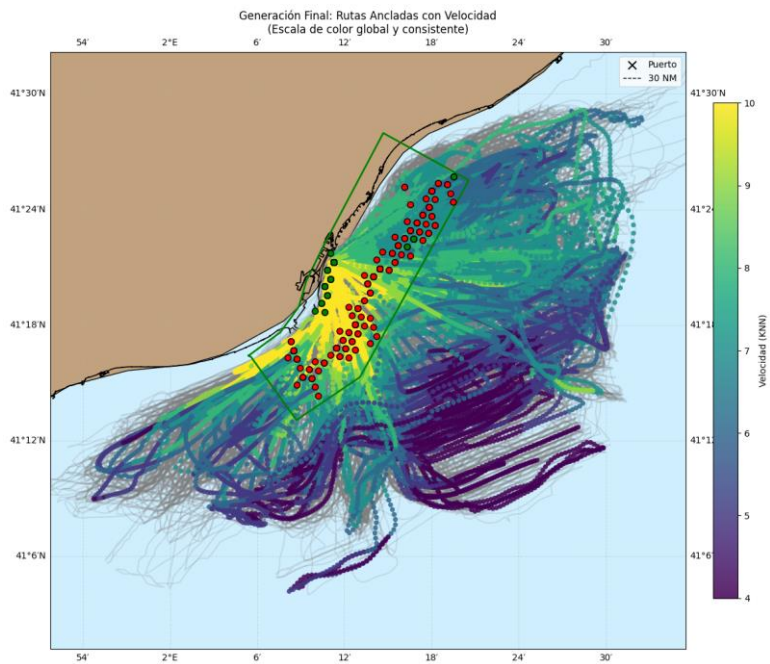


Figura 3-48 Rutas generadas con asignación de velocidad (Fuente: propia)

En la Figura 3-49 se muestra el resultado final de una trayectoria generada, añadiéndole como ayuda visual sus respectivos gráficos de velocidad y rumbo.

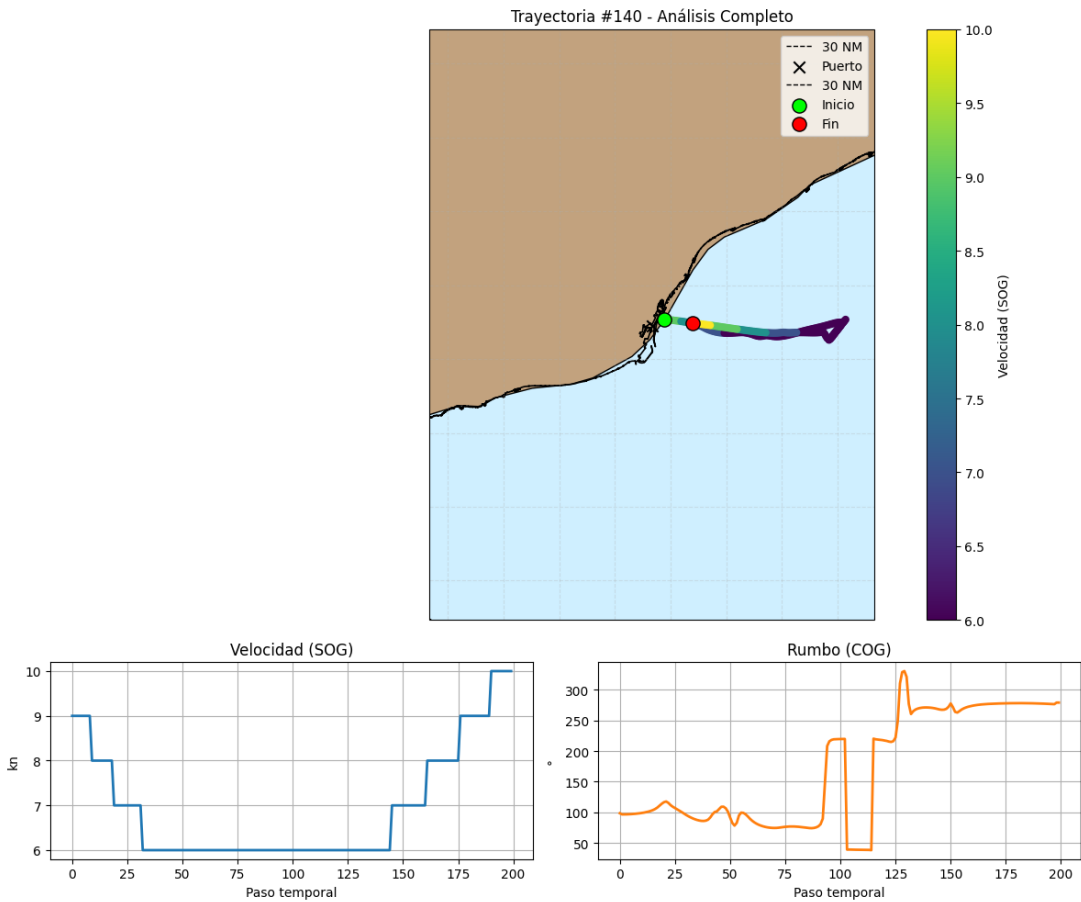


Figura 3-49 Presentación de trayectoria sintética completa (Fuente: propia)

4 RESULTADOS Y DISCUSIÓN

Una vez implementadas las distintas metodologías de generación de trayectorias sintéticas a partir de datos AIS reales, el presente capítulo tiene como objetivo evaluar de forma cuantitativa y comparativa su desempeño, con el fin de determinar cuál realiza la mejor aproximación para cada tipología de buque analizada. En particular, se pretende identificar el método más eficaz tanto para buques tipo *Cargo*, caracterizados por trayectorias estructuradas y previsibles, como para buques tipo *Fishing*, cuyo comportamiento presenta mayor variabilidad espacial y cinemática.

La validación se realiza mediante el contraste entre trayectorias reales y sintéticas, empleando métricas geométricas y dinámicas que permitan medir objetivamente la similitud entre ambas. Se analizan la coherencia espacial de las rutas generadas, la conservación de patrones de navegación, la distribución del tráfico y la cinemática asociada a velocidades y transiciones. Este enfoque garantiza que la evaluación no se limite al realismo visual, sino que se base en criterios cuantificables.

A partir de este análisis comparativo, se identifican las fortalezas y limitaciones de cada modelo (basado en discretización espacial H3 y cadenas de Markov, cGAN con KNN, generación conjunta espacio-dinámica mediante cGAN y modelo LSTM) con el propósito de seleccionar el método más robusto y adecuado para cada tipo de buque.

4.1 Métricas para la comparación de trayectorias

Con el objetivo de evaluar el grado de realismo de las trayectorias sintéticas generadas a lo largo del desarrollo del trabajo, resulta necesario establecer un conjunto de métricas que permitan comparar de forma objetiva dichas trayectorias con las trayectorias reales obtenidas a partir de los datos AIS. Estas métricas, permiten cuantificar hasta qué punto los modelos desarrollados son capaces de reproducir los patrones espaciales y dinámicos observados en el tráfico marítimo real.

Las métricas empleadas se centran, principalmente, en la comparación de la distribución espacial de las trayectorias, la coherencia de los patrones de movimiento y la similitud entre trayectorias reales y sintéticas. Este análisis se realiza mediante dos enfoques. Por un lado, se utilizan representaciones visuales que permiten comparar de forma directa el comportamiento global de los conjuntos de trayectorias. Por otro lado, se analizan determinadas magnitudes estadísticas asociadas a las trayectorias, lo que permite evaluar de forma cuantitativa el grado de similitud entre los datos generados y los datos reales.

Estas métricas se aplicarán de forma independiente a las dos tipologías de buques consideradas, buques tipo *Cargo* y buques tipo *Fishing*. Todo esto con el objetivo de evaluar el rendimiento de los distintos modelos generativos desarrollados. De este modo, se analizará hasta qué punto las trayectorias sintéticas generadas reproducen de forma realista los patrones de navegación de los datos AIS originales,

permitiendo identificar qué metodología resulta más adecuada para representar el comportamiento de cada tipo de tráfico marítimo.

4.1.1 Distancia de Fréchet

La distancia de Fréchet es una medida matemática que se utiliza para cuantificar la similitud entre dos curvas teniendo en cuenta tanto la forma como el orden de los puntos que las componen. A diferencia de otras métricas de distancia, la distancia de Fréchet no solo considera qué puntos están más cerca entre sí, sino cómo se recorren las curvas, preservando su estructura y continuidad, tal y como se puede observar en la Figura 4-1.

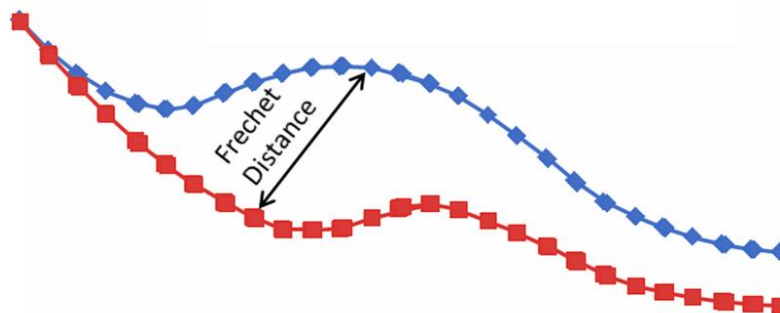


Figura 4-1 Ejemplo de distancia Fréchet [72]

De forma intuitiva, la distancia de Fréchet suele explicarse mediante la analogía de una persona paseando a un perro (Figura 4-2): la persona recorre una curva y el perro recorre otra distinta, pudiendo variar la velocidad de ambos, pero sin retroceder. La distancia de Fréchet es la longitud mínima de la correa necesaria para que ambos puedan completar sus trayectorias desde el inicio hasta el final sin separarse más de esa distancia máxima en ningún instante. Esta interpretación muestra que la métrica tiene en cuenta la correspondencia secuencial entre los puntos de ambas curvas, no solo su proximidad geométrica en cada punto [70].

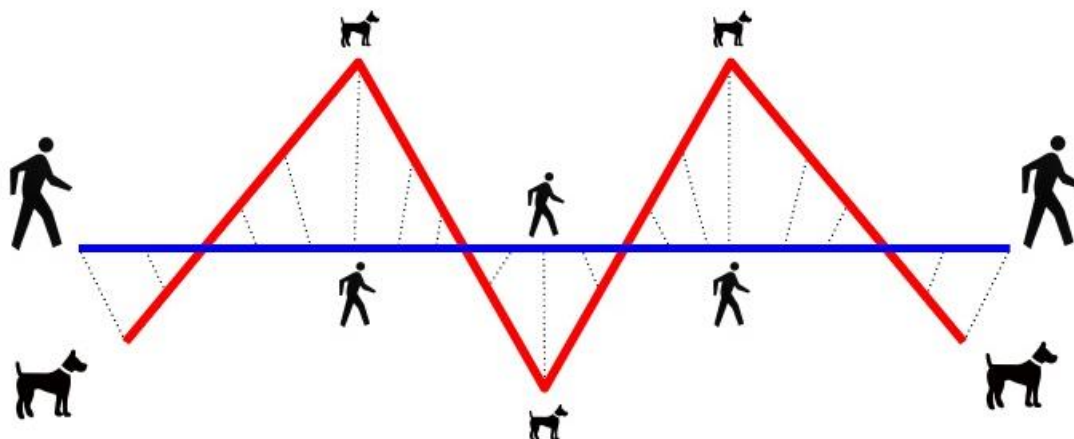


Figura 4-2 Distancia Fréchet semejada a un perro y una correa[72]

La distancia Fréchet se utiliza para medir hasta qué punto los resultados generados por un modelo se parecen a los datos reales, comparando cómo se distribuyen estadísticamente ambos conjuntos. Cuando el valor es bajo, significa que las diferencias entre las imágenes reales y las generadas son menores, lo que indica que el modelo es capaz de producir resultados más realistas [71].

En el caso de este Trabajo de Fin de Grado, la métrica basada en la distancia de Fréchet se empleará para medir la similitud entre dos conjuntos de trayectorias entre trayectorias reales de buques y las trayectorias generadas.

Para poder aplicar esta métrica, cada trayectoria se modela como una secuencia ordenada de puntos. A partir de estas secuencias, se evalúa la similitud entre las trayectorias reales y las trayectorias generadas sintéticamente comparando su comportamiento global mediante la distancia de Fréchet.

Un valor bajo de la métrica basada en la distancia de Fréchet entre las trayectorias reales y las trayectorias generadas indica que presentan un comportamiento similar desde el punto de vista estadístico y geométrico. Por el contrario, valores elevados de esta métrica reflejan una mayor diferencia entre ambos conjuntos de trayectorias, indicando que las trayectorias generadas pueden presentar desviaciones significativas respecto a las reales.

El uso de la distancia de Fréchet como métrica para evaluar la calidad de modelos generativos, como las redes GAN, ya ha sido aplicado y estudiado en contextos reales. Un ejemplo de ello se encuentra en el ámbito de la imagen médica. En la citada Tesis[70], se emplean GAN condicionales (cGAN) para el aumento de datos con el fin de mitigar problemas derivados de la escasez de muestras disponibles en tareas de clasificación médica. En este caso, la Distancia de Fréchet se emplea para comparar las imágenes reales con las imágenes generadas por el modelo, permitiendo evaluar de forma cuantitativa la calidad y el realismo de las imágenes sintetizadas. De este modo, la distancia de Fréchet se convierte en una herramienta útil para validar que las imágenes generadas son adecuadas para su posterior uso en análisis médicos, contribuyendo al desarrollo de modelos de diagnóstico más fiables y precisos.

4.1.2 Comparación espacial mediante mapas de calor

Para analizar los resultados, se utilizará una comparación visual mediante mapas de calor con el objetivo de analizar la distribución espacial de las trayectorias generadas por los distintos modelos. Los mapas de calor son una técnica de visualización que permite representar la concentración de datos en una determinada zona mediante variaciones de color, de forma que las áreas con mayor presencia de datos aparecen más destacadas visualmente. Este tipo de representación se emplea con frecuencia para analizar distribuciones espaciales, ya que permite identificar de manera sencilla dónde se concentra una mayor cantidad de información [73].

Para llevar a cabo este análisis se generará un mapa de calor independiente para cada uno de los métodos de generación de trayectorias utilizados, así como un mapa adicional construido a partir de las trayectorias reales obtenidas de los datos AIS. De esta manera, cada mapa representará la distribución espacial del conjunto de trayectorias correspondiente, permitiendo observar de forma clara por qué zonas del área de estudio se concentran más trayectorias y cuáles presentan menor presencia.

Una vez generados los mapas, se realizará una comparación visual entre los mapas obtenidos mediante los distintos métodos y el mapa correspondiente a las trayectorias reales. El objetivo de esta comparación será comprobar hasta qué punto los modelos son capaces de reproducir un comportamiento espacial similar al observado en los datos reales. En particular, se analizará si las trayectorias representadas en los mapas generados siguen líneas de navegación parecidas a las presentes en el mapa real y si la distribución general de las trayectorias dentro del área de estudio mantiene una forma similar. Esta comparación permitirá valorar de forma cualitativa el grado de similitud entre las trayectorias reales y las generadas por cada método.

4.1.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos

4.1.3.1 MAE (Mean Absolute Error)

El error absoluto medio (Mean Absolute Error, MAE) es una métrica utilizada para medir la diferencia entre los valores reales y los valores estimados por un modelo. Esta medida calcula el promedio de las diferencias absolutas entre ambos conjuntos de datos, proporcionando una forma

sencilla de cuantificar el error cometido por un método de predicción o estimación. Una de sus principales ventajas es que resulta fácil de interpretar, ya que el error se expresa en las mismas unidades que la variable analizada, lo que permite evaluar directamente la magnitud de la desviación entre los valores reales y los generados por el modelo [77][78].

Matemáticamente, el MAE se define como el promedio del valor absoluto de las diferencias entre los valores reales y los valores estimados:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Ecuación 6

Donde:

- y_i representa el valor real.
- $\hat{y}_i$ representa el valor estimado por el modelo.
- n representa el número total de observaciones consideradas.

El uso del valor absoluto evita que los errores positivos y negativos se cancelen entre sí, permitiendo obtener una medida directa de la magnitud media del error.

Un valor bajo de MAE indica que las predicciones del modelo se aproximan correctamente a los valores reales, mientras que valores más elevados reflejan una mayor diferencia entre ambos conjuntos de datos.

En el contexto de este trabajo, el MAE se empleará para comparar los valores de velocidad de las trayectorias reales con los generados por los distintos métodos. De esta forma, esta métrica permitirá cuantificar el error medio existente entre ambas series de velocidades, proporcionando una medida clara del grado de similitud entre los perfiles de velocidad reales y los generados por los modelos.

4.1.3.2 DTW (Dynamic Time Warping)

El Dynamic Time Warping o alineamiento dinámico del tiempo (DTW) es un método utilizado para medir la similitud entre dos series temporales que pueden presentar diferencias en su evolución a lo largo del tiempo. A diferencia de otras métricas que comparan punto a punto dos secuencias, DTW permite alinear las series de forma flexible, de manera que elementos que ocurren en momentos ligeramente distintos puedan compararse entre sí. Esto resulta especialmente útil cuando dos series presentan un comportamiento similar, pero no están perfectamente sincronizadas temporalmente. El objetivo del algoritmo es encontrar una alineación óptima entre ambas secuencias minimizando una distancia acumulada entre sus valores correspondientes [74][75]. Esto se aprecia en la Figura 4-3.

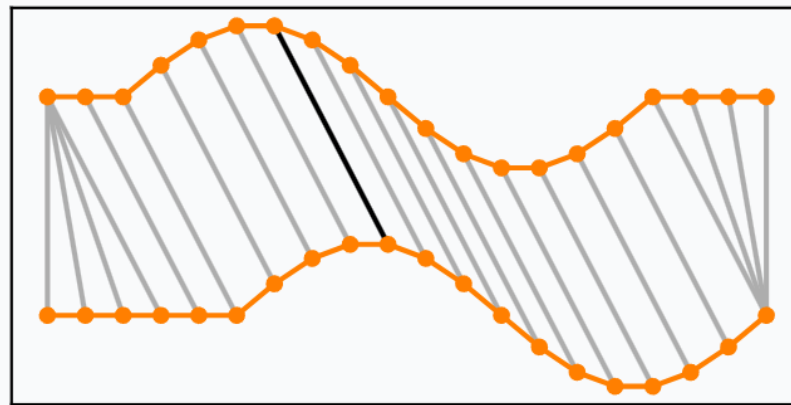


Figura 4-3 Ejemplo de aplicación de DTW entre dos trayectorias[76]

El funcionamiento de DTW se basa en la construcción de una matriz de distancias entre los puntos de las dos series temporales que se desean comparar. Cada elemento de esta matriz representa la distancia

entre un punto de la primera serie y un punto de la segunda. A partir de esta matriz, el algoritmo busca el camino de deformación óptimo que conecta el inicio y el final de ambas secuencias minimizando el coste total acumulado. Este camino permite que un punto de una serie se alinee con uno o varios puntos de la otra, lo que hace posible comparar correctamente series que tienen distinta duración o que evolucionan a velocidades diferentes [76]. Matemáticamente, el DTW puede expresarse como la distancia mínima acumulada a lo largo del camino de alineación óptimo entre ambas series.

$$DTW(X, Y) = P_{min} \sum_{(i,j) \in P} d(x_i, y_j) \quad \text{Ecuación 7}$$

Donde:

- P es el camino de alineación (warping path)
- $d(x_i, y_j)$ es la distancia entre puntos
- el algoritmo busca el camino con coste mínimo

En el contexto de este trabajo, DTW se empleará para comparar los perfiles de velocidad de las trayectorias reales y las trayectorias generadas por los distintos modelos. Valores más bajos indican una mayor similitud entre los perfiles de velocidad comparados, por el contrario, valores más altos señalan una mayor diferencia. Esta métrica permite analizar si la evolución de la velocidad a lo largo de las trayectorias sigue un comportamiento similar al observado en los datos AIS reales, incluso cuando existen pequeñas diferencias en el momento en el que se producen aceleraciones o cambios de velocidad. Por este motivo, DTW resulta especialmente adecuado para evaluar la similitud entre series temporales de velocidad, ya que permite comparar la forma general del perfil de velocidad sin exigir que ambas secuencias estén perfectamente alineadas en el tiempo.

4.1.3.3 Cálculos estadísticos de la velocidad

Con el objetivo de analizar el comportamiento general de la velocidad en las trayectorias generadas por los distintos métodos, se realizará un análisis estadístico basado en dos medidas: la velocidad media y la desviación estándar. Estas métricas se calcularán tanto para las trayectorias reales obtenidas a partir de los datos AIS como para las trayectorias generadas por cada uno de los modelos, permitiendo comparar su comportamiento global.

La **velocidad media** permite obtener un valor representativo de la velocidad del conjunto de trayectorias analizado. Esta se calcula como el promedio de las velocidades medias de todas las trayectorias consideradas:

$$\bar{v} = \frac{1}{N} \sum_{i=1}^N v_i \quad \text{Ecuación 8}$$

Donde:

- v_i representa la velocidad media de cada trayectoria.
- N representa el número total de trayectorias analizadas.

Por otro lado, la **desviación estándar** mide la variabilidad de estas velocidades respecto al valor medio, permitiendo evaluar si las trayectorias generadas presentan una dispersión de velocidades similar a la observada en las trayectorias reales. Su expresión matemática es:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2} \quad \text{Ecuación 9}$$

Donde:

- v_i representa la velocidad media de cada trayectoria.
- $\bar{v}$ representa la velocidad media.
- N representa el número total de trayectorias analizadas.

Estas métricas permiten comparar de forma sencilla si los distintos métodos generan niveles medios de velocidad y variabilidad similares a los observados en las trayectorias reales.

4.2 Validación en buques tipo Cargo

En este apartado se analizan los resultados obtenidos para los buques tipo *Cargo*, evaluando la capacidad de los distintos modelos desarrollados para reproducir los patrones de navegación observados en los datos AIS reales.

La validación se realiza mediante la comparación entre las trayectorias reales y las trayectorias generadas por cada uno de los modelos implementados, utilizando las métricas descritas en el apartado anterior. Asimismo, se lleva a cabo una comparación entre los distintos métodos de generación con el fin de analizar su rendimiento y determinar cuál de ellos reproduce con mayor fidelidad el comportamiento de navegación característico de los buques tipo *Cargo*.

Con este objetivo, se seleccionará un conjunto de 200 trayectorias reales procedentes de los datos AIS y se generarán 200 trayectorias sintéticas para cada uno de los métodos analizados. De esta forma, todas las comparaciones se realizarán utilizando conjuntos de trayectorias del mismo tamaño, lo que permite llevar a cabo un análisis equilibrado entre los distintos métodos utilizados. Así mismo, al igual que en los cargueros, se discretizarán las velocidades de las trayectorias reales con el objetivo de que las gráficas y, por lo tanto, el comportamiento del buque no difiera demasiado con las velocidades asignadas en los modelos y se pueda realizar una comparación realista y adecuada de estas (las velocidades sintéticas fueron discretizadas). Esta selección de trayectorias servirá como base para el cálculo de las distintas métricas de evaluación y para la posterior comparación de los resultados obtenidos.

4.2.1 Distancia de Fréchet

4.2.1.1 Unidad de cálculo

En este trabajo, la distancia de Fréchet se ha calculado directamente utilizando las coordenadas geográficas en latitud y longitud, sin convertir previamente las posiciones a metros o kilómetros. Esto se debe a que esta métrica se emplea principalmente para comparar la forma y similitud geométrica entre trayectorias, por lo que la unidad de medida no afecta al resultado, siempre que todas las trayectorias se encuentren en el mismo sistema de coordenadas.

Dado que el área de estudio se limita al entorno del litoral de Barcelona, los errores asociados a la curvatura terrestre son muy pequeños, por lo que trabajar en grados proporciona resultados comparables a los que se obtendrían utilizando unidades métricas. Además, evitar la conversión a metros simplifica el cálculo y evita introducir posibles errores derivados de transformaciones geodésicas.

Por tanto, utilizar directamente las coordenadas geográficas permite realizar una comparación geométrica consistente y eficiente entre trayectorias, manteniendo la validez del análisis de similitud espacial mediante la distancia de Fréchet.

4.2.1.2 Obtención de resultados

Durante el análisis de los resultados obtenidos mediante la distancia de Fréchet se detectó una limitación en la interpretación de las métricas calculadas. Aunque se disponía de los valores de distancia correspondientes a cada método, estos resultados no permitían determinar por sí solos si las trayectorias generadas presentaban una similitud adecuada con respecto a las trayectorias reales.

Con el objetivo de disponer de una referencia que permitiera interpretar correctamente estos valores, se decidió realizar una comparación adicional entre trayectorias reales. Para ello, se calcularon las distancias de Fréchet entre un conjunto de 200 trayectorias reales. Esta comparación permite establecer un punto de referencia sobre el nivel de similitud entre trayectorias reales, ya que en la práctica es muy poco probable encontrar dos trayectorias exactamente iguales. De esta forma, los valores obtenidos en la comparación real-real permiten poner en contexto los resultados obtenidos al comparar trayectorias reales con trayectorias sintéticas. Si las distancias de Fréchet obtenidas para los modelos generativos se aproximan a las observadas entre trayectorias reales, podrá considerarse que las trayectorias generadas reproducen de forma adecuada los patrones de navegación presentes en los datos AIS.

Por otro lado, a la hora de calcular esta distancia, únicamente se tendrán en cuenta las 20 trayectorias sintéticas más cercanas a la trayectoria real analizada, excluyendo, como es lógico, la propia trayectoria comparada. Este procedimiento permite mantener comparaciones espacialmente relevantes, reduciendo de forma significativa el número de cálculos necesarios y, por tanto, el coste computacional asociado al proceso.

Este conjunto de 20 trayectorias se selecciona de forma rápida mediante el cálculo del centroide de cada trayectoria. El centroide se define como la posición media de todos los puntos que componen una trayectoria, lo que permite representar de forma simplificada su localización espacial y facilitar la identificación de las trayectorias más cercanas. Este procedimiento permite realizar un filtrado previo de trayectorias candidatas antes de calcular la distancia de Fréchet.

Una vez calculadas las distancias de Fréchet para cada trayectoria, se obtuvieron diferentes medidas estadísticas con el objetivo de analizar el comportamiento global de cada método. En particular, se calcularon la media, la mediana y el percentil 95 de las distancias obtenidas.

- **Media:** permite conocer el nivel promedio de similitud entre las trayectorias reales y las generadas por cada modelo.
- **Mediana:** por su parte, proporciona una medida más robusta frente a valores extremos, permitiendo evaluar el comportamiento típico de las trayectorias sin que algunos casos atípicos influyan de forma excesiva en el resultado.
- **Percentil 95:** permite analizar los casos más desfavorables, mostrando el valor por debajo del cual se encuentran el 95% de las distancias calculadas.

En conjunto, estas métricas permiten evaluar tanto el comportamiento medio de los modelos como la presencia de posibles trayectorias con mayores diferencias respecto a los datos reales.

4.2.1.3 Resultados

Estos son los resultados obtenidos para cada uno de los modelos empleados, así como la comparación real-real:

Método	Media	Mediana	Percentil 95%
Real/Real	0,024	0,018	0,062
GAN/Real	0,027	0,022	0,061
GAN+KNN/Real	0,027	0,019	0,059
Mapa de calor+H3/ Real	0,035	0,026	0,099

Tabla 4-1 Resultados de Distancia de Fréchet

Método	Cálculo	Porcentaje diferencia
Real/Real	$\frac{0,024}{0,024} = 1$	0%
GAN/Real	$\frac{0,027}{0,024} = 1,12$	12%
GAN+KNN/Real	$\frac{0,027}{0,024} = 1,12$	12%
Mapa de calor+H3/ Real	$\frac{0,035}{0,024} = 1,46$	46%

Tabla 4-2 Cálculo porcentaje diferencia respecto trayectorias reales

➤ **Real-Real:**

Mide el grado de similitud existente entre trayectorias reales obtenidas a partir de los datos AIS. Los resultados muestran una media de 0,024, lo que indica que incluso entre trayectorias reales existen diferencias geométricas apreciables. Este valor refleja la variabilidad natural del tráfico marítimo, ya que las trayectorias de los buques no son exactamente idénticas debido a factores como pequeñas variaciones en la navegación, condiciones operativas o maniobras realizadas durante el tránsito. La mediana, situada en 0,018, confirma que la mayoría de las trayectorias reales presentan una similitud relativamente elevada entre sí. Por su parte, el percentil 95 alcanza un valor de 0,062, lo que indica que, incluso en los casos más desfavorables, las diferencias entre trayectorias reales se mantienen dentro de ese rango.

En consecuencia, este conjunto de valores puede considerarse como una referencia o nivel base de similitud para el problema analizado. Es decir, valores de distancia de Fréchet próximos a estos niveles pueden interpretarse como resultados coherentes con la variabilidad real presente en los datos AIS.

➤ **GAN-Real:**

Se obtiene una media de 0,027, lo que supone un incremento muy reducido respecto al valor observado en la comparación real-real (0,024). Esta pequeña diferencia indica que, en promedio, las trayectorias generadas por el modelo presentan un nivel de similitud muy cercano al existente entre trayectorias reales, solo un 12% por encima del ruido de las trayectorias reales. La mediana se sitúa en 0,022, valor también muy próximo al observado en la comparación real-real, lo que sugiere que la mayor parte de las trayectorias generadas reproducen de forma adecuada la geometría de las trayectorias reales. Finalmente, el percentil 95 alcanza un valor de 0,061, prácticamente idéntico al obtenido en la comparación entre trayectorias reales (0,062). Este resultado indica que incluso en los casos menos favorables las trayectorias generadas por el modelo mantienen un grado de similitud muy similar al observado entre trayectorias reales.

En conjunto, estos resultados muestran que el modelo GAN Total es capaz de generar trayectorias cuya geometría reproduce de forma muy fiel los patrones de navegación presentes en los datos AIS.

➤ **GAN+KNN-Real:**

La media obtenida es de 0,027, prácticamente idéntica a la del modelo GAN Total, lo que indica que el error medio respecto a las trayectorias reales se mantiene en niveles muy similares. Al igual que en el caso anterior, solo un 12% de diferencia respecto a las trayectorias reales. Sin embargo, la mediana alcanza un valor de 0,019, muy cercano al observado en la comparación real-real (0,018). Este resultado sugiere que, en un número elevado de casos, las trayectorias generadas por este modelo presentan una similitud geométrica prácticamente equivalente a la existente entre trayectorias reales. Además, el

percentil 95 alcanza un valor de 0,059, ligeramente inferior al observado en la comparación real-real. Este resultado indica que incluso en los casos más desfavorables el modelo mantiene un nivel de similitud muy elevado.

En conjunto, estos resultados sugieren que el modelo GAN + KNN es capaz de generar trayectorias altamente realistas en un gran número de casos, aunque la media aumente ligeramente debido a la presencia de algunas trayectorias con mayor desviación. Desde el punto de vista geométrico, tanto GAN como GAN + KNN muestran un rendimiento muy similar en la generación de trayectorias sintéticas, teniendo GAN+KNN mejores resultados y, por lo tanto, siendo más parecido.

➤ **Mapa de calor+H3-Real:**

La media obtenida es de 0,035, lo que supone un incremento aproximado del 45 % respecto al valor observado en la comparación real-real. Este resultado indica que, en promedio, las trayectorias generadas por este modelo se alejan más de las trayectorias reales que en el caso de los modelos basados en redes generativas. La mediana alcanza un valor de 0,026, claramente superior al observado en la comparación real-real (0,018), lo que sugiere que esta mayor diferencia no se limita únicamente a algunos casos aislados, sino que afecta a una parte significativa de las trayectorias generadas. Finalmente, el percentil 95 alcanza un valor de 0,099, considerablemente superior al observado en los otros modelos. Este resultado indica que en los casos más desfavorables las trayectorias generadas por el modelo H3 presentan desviaciones geométricas notablemente mayores respecto a las trayectorias reales. En conjunto, estos resultados sugieren que, aunque el modelo H3 es capaz de reproducir de forma general la estructura del tráfico marítimo, las trayectorias generadas presentan una mayor dispersión y se alejan con mayor frecuencia de los patrones reales de navegación.

En conjunto, el análisis de la distancia de Fréchet muestra que los modelos basados en redes generativas (GAN Total y GAN + KNN) son capaces de generar trayectorias cuya similitud geométrica con las trayectorias reales se aproxima al nivel de variabilidad presente en los propios datos AIS, siendo el modelo GAN+KNN un poco más superior. Por el contrario, el modelo basado en discretización espacial H3 presenta mayor desviación, lo que sugiere una menor capacidad para reproducir con precisión la geometría de las trayectorias observadas en el tráfico marítimo real.

Con el objetivo de obtener una comparación directa final entre los distintos métodos, se usó el valor medio de las distancias de Fréchet obtenidas para cada modelo para ordenarlos en función de su desempeño, considerando que menores distancias de Fréchet indican una mayor similitud geométrica entre trayectorias y, por tanto, una mayor capacidad del modelo para reproducir el comportamiento observado en los datos AIS.

Método	Media	Puesto
Real/Real	0,0239	-
GAN+KNN/Real	0,0266	1º
GAN/Real	0,0273	2º
Mapa de calor+H3/ Real	0,0352	3º

Tabla 4-3 Clasificación de modelos

En la clasificación final se observa que el método GAN + KNN, considerando el valor medio de las distancias de Fréchet obtenidas, es el que ofrece el mejor rendimiento entre los modelos analizados, al presentar el menor valor medio de distancia y, por tanto, la mayor similitud geométrica con las trayectorias reales. Esto coincide con lo analizado anteriormente.

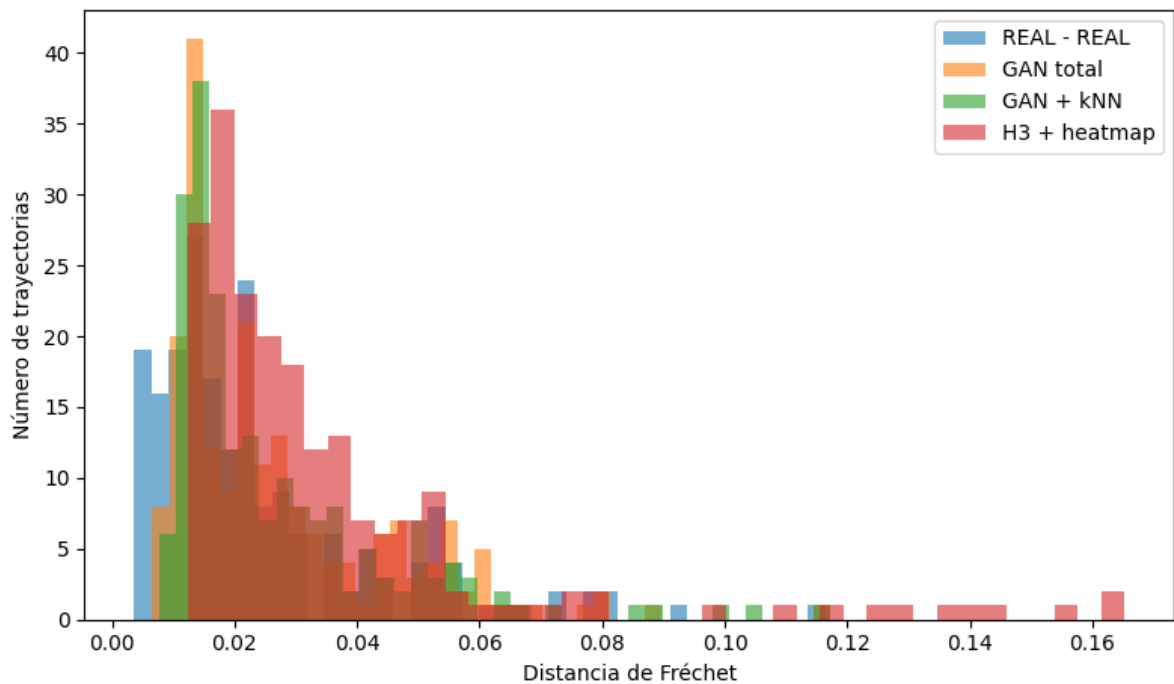


Figura 4-4 Gráfica resultados de distancia de Fréchet (Fuente: propia)

En la Figura 4-4 se puede observar una gráfica con los resultados obtenidos. En ella, los resultados que se encuentren más a la izquierda supondrán mejores resultados y, por lo tanto, trayectorias más parecidas entre sí. Por el contrario, colas más largas y barras más a la derecha representan resultados más dispares de lo normal. Se puede comprobar que incluso el conjunto de datos reales tiene alguna trayectoria hacia la derecha, lo que representa la variabilidad inherente al tráfico marítimo.

En esta gráfica se hace perceptible cómo las trayectorias que emplean GAN tienen resultados semejantes y más parecidos con la realidad, así como más agrupados en la zona izquierda y con menor cola. Aun así, como se puede observar, la concentración en la zona izquierda del modelo GAN + KNN es ligeramente mejor que el modelo GAN, al tener menor cola. Por otro lado, las trayectorias generadas con mapas de calor y celdas H3, aun teniendo buenos resultados, tiene una mayor cola y menor concentración de trayectorias a la izquierda de la gráfica.

En conclusión, se observa que las distribuciones correspondientes a los modelos basados en cGAN, sobre todo el modelo GAN + KNN, se solapan ampliamente con la distribución de similitud entre trayectorias reales, lo que indica que las trayectorias generadas presentan una geometría comparable a la variabilidad natural del sistema. En contraste, el modelo basado en H3 muestra una distribución con mayores valores de distancia, evidenciando una menor similitud geométrica.

4.2.1.4 Método seleccionado según la distancia Fréchet

A partir de los resultados obtenidos mediante la distancia de Fréchet, se observa que los modelos basados en redes generativas presentan una capacidad notable para reproducir la geometría de las trayectorias reales. En particular, el método GAN + KNN muestra el mejor comportamiento global, al obtener los valores más bajos de distancia media y mantener una elevada similitud con las trayectorias reales. Por ello, en términos de similitud geométrica entre trayectorias, GAN + KNN se identifica como el método que ofrece los mejores resultados en este análisis.

4.2.2 Comparación espacial mediante mapas de calor

Con el objetivo de analizar la distribución espacial de las trayectorias generadas, se construyeron mapas de calor a partir de todos los puntos que componen las trayectorias reales y las generadas por cada modelo. Para ello se empleó una representación hexagonal de densidad, en la que cada celda muestra la

concentración de puntos de trayectoria en esa zona. Esta visualización permite comparar de forma directa la distribución espacial del tráfico generado por cada modelo con la observada en los datos AIS reales. El resultado se muestra en la Figura 4-5:

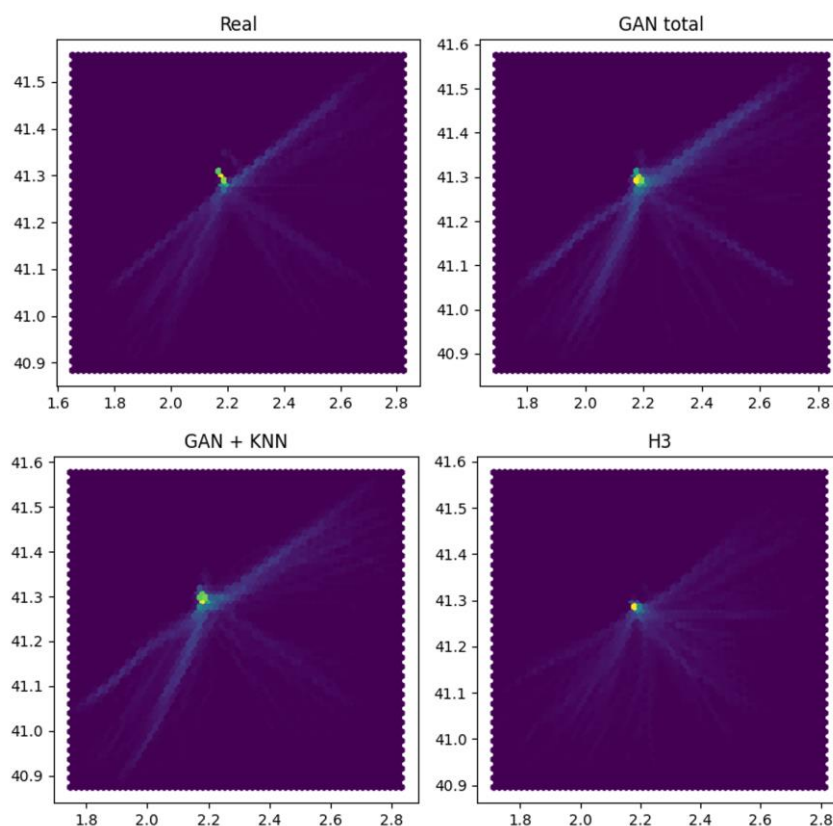


Figura 4-5 Comparación visual de la distribución espacial de trayectorias mediante mapas de calor (Fuente: propia)

En el mapa correspondiente a las trayectorias reales se observa una estructura espacial claramente definida. Destaca una zona de elevada densidad situada en el entorno del acceso portuario, a partir de la cual se organizan varias líneas de navegación que se extienden hacia mar abierto. Entre ellas sobresale una ruta dominante en dirección noreste (NE), acompañada por varios ramales secundarios que reflejan la dispersión natural del tráfico en el área de estudio. En conjunto, el mapa real muestra un patrón coherente que se caracteriza por una fuerte concentración de trayectorias en las zonas más transitadas y una progresiva dispersión a medida que aumenta la distancia al puerto.

- **Mapa generado mediante GAN:** reproduce de forma notable esta organización espacial general. La posición del núcleo de mayor densidad coincide de manera muy aproximada con la observada en los datos reales y la ruta principal hacia el noreste mantiene una orientación y continuidad similares. Asimismo, pueden identificarse varios de los ramales secundarios presentes en el mapa real, aunque en este caso aparecen ligeramente más suavizados y con menor definición. A pesar de ello, el modelo consigue capturar correctamente la estructura global del tráfico marítimo en el área analizada.
- **Mapa generado mediante GAN + KNN:** la similitud visual con el patrón real resulta todavía más evidente. Además de conservar la localización del núcleo central de alta densidad, este enfoque reproduce con mayor claridad la geometría de la línea principal de navegación y mantiene una definición más marcada en algunos de los ramales secundarios. En términos generales, la distribución espacial obtenida mediante este método presenta una correspondencia visual muy próxima a la observada en los datos AIS, manteniendo tanto la

concentración de trayectorias en las zonas principales como una dispersión coherente hacia las áreas periféricas.

- **Mapa generado mediante el modelo H3:** muestra una mayor diferencia respecto al patrón real. Aunque se mantiene la presencia de un núcleo central y varias direcciones de salida reconocibles, la distribución espacial aparece más difusa y menos ajustada a la estructura observada en los datos AIS. Algunas de las líneas de navegación presentan una menor definición, mientras que otras muestran una dispersión mayor de la esperada, lo que da lugar a una representación más abierta del tráfico marítimo.

En conjunto, la comparación visual de los mapas de calor indica que los modelos basados en redes generativas son los que mejor reproducen la distribución espacial del tráfico marítimo real. En particular, el método GAN + KNN es el que presenta la correspondencia más próxima con el patrón observado en los datos AIS. Este modelo mantiene con mayor claridad tanto la localización del núcleo principal de densidad como la geometría de la ruta dominante y la disposición de los ramales secundarios. El modelo GAN también reproduce de forma adecuada la estructura general del tráfico, aunque con una ligera pérdida de definición en algunas de las líneas de navegación. Por el contrario, el enfoque basado en discretización H3 muestra una mayor dispersión espacial y una menor precisión en la representación de las rutas principales. Por tanto, desde el punto de vista visual, GAN + KNN se presenta como el método que mejor aproxima la distribución espacial del tráfico real en el área de estudio, resultado que además se encuentra en consonancia con el análisis cuantitativo realizado mediante la distancia de Fréchet.

4.2.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos

4.2.3.1 Análisis gráfico de la velocidad

A continuación, se analiza el perfil medio de velocidad a lo largo de las trayectorias con el objetivo de evaluar hasta qué punto los distintos modelos son capaces de reproducir el comportamiento cinemático observado en los datos AIS reales. Para ello, se calcula la velocidad media en cada punto de la trayectoria considerando el conjunto de trayectorias disponibles para cada método. Con el fin de poder comparar trayectorias de distinta longitud, el eje horizontal se normaliza entre 0 y 1, representando el progreso relativo a lo largo de la trayectoria desde el punto inicial hasta el punto final. Además, se representa la desviación estándar asociada al perfil medio de velocidad de los datos reales (área azul cielo), lo que permite visualizar la variabilidad del comportamiento observado. Por otro lado, el eje vertical representa la velocidad en nudos.

El resultado de esta gráfica se muestra en la Figura 4-6.

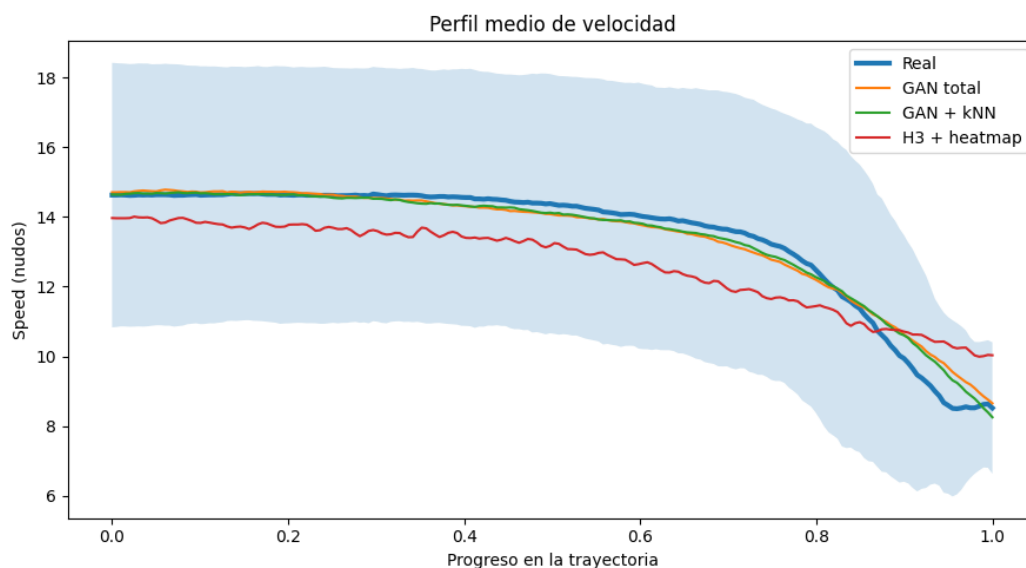


Figura 4-6 Perfil de velocidades según método (Fuente: propia)

La curva azul representa el perfil medio de velocidad de las trayectorias reales en función del progreso relativo a lo largo del recorrido. En ella pueden distinguirse tres fases principales del comportamiento de navegación.

En el **inicio de la trayectoria** (aproximadamente entre 0 y 0,3 del progreso normalizado), la velocidad se mantiene prácticamente constante, con valores en torno a 14-15 nudos. Este comportamiento es consistente con una fase de navegación relativamente estable en mar abierto.

En el **tramo intermedio** de la trayectoria (entre 0,3 y aproximadamente 0,75), se observa un descenso gradual de la velocidad media. Esta reducción progresiva puede asociarse al proceso de aproximación al entorno portuario, donde los buques comienzan a ajustar su velocidad en preparación para las maniobras posteriores.

Finalmente, en el **tramo final** de la trayectoria (entre 0,75 y 1,0), la velocidad experimenta una disminución más pronunciada, alcanzando valores próximos a 8-9 nudos. Esta fase se corresponde con la etapa final de aproximación y con las maniobras asociadas a la entrada al puerto.

La zona sombreada alrededor de la curva representa la desviación estándar de las velocidades observadas en los datos reales. Esta banda permite visualizar la variabilidad existente entre trayectorias, mostrando que existen trayectorias con velocidades superiores o inferiores al valor medio, lo que refleja la diversidad de comportamientos de navegación en esta fase de la aproximación portuaria.

A partir del perfil medio de velocidad representado en la figura, es posible analizar el comportamiento de los distintos modelos en relación con la evolución de la velocidad observada en las trayectorias reales.

➤ Modelos basados en cGAN:

Las curvas correspondientes a GAN Total (naranja) y GAN + KNN (verde) presentan un comportamiento muy similar entre sí, lo que justifica su análisis conjunto. Ambos modelos reproducen adecuadamente el nivel de velocidad inicial y mantienen una forma general del perfil, muy próxima a la observada en los datos reales.

No obstante, se aprecia una diferencia en la dinámica de reducción de velocidad a lo largo de la trayectoria. Mientras que en los datos reales la disminución de la velocidad se produce de forma más brusca en el tramo final, las curvas generadas por los modelos GAN muestran una reducción más progresiva de la velocidad. Como consecuencia, la transición hacia velocidades menores aparece más suavizada en las trayectorias generadas. Además, en el perfil real se observa que, tras la desaceleración final, la velocidad tiende a mantenerse constante durante un pequeño tramo al final de la trayectoria, comportamiento que no se reproduce con la misma claridad en los modelos generativos.

La principal diferencia entre GAN y GAN + KNN se observa en la velocidad final. En el caso de GAN, la curva se sitúa ligeramente por encima del perfil real, mientras que el modelo GAN + KNN se mantiene por debajo. Más allá de esta diferencia al final de la trayectoria, ambos perfiles presentan una evolución prácticamente idéntica.

En conjunto, puede afirmarse que los modelos basados en GAN reproducen correctamente la tendencia general del perfil de velocidad, aunque tienden a generar una desaceleración más gradual que la observada en los datos reales. Considerado esto, el modelo GAN es mejor que el GAN+KNN, ya que la curva acaba más cerca que la de trayectorias reales.

➤ Modelo basado en H3:

El perfil de velocidad generado mediante el modelo H3 y mapas de calor, presenta diferencias más evidentes respecto al comportamiento observado en las trayectorias reales. Durante la mayor parte del recorrido, las velocidades obtenidas con este método se sitúan sistemáticamente por debajo de las registradas en los datos AIS. Sin embargo, en el tramo final de la trayectoria se observa el

comportamiento contrario, con valores que llegan a situarse ligeramente por encima de los correspondientes al perfil real.

Además, la evolución del perfil generado por este modelo se caracteriza por ser más constante a lo largo de la trayectoria, mostrando una variabilidad menor que la observada tanto en los datos reales como en los modelos basados en GAN. Sobre esta tendencia más uniforme aparecen pequeños picos de velocidad intermedios, que generan un perfil menos suave y coherente con la evolución progresiva que presentan las trayectorias reales. Por otra parte, el modelo muestra una menor capacidad para reproducir la dinámica de desaceleración observada en los datos reales, donde la reducción de velocidad hacia el final de la trayectoria se produce de forma más marcada. En el caso del modelo H3, esta transición aparece más suavizada y menos ajustada al patrón real, lo que demuestra una representación más imprecisa del comportamiento cinemático de los buques a lo largo del recorrido.

En conjunto, estas diferencias indican que el modelo basado en H3 y mapas de calor presenta una menor capacidad para reproducir el comportamiento cinemático de los buques, dando lugar a una representación menos realista de la evolución de la velocidad a lo largo de la trayectoria.

Con el objetivo de analizar el comportamiento de los modelos en la fase final de las trayectorias, se estudia la distribución de la velocidad en el último punto de cada trayectoria. Para ello se representa un histograma comparando las velocidades finales observadas en los datos reales con las generadas por los distintos modelos, tal y como se muestra en la Figura 4-7.

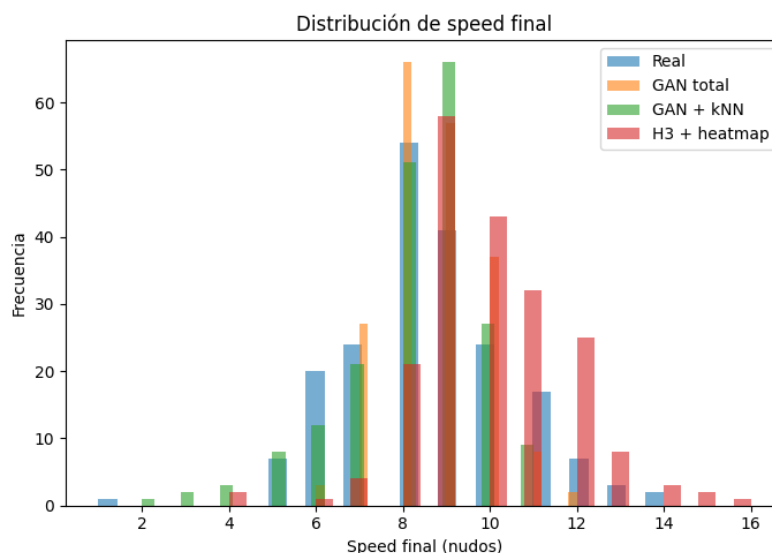


Figura 4-7 Gráfica de las velocidades finales según método (Fuente: propia)

La distribución de velocidades observada en los datos reales muestra un comportamiento característico de la fase final de aproximación al puerto. La mayor concentración de valores se sitúa aproximadamente entre 7 y 10 nudos, con un pico principal en torno a 8-9 nudos. Este patrón refleja la variabilidad natural que se produce durante la aproximación portuaria, donde los buques reducen progresivamente su velocidad, pero mantienen ciertas diferencias dependiendo de factores como la densidad de tráfico marítimo, las condiciones meteorológicas o las maniobras necesarias para la entrada al puerto.

Las distribuciones generadas por los modelos GAN y GAN + KNN presentan una forma muy similar a la observada en los datos reales. En ambos casos, el pico principal de la distribución se mantiene en el intervalo de 8-9 nudos, y la dispersión de los valores se sitúa en un rango comparable al de las trayectorias reales. Además, la forma general del gráfico reproduce adecuadamente la distribución de velocidades observada en el conjunto de datos AIS original, lo que indica que estos modelos consiguen capturar de forma realista el comportamiento dinámico de los buques durante la aproximación final al puerto.

Por el contrario, la distribución generada mediante el modelo H3 y mapa de calor presenta diferencias más notables respecto a los datos reales. En este caso se observa una mayor concentración de velocidades en el intervalo aproximado de 9 a 12 nudos, junto con una cola derecha más pronunciada, lo que indica la presencia de velocidades más elevadas que las observadas con mayor frecuencia en las trayectorias reales. Este comportamiento sugiere que el modelo H3 tiende a generar trayectorias en las que los buques mantienen velocidades finales ligeramente superiores, lo que apunta a una menor capacidad de este enfoque para reproducir correctamente el proceso de desaceleración característico de la aproximación al puerto.

4.2.3.2 Error absoluto medio (MAE)

Los valores obtenidos para el error absoluto medio (MAE) permiten cuantificar la diferencia entre el perfil medio de velocidad real y el generado por cada modelo a lo largo de la trayectoria. Los resultados obtenidos para cada uno de los modelos empleados se pueden ver en la tabla que se muestra a continuación:

Método	MAE (nudos)
GAN	0,2684499999999999
GAN+KNN	0,2211249999999999
Mapa de calor+H3	1,098425

Tabla 4-4 Resultados MAE

Se puede observar que ambos modelos GAN obtienen resultados por debajo de los 0,3 nudos, siendo el modelo GAN+KNN ligeramente mejor. Este resultado muestra que el modelo reproduce con precisión la evolución de la velocidad observada en los datos reales, con pequeñas variaciones.

Por el contrario, el modelo que emplea celdas H3, obtiene resultados por encima de 1 nudo, significativamente peor que los casos anteriores. Este valor indica que, en promedio, la diferencia entre el perfil generado por este modelo y el perfil real supera 1 nudo, lo que refleja una menor capacidad para reproducir con precisión la dinámica de velocidad observada en los datos AIS.

En definitiva, estos resultados muestran que los modelos basados en GAN presentan un rendimiento claramente superior en la reproducción del perfil de velocidad real. Entre ellos, GAN + KNN obtiene el mejor ajuste, mientras que el modelo H3 muestra una desviación considerablemente mayor, lo que confirma las diferencias observadas previamente en el análisis visual de los perfiles de velocidad.

4.2.3.3 Alineamiento dinámico del tiempo (DTW)

Los resultados obtenidos mediante la distancia DTW permiten evaluar la similitud entre la evolución temporal del perfil de velocidad real y la generada por cada modelo. Valores más bajos indican una mayor similitud entre los perfiles de velocidad comparados, por el contrario, valores más altos señalan una mayor diferencia. En la siguiente tabla se pueden ver los resultados obtenidos.

Método	DTW
GAN	0,735102033734094
GAN+KNN	0,47986977400124
Mapa de calor+H3	8,102046037884513

Tabla 4-5 Resultados DTW

Los resultados obtenidos mediante la distancia DTW permiten evaluar la similitud entre la evolución temporal del perfil de velocidad real y el generado por cada modelo. El modelo GAN + KNN presenta la menor distancia, lo que indica que es el que reproduce con mayor precisión la evolución de la velocidad a lo largo de la trayectoria. El modelo GAN total también muestra un comportamiento muy próximo al observado en los datos AIS, con valores igualmente bajos que reflejan una forma de curva similar a la real.

Por el contrario, el modelo H3 presenta una distancia significativamente mayor, lo que indica una diferencia mayor respecto al perfil de velocidad real. Esto nos muestra una menor capacidad para reproducir la dinámica temporal de la velocidad, siguiendo lo observado en el análisis visual de los perfiles.

Estas diferencias pueden deberse a que mientras que los modelos basados en GAN aprenden directamente los patrones temporales de las trayectorias AIS, el modelo H3 se basa principalmente en transiciones entre celdas, lo que puede limitar su capacidad para representar con precisión la evolución temporal de la velocidad.

4.2.3.4 Cálculos estadísticos

Método	Media(nudos)	Desviación estándar(nudos)
Real	13,42415	4,117450276263212
GAN	13,39665	2,3023289898491917
GAN+KNN	13,380775	2,376054586783519
Mapa de calor+H3	12,646925	2,1044032513696136

Tabla 4-6 Media y desviación estándar según método

La velocidad media de las trayectorias reales es reproducida con gran precisión por los modelos GAN total y GAN + KNN, cuyos valores se mantienen muy próximos a los observados en los datos AIS. Entre ambos, el modelo GAN + KNN presenta un ajuste ligeramente mejor al comportamiento real, mostrando una mayor proximidad a la velocidad media observada en las trayectorias reales. Por el contrario, el modelo H3 presenta una media inferior, lo que indica una mayor desviación respecto al comportamiento real de los buques.

En cuanto a la variabilidad de las velocidades, la desviación estándar de los datos reales es notablemente mayor que la obtenida en todos los modelos generativos. Esto indica que las trayectorias sintéticas presentan una dispersión menor y perfiles de velocidad más suavizados que los observados en los datos reales. Esto puede ser debido a que aprenden los patrones más frecuentes del conjunto de datos. Como consecuencia, las situaciones más extremas o menos frecuentes aparecen con menor frecuencia en las trayectorias generadas.

En conjunto, los resultados muestran que los modelos basados en GAN reproducen correctamente el nivel medio de velocidad, siendo GAN + KNN el que presenta el mejor resultado, mientras que el modelo H3 se aleja más del comportamiento real. Así mismo, se aprecia que todos los modelos tienden a reducir la variabilidad de velocidad observada en los datos AIS.

4.2.3.5 Método seleccionado según velocidad

A partir del análisis realizado sobre los perfiles de velocidad, las distribuciones de velocidad final y las métricas estadísticas obtenidas, se observa que los modelos basados en cGAN son los que reproducen mejor el comportamiento dinámico de las trayectorias reales. En particular, el modelo GAN + KNN

presenta el mejor resultado, ya que reproduce de forma más precisa la evolución temporal de la velocidad y mantiene una distribución de velocidades finales muy similar a la observada en los datos AIS reales.

En consecuencia, considerando el conjunto de resultados obtenidos en el análisis de velocidades, se selecciona GAN + KNN como el método que mejor representa la dinámica real de navegación, al ser el que reproduce con mayor fidelidad la evolución temporal de la velocidad.

4.2.4 Selección del método óptimo para buques tipo Cargo

Antes de seleccionar el modelo más adecuado para la generación de trayectorias de buques tipo Cargo, es necesario analizar de forma conjunta los resultados obtenidos mediante las diferentes métricas empleadas.

Para facilitar esta comparación, a continuación, se presenta una tabla resumen en la que se recoge la clasificación obtenida por cada modelo en los distintos métodos de evaluación, permitiendo identificar de forma clara cuál reproduce con mayor fidelidad el comportamiento observado en los datos AIS reales.

Método	GAN	GAN+KNN	Mapa de calor+H3
Distancia Fréchet			
Mapas de calor			
Análisis gráfico de la velocidad			
Velocidad final			
MAE			
DTW			
Cálculos estadísticos			

Tabla 4-7 Evaluación comparativa de los modelos analizados

MEJOR	
INTERMEDIO	
PEOR	

Tabla 4-8 Leyenda

A partir de los resultados recogidos en la Tabla 4-7, se observa que el modelo GAN + KNN obtiene el mejor comportamiento global en la mayoría de los métodos de evaluación considerados. Este modelo muestra la mayor similitud con las trayectorias reales tanto en las métricas cuantitativas como en los análisis visuales realizados, reproduciendo con mayor precisión la distribución espacial de las trayectorias, la evolución temporal de la velocidad y las características estadísticas del movimiento.

Aunque el modelo GAN presenta también un buen rendimiento y resultados cercanos a los datos reales, su comportamiento es ligeramente inferior al obtenido mediante GAN + KNN. Por el contrario, el método basado en mapas de calor con celdas H3 presenta desviaciones más significativas respecto al comportamiento observado en los datos AIS.

En consecuencia, considerando de forma conjunta todos los métodos de análisis empleados, se concluye que GAN + KNN es el modelo que mejor reproduce el comportamiento real de las trayectorias

de buques tipo Cargo, siendo por tanto el método seleccionado para representar de forma más realista la dinámica de navegación observada en los datos AIS.

En la Figura 4-8 se muestra un mapa con trayectorias generadas de las cuatro maneras distintas.

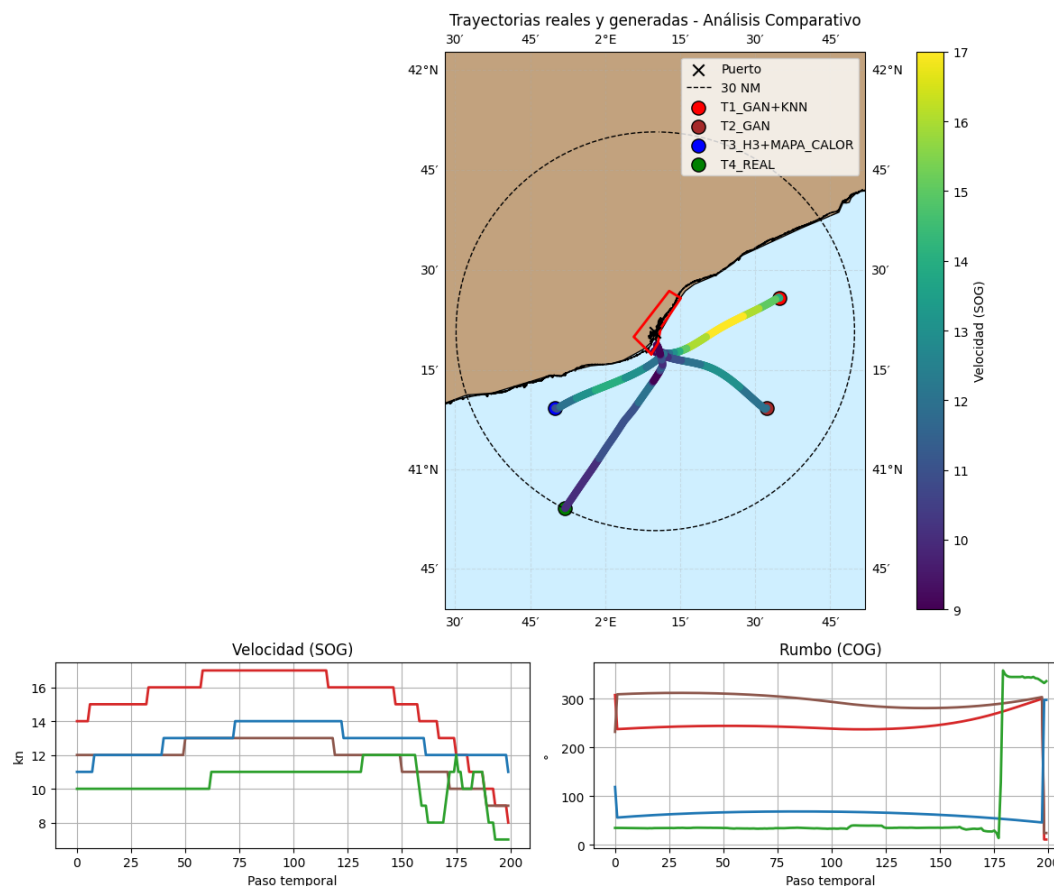


Figura 4-8 Trayectorias generadas con los distintos métodos Cargo (Fuente: propia)

4.3 Validación en buques tipo Fishing

En este apartado se analizan los resultados obtenidos para los buques tipo *Fishing*, evaluando la capacidad de los modelos desarrollados para reproducir el comportamiento de navegación observado en los datos AIS reales. A diferencia de los buques tipo *Cargo*, las embarcaciones pesqueras presentan trayectorias más irregulares y variables debido a la actividad de pesca, lo que supone una mayor complejidad en el proceso de modelado.

La validación se realiza comparando las trayectorias reales con las trayectorias sintéticas generadas por los distintos modelos implementados, utilizando las métricas descritas en el apartado “Métricas para la comparación de trayectorias”. Asimismo, se comparan los resultados obtenidos por cada modelo con el fin de identificar qué metodología reproduce con mayor fidelidad las dinámicas de navegación características de los buques pesqueros.

Con este objetivo, al igual que en los buques tipo *Cargo*, se seleccionará un conjunto de 200 trayectorias reales procedentes de los datos AIS y se generarán 200 trayectorias sintéticas para cada uno de los métodos analizados. De esta forma, todas las comparaciones se realizarán utilizando conjuntos de trayectorias del mismo tamaño, lo que permite llevar a cabo un análisis coherente entre los distintos métodos. Así mismo, al igual que en los cargueros, se discretizarán las velocidades de las trayectorias reales con el objetivo de que las gráficas y, por lo tanto, el comportamiento del buque no difiera demasiado con las velocidades asignadas en los modelos y se pueda realizar una comparación realista y adecuada de estas (las velocidades sintéticas cambian de 0,5 en 0,5 nudos, de esta manera las velocidades

reales se comportarán de la misma forma). Esta selección de trayectorias sirve como base para el cálculo de las distintas métricas de evaluación y para la posterior comparación de los resultados.

Dado que las métricas y los procedimientos de análisis empleados son los mismos que los descritos previamente para los buques tipo *Cargo*, en este apartado se presentarán los resultados de forma más concisa, centrándose en su interpretación.

4.3.1 Distancia de Fréchet

4.3.1.1 Resultados

Método	Media	Mediana	Percentil 95%
Real/Real	0,061	0,056	0,122
Mapa de calor+H3	0,087	0,077	0,147
LSTM	0,089	0,074	0,161

Tabla 4-9 Resultados distancia Fréchet

Método	Cálculo	Porcentaje diferencia
Real/Real	$\frac{0,061}{0,061} = 1$	0%
Mapa de calor+H3	$\frac{0,087}{0,061} = 1,42$	42%
LSTM	$\frac{0,089}{0,061} = 1,46$	46%

Tabla 4-10 Cálculo porcentaje diferencia respecto trayectorias reales

La variabilidad natural de las trayectorias de pesca puede evaluarse a partir de la comparación real-real, que proporciona una referencia del grado de similitud existente entre las propias trayectorias reales. En este caso, los valores obtenidos muestran que incluso dos trayectorias reales de pesca presentan una distancia de Fréchet media relativamente elevada. Si se compara con el caso de los buques tipo *Cargo*, donde la distancia media real-real era notablemente menor, se observa que las trayectorias de pesca presentan una dispersión aproximadamente 2,5 veces mayor. Este resultado indica que el comportamiento de los buques pesqueros es inherentemente más variable que el tráfico de entrada en puerto característico de los buques de carga.

Al comparar los modelos generativos con esta referencia, se observa que tanto H3 como LSTM presentan distancias medias superiores a las del conjunto de datos reales, en torno a un 42-46%.

➤ Mapa de calor+H3

El modelo basado en mapas de calor y H3 presenta valores de distancia de Fréchet superiores a los observados en la comparación real-real. La media muestra que las trayectorias generadas por el modelo presentan una desviación mayor respecto a las trayectorias reales que la existente entre trayectorias reales entre sí, en este caso un 42% mayor. Sin embargo, al analizar la mediana, se observa que el comportamiento típico del modelo se mantiene relativamente cercano al patrón real, lo que indica que una parte significativa de las trayectorias generadas reproduce razonablemente la geometría de las trayectorias pesqueras observadas en los datos AIS. Por su parte, el percentil 95 es más elevado que en el caso real-real, lo que sugiere la existencia de algunos casos extremos donde la desviación es mayor.

Este comportamiento puede deberse a que el método H3 modela las trayectorias como transiciones entre celdas espaciales discretas, lo que permite capturar correctamente las zonas de actividad y las tendencias espaciales generales del movimiento. No obstante, las trayectorias de pesca suelen incluir movimientos locales complejos, como cambios frecuentes de rumbo, trayectorias en zig-zag o pequeños bucles asociados a la actividad pesquera. Este tipo de patrones puede resultar más difícil de representar mediante una discretización espacial, haciendo que, aunque el modelo reproduzca adecuadamente la estructura global del movimiento, en determinados casos aparezcan desviaciones geométricas más grandes.

➤ **Modelo LSTM**

El modelo LSTM presenta resultados similares a los obtenidos por el modelo H3. La media de las distancias indica que el error medio respecto a las trayectorias reales es del mismo orden que el observado en el modelo anterior, lo que sugiere un rendimiento similar, aunque con un porcentaje de diferencia respecto a las reales de un 46%. La mediana, ligeramente inferior a la obtenida con H3, indica que el comportamiento típico de muchas trayectorias generadas por el modelo LSTM se aproxima razonablemente al patrón real observado en los datos AIS. Esto sugiere que el modelo es capaz de capturar parte de la estructura secuencial del movimiento pesquero.

Sin embargo, el percentil 95 muestra valores algo superiores, lo que indica la presencia de algunos casos en los que la desviación espacial es mayor. Esto sugiere que, aunque el modelo puede generar trayectorias relativamente similares a las reales en muchos casos, al partir también de celdas H3, aparecen algunas trayectorias con diferencias mayores.

➤ **Interpretación global**

En conjunto, los resultados obtenidos muestran que ambos modelos son capaces de reproducir parcialmente la geometría de las trayectorias reales de buques pesqueros, aunque con una desviación mayor que la observada en el caso de los buques tipo *Cargo*. Esta diferencia es coherente con la propia naturaleza de la actividad pesquera, caracterizada por trayectorias menos estructuradas, mayores cambios de dirección y una variabilidad espacial significativamente más elevada.

No obstante, al comparar de forma conjunta la media, la mediana y el percentil 95, se observa que el modelo H3 ofrece un comportamiento global ligeramente más favorable. Aunque el modelo LSTM presenta una mediana algo inferior, lo que indica un comportamiento típico muy similar al del modelo H3, este último obtiene una media menor y un percentil 95 más reducido, siendo ligeramente más estable y con una menor desviación en los casos más extremos.

En consecuencia, para el caso de los buques tipo *Fishing*, se considera que el modelo basado en discretización espacial H3 es el que ofrece el mejor comportamiento global en términos de similitud geométrica respecto a las trayectorias reales, si bien la diferencia respecto al modelo LSTM no es tan marcada. Esto sugiere que, en trayectorias pesqueras, ambos métodos resultan similares, aunque el método H3 presenta una ligera ventaja como mejor modelo dentro del conjunto de resultados analizados.

Al igual que en el caso de los cargueros, los modelos se ordenaron según el valor medio de las distancias Fréchet, considerando que las más bajas reflejan una mayor proximidad geométrica a las trayectorias observadas en los datos AIS y, por tanto, una mejor capacidad del modelo para representar el comportamiento real del tráfico analizado.

Método	Media	Puesto
Real/Real	0,0610	-
Mapa de calor+H3	0,0871	1º
LSTM	0,0888	2º

Tabla 4-11 Clasificación de los modelos

En la clasificación final se observa que el método Mapa de calor+H3 es el que ofrece el mejor rendimiento entre los modelos analizados, al presentar el menor valor medio de distancia y, por tanto, la mayor similitud geométrica con las trayectorias reales. Esto, coincide con lo analizado anteriormente.

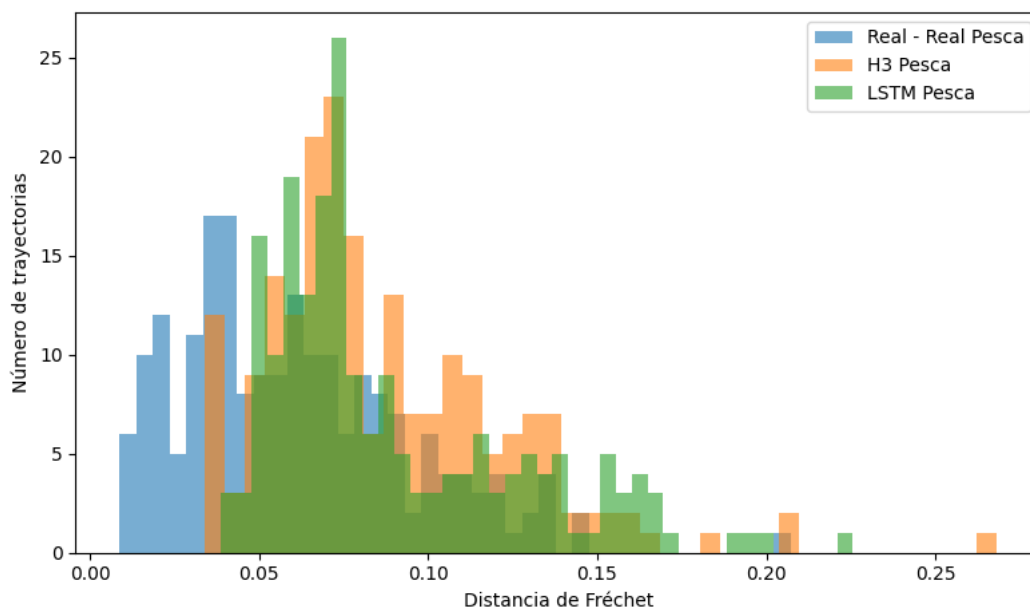


Figura 4-9 Distribución de la distancia de Fréchet para trayectorias de buques tipo *Fishing* (Fuente: propia)

La Figura 4-9 muestra la distribución de las distancias de Fréchet obtenidas al comparar trayectorias reales de buques pesqueros con trayectorias generadas por los modelos H3 y LSTM, así como la comparación real-real, que se utiliza como referencia de la variabilidad natural presente en los datos AIS. A diferencia del caso de los buques tipo *Cargo*, las trayectorias de pesca presentan un comportamiento espacial más irregular, por lo que es esperable observar valores de distancia de Fréchet más elevados incluso entre trayectorias reales.

En la comparación real-real pesca, la mayor parte de los valores se concentra aproximadamente entre 0.03 y 0.08, con una cola que se extiende hasta valores cercanos a 0.12-0.13. Esto indica que, incluso entre trayectorias reales, la similitud geométrica es limitada debido a la naturaleza variable de la actividad pesquera. Este resultado confirma que las trayectorias de pesca presentan una dispersión espacial mayor que las trayectorias de navegación hacia puerto, donde los recorridos suelen ser más regulares y repetitivos.

En cuanto al modelo H3, su distribución aparece desplazada hacia valores ligeramente mayores de distancia de Fréchet respecto a la referencia real-real. La mayor parte de las trayectorias generadas se concentra aproximadamente entre 0.06 y 0.13, aunque se observa una cola que alcanza valores superiores a 0.25. Este comportamiento indica que el modelo introduce una cierta desviación geométrica respecto a las trayectorias reales. No obstante, el núcleo principal de la distribución mantiene un solapamiento significativo con la variabilidad natural de las trayectorias reales, lo que sugiere que el modelo reproduce parcialmente los patrones espaciales del movimiento pesquero.

El modelo LSTM presenta una distribución similar a la del modelo H3, aunque con algunas diferencias destacables. En este caso se observa una mayor concentración de trayectorias en el intervalo aproximado de 0.05-0.08, lo que indica que muchas trayectorias generadas presentan una similitud geométrica razonable con las trayectorias reales. Sin embargo, también se aprecia la presencia de valores más elevados en la cola de la distribución, lo que refleja la existencia de algunos casos con desviaciones espaciales más pronunciadas, aunque la cola es de menor tamaño que en el caso anterior.

En conjunto, el análisis del histograma muestra que ambos modelos son capaces de capturar parcialmente la estructura geométrica de las trayectorias de pesca, aunque con una desviación mayor

que en el caso del tráfico de buques de carga. Así mismo el modelo LSTM presenta un mayor solapamiento con la distribución correspondiente a las trayectorias reales, especialmente en el intervalo donde se concentra la mayor parte de los valores. Esto sugiere que el modelo reproduce ligeramente mejor la similitud geométrica típica de las trayectorias de pesca, aunque ambos modelos muestran un comportamiento global comparable.

4.3.1.2 Método seleccionado según la distancia Fréchet

Los resultados obtenidos a partir de la distancia de Fréchet muestran que ambos modelos presentan un comportamiento similar en la reproducción de la geometría de las trayectorias de buques pesqueros. No obstante, al comparar los valores medios obtenidos, se observa que el modelo H3 presenta una distancia de Fréchet ligeramente inferior a la del modelo LSTM, lo que indica una mayor similitud geométrica entre las trayectorias generadas y las trayectorias reales.

Asimismo, el análisis del percentil 95 muestra que el modelo H3 presenta valores algo menores que el modelo LSTM, lo que sugiere una menor presencia de trayectorias con desviaciones geométricas elevadas. En conjunto, estos resultados indican que, aunque la diferencia entre ambos modelos es reducida, el modelo H3 ofrece un comportamiento global ligeramente más favorable en términos de similitud geométrica respecto a las trayectorias reales de buques de pesca.

4.3.2 Comparación espacial mediante mapas de calor

La Figura 4-10 muestra una comparación de los mapas de calor de densidad de trayectorias correspondientes a los buques tipo *Fishing*. En ella se representan, mediante una discretización espacial hexagonal, las zonas donde se concentran con mayor frecuencia los puntos de las trayectorias reales y de las trayectorias generadas por los modelos H3 y LSTM. Este tipo de representación permite visualizar la distribución espacial del tráfico pesquero y evaluar hasta qué punto los modelos son capaces de reproducir los patrones observados en los datos AIS.

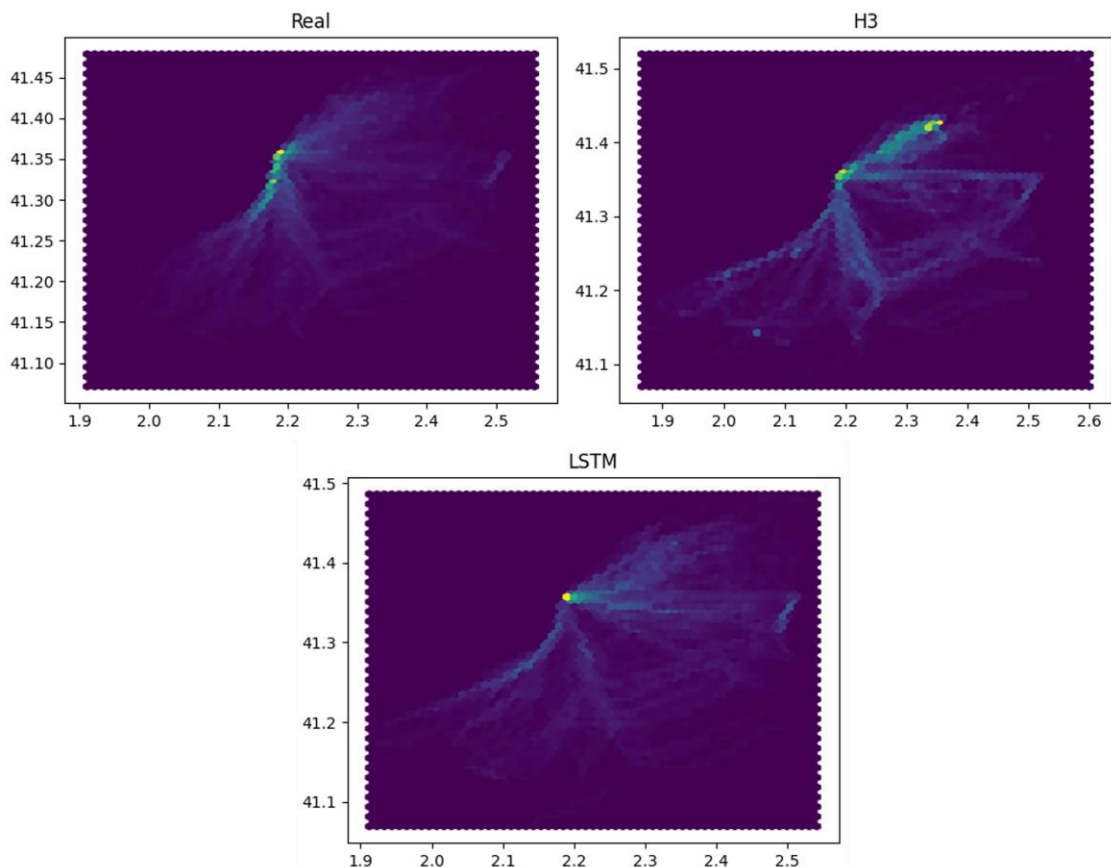


Figura 4-10 Comparación modelos con mapas de calor (Fuente: propia)

En el caso de las trayectorias reales, se observa un núcleo de alta densidad situado en el punto de salida del puerto, desde el cual se extienden múltiples direcciones hacia distintas zonas de actividad. A diferencia de las trayectorias de buques de carga, donde las rutas suelen concentrarse en corredores relativamente definidos, las trayectorias de pesca muestran una dispersión espacial mayor, con varios ramales que se expanden hacia distintas áreas del entorno marítimo. Este patrón refleja el carácter exploratorio de la actividad pesquera, donde los buques se desplazan entre diferentes zonas de pesca y realizan movimientos menos estructurados.

➤ **Mapas de calor+H3:**

En el caso del modelo H3, el mapa de calor muestra una tendencia clara de las trayectorias generadas a dirigirse hacia una zona de actividad situada al noreste del puerto de Barcelona, donde se observa una mayor concentración de puntos. Este comportamiento sugiere que el modelo identifica esta área como una zona principal de actividad pesquera, generando una especie de ruta predominante entre el puerto y dicha zona.

Además, en comparación con las trayectorias reales, se aprecia que las rutas o ramales secundarios aparecen más marcados y definidos. Esto indica que el modelo tiende a generar trayectorias más estructuradas y repetitivas, lo que se traduce en una representación espacial más estable de los desplazamientos de los pesqueros. Como consecuencia, las trayectorias generadas por H3 muestran movimientos más regulares y menos dispersos que los observados en los datos AIS reales.

Este comportamiento puede ser debido a que el modelo basado en celdas H3, favorezca la generación de trayectorias alineadas con determinadas transiciones espaciales frecuentes. De esta forma, el modelo tiende a reforzar las rutas más comunes observadas en los datos, produciendo patrones de movimiento más definidos que los presentes en las trayectorias reales.

➤ **Modelo LSTM:**

Por su parte, el modelo LSTM presenta una distribución espacial próxima a la observada en los datos reales. Las trayectorias generadas muestran una dispersión relativamente suave y continua, que reproduce de forma razonable la estructura general del movimiento pesquero en el área de estudio. En términos generales, la distribución espacial obtenida mantiene cierta similitud con la observada en las trayectorias reales, reflejando razonablemente las principales zonas de actividad y las direcciones predominantes de desplazamiento.

No obstante, se aprecia una tendencia del modelo a concentrar trayectorias en determinadas direcciones preferentes. En particular, se observa que las embarcaciones generadas tienden a abandonar el puerto siguiendo un rumbo predominante hacia el Este, mientras que las trayectorias de retorno muestran una orientación mayoritaria en dirección Oeste. Este comportamiento genera una pequeña ruta más marcada en la zona situada al este del puerto de Barcelona, donde se concentra una mayor densidad de trayectorias generadas.

En conjunto, el patrón espacial producido por el modelo LSTM resulta similar al observado en los datos reales, reproduciendo razonablemente la dispersión general de las trayectorias pesqueras. Sin embargo, algunas rutas aparecen representadas con mayor intensidad, lo que sugiere que el modelo tiende a reforzar determinadas direcciones de desplazamiento presentes en los datos AIS, produciendo una ligera concentración de trayectorias en esos corredores de movimiento predominantes.

➤ **Interpretación final:**

En conjunto, a partir de la comparación de los mapas de calor se observa que el modelo LSTM ofrece una representación espacial más próxima a la observada en los datos AIS. Aunque ambos métodos reproducen la estructura general del movimiento, el modelo LSTM mantiene una dispersión espacial más similar a la de las trayectorias reales, mientras que el modelo H3 tiende a generar rutas más definidas

y estructuradas. Por este motivo, desde el punto de vista de la distribución espacial del tráfico pesquero, el modelo LSTM proporciona el ajuste más adecuado.

4.3.3 Análisis cinemático de las trayectorias mediante MAE, DTW y cálculos estadísticos

4.3.3.1 Análisis gráfico de la velocidad

La Figura 4-11 muestra el perfil medio de velocidad a lo largo de la trayectoria normalizada para los buques tipo *Fishing*, comparando las trayectorias reales con las generadas mediante los modelos H3 y LSTM-KNN. La zona sombreada representa la variabilidad observada en los datos reales, lo que permite contextualizar el comportamiento de los modelos dentro del rango de velocidades presentes en el conjunto AIS.

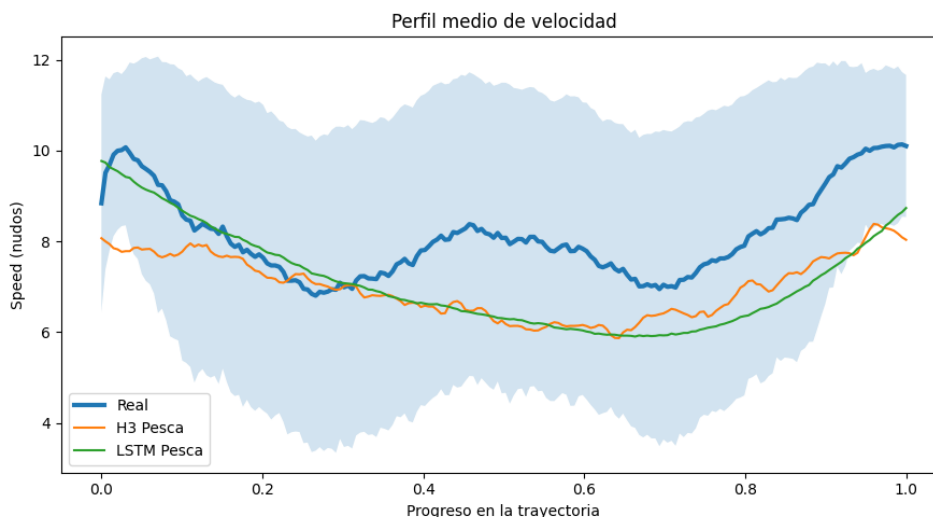


Figura 4-11 Perfil de velocidad buques pesqueros (Fuente: propia)

En las trayectorias reales se observa un comportamiento variable de la velocidad a lo largo del recorrido. Tras una fase inicial con valores cercanos a los 10 nudos, la velocidad disminuye progresivamente hasta situarse en torno a los 7 nudos. Sin embargo, en la parte central de la trayectoria se aprecia un nuevo incremento de la velocidad hasta valores próximos a los 8 nudos, lo que podría estar asociado a fases de operación o desplazamientos entre posibles zonas de pesca, donde los buques modifican su velocidad durante la búsqueda de caladeros. Finalmente, en el tramo final se observa un nuevo aumento de la velocidad a medida que las embarcaciones se aproximan de nuevo al puerto.

➤ Mapas de calor+H3:

El modelo H3 reproduce de forma razonable la tendencia general del perfil de velocidad, mostrando una disminución progresiva en la fase inicial de la trayectoria y un aumento en el tramo final, en línea con el comportamiento observado en los datos reales. Sin embargo, el modelo no representa adecuadamente la fase central de la trayectoria, donde en los datos reales se aprecia un incremento de velocidad hasta valores próximos a los 8 nudos, que puede estar asociado a desplazamientos entre zonas de pesca o a fases de búsqueda de caladeros.

A lo largo del recorrido, el perfil generado por H3 presenta además un comportamiento más suavizado que el observado en las trayectorias reales, aunque sí es capaz de reflejar pequeñas variaciones locales de velocidad, visibles en forma de ligeros picos en la curva. No obstante, en prácticamente toda la trayectoria las velocidades generadas se mantienen por debajo de las reales, lo que indica una tendencia del modelo a subestimar la velocidad media de navegación. Esta diferencia resulta especialmente evidente en algunos tramos finales, donde la curva generada incluso se sitúa fuera del rango de variabilidad representado por la zona sombreada, mostrando una desviación mayor respecto al comportamiento observado en los datos AIS.

➤ **Modelo LSTM-KNN:**

Por su parte, el modelo LSTM-KNN sigue de manera más cercana la evolución general del perfil real de velocidad a lo largo de la trayectoria. En particular, se observa que las velocidades generadas son más altas en los tramos inicial y final, lo que reproduce de forma similar los patrones observados en los datos reales durante las fases de salida del puerto y retorno. Esta característica hace que la forma global de la curva resulte más próxima al comportamiento real en comparación con el modelo H3.

No obstante, al igual que ocurría con el modelo anterior, el perfil generado por LSTM-KNN tampoco reproduce el incremento de velocidad presente en la fase central de la trayectoria, donde en los datos reales se observa un aumento hasta valores cercanos a los 8 nudos. A pesar de esta limitación, la tendencia general del perfil generado se aproxima mejor al observado en las trayectorias reales, situándose además con mayor frecuencia dentro del rango de variabilidad representado por la zona sombreada. En conjunto, el modelo LSTM-KNN ofrece una representación global más coherente del perfil de velocidad, mostrando un comportamiento más próximo al observado en los datos AIS que el obtenido mediante el modelo H3.

➤ **Interpretación general:**

De forma general, puede concluirse que tanto el modelo H3 como el modelo LSTM-KNN son capaces de capturar la tendencia general del perfil de velocidad, reproduciendo las fases principales de la trayectoria. No obstante, ambos modelos tienden a subestimar ligeramente las velocidades medias observadas en los datos reales, generando perfiles algo más suavizados que los registrados en las trayectorias AIS.

Este resultado es coherente con lo observado previamente en el análisis geométrico de las trayectorias. Las trayectorias de buques pesqueros presentan una mayor variabilidad espacial y operativa, asociada a cambios frecuentes de rumbo, desplazamientos entre zonas de pesca y fases de búsqueda de caladeros. Esta mayor complejidad del comportamiento pesquero dificulta la reproducción exacta tanto de trayectorias como de perfiles de velocidad, como se puede ver en las pequeñas desviaciones observadas en los perfiles generados por los modelos.

Sin embargo, perfiles generados se mantienen generalmente dentro del rango de variabilidad observado en los datos reales, representado por la banda sombreada de la figura. Esto indica que, aunque los modelos no reproduzcan con total precisión todas las fluctuaciones de la velocidad real, las velocidades generadas se mantienen en valores físicamente plausibles y coherentes con el comportamiento de los buques pesqueros registrado en los datos AIS.

4.3.3.2 Error absoluto medio (MAE)

Método	MAE (nudos)
Mapa de calor+H3	1.15425
LSTM-KNN	1.10705

Tabla 4-12 Resultados MAE pesqueros

Los resultados obtenidos indican que el error absoluto medio (MAE) entre el perfil de velocidad real y los perfiles generados por los modelos se sitúa aproximadamente en torno a 1 nudo. En concreto, el modelo LSTM-KNN presenta un error ligeramente menor que el modelo H3, lo que sugiere que reproduce con algo más de precisión la evolución media de la velocidad a lo largo de la trayectoria.

Si se compara este error con la velocidad media real, que se sitúa aproximadamente en torno a los 8 nudos, se obtiene un error relativo cercano al 13%. Este valor indica que los modelos son capaces de aproximar de forma razonable la dinámica de velocidad observada en los datos AIS, aunque existe cierta desviación respecto al comportamiento real.

En conjunto, estos resultados confirman lo observado previamente en el análisis gráfico del perfil de velocidad: ambos modelos capturan la tendencia global del comportamiento dinámico, pero presentan dificultades para reproducir con precisión algunas variaciones locales de velocidad, especialmente en la parte central de la trayectoria. Dentro de esta comparación, el modelo LSTM-KNN ofrece un ajuste ligeramente mejor, al presentar una menor desviación media respecto al perfil de velocidad real.

4.3.3.3 Alineamiento dinámico del tiempo (DTW)

Método	DTW
Mapa de calor+H3	11.621991008428807
LSTM-KNN	8.96760280119498

Tabla 4-13 Resultados DTW pesqueros

Los resultados muestran que el modelo LSTM-KNN obtiene una distancia DTW menor que el modelo H3, lo que indica que el perfil de velocidad generado por LSTM-KNN reproduce con mayor fidelidad la evolución temporal del perfil real.

Los valores obtenidos son relativamente elevados, especialmente en comparación con los resultados obtenidos previamente para los buques tipo *Cargo*. Esto indica que la reproducción de la dinámica de velocidad resulta más compleja en el caso de los pesqueros, donde sus trayectorias presentan una mayor variabilidad por las características de la actividad pesquera.

En conclusión, el análisis mediante DTW confirma que el modelo LSTM-KNN reproduce con mayor precisión la dinámica temporal de la velocidad, aunque ambos modelos presentan mayores desviaciones que las observadas en el caso de trayectorias más estructuradas como las de los buques tipo *Cargo*.

4.3.3.4 Cálculos estadísticos

Método	Media(nudos)	Desviación estándar(nudos)
Real	8.132425	3.259622465773452
Mapa de calor+H3	6.998725	2.453532835397766
LSTM-KNN	7.130125	1.69462163457658

Tabla 4-14 Cálculos estadísticos de pesqueros

Los resultados estadísticos muestran que ambos modelos subestiman ligeramente la velocidad media observada en los datos reales. Mientras que el conjunto de trayectorias AIS presenta una velocidad media de 8.13 nudos, los valores generados por los modelos se sitúan alrededor de 7 nudos, lo que supone una diferencia aproximada de 1 nudo. Este comportamiento coincide con lo observado previamente en el análisis gráfico del perfil de velocidad. Una posible explicación es que los modelos generen una mayor proporción de trayectorias en fases de operación o pesca, caracterizadas por velocidades más bajas, en comparación con las fases de tránsito.

En cuanto a la variabilidad de la velocidad, representada por la desviación estándar, los resultados indican que los modelos generan perfiles de velocidad menos variables que los datos reales. El conjunto real presenta una desviación estándar mayor, lo que refleja cambios de velocidad más pronunciados asociados a las distintas fases de la actividad pesquera. Por el contrario, los modelos tienden a producir una dinámica más suavizada, reduciendo la amplitud de los cambios de velocidad. Este comportamiento es especialmente notable en el modelo LSTM-KNN, que presenta la menor variabilidad entre los tres conjuntos analizados.

En conjunto, a partir de los resultados obtenidos puede concluirse que el modelo H3 representa de forma más adecuada el comportamiento de velocidad de los buques pesqueros según los cálculos estadísticos realizados. Aunque este modelo presenta una diferencia algo mayor respecto a la velocidad media real, ofrece un resultado más próximo en términos de desviación estándar, lo que indica una mayor variabilidad en las velocidades generadas. Esta característica resulta clave en el caso de las trayectorias de pesca, donde el comportamiento real se caracteriza por cambios frecuentes de velocidad asociados a las distintas fases de operación y desplazamiento entre zonas de pesca. Por este motivo, la mayor variabilidad representada por el modelo H3 puede considerarse más representativa del comportamiento real de los buques pesqueros que la obtenida mediante el modelo LSTM-KNN.

4.3.3.5 Método seleccionado según velocidad

En el caso de los buques tipo *Fishing*, la reproducción del comportamiento dinámico de la velocidad resulta más compleja que en los buques tipo *Cargo*. Esto se debe a la naturaleza propia de la actividad pesquera, caracterizada por cambios frecuentes de velocidad asociados a las distintas fases de operación, como el tránsito hacia las zonas de pesca, la actividad de faena y el regreso a puerto. Como consecuencia, los resultados obtenidos en la modelización dinámica muestran un ajuste inferior al observado en los buques mercantes, cuyos perfiles de velocidad son más estables y predecibles.

A pesar de ello, ambos modelos analizados generan velocidades físicamente plausibles y dentro del rango observado en los datos reales. Entre ellos, el modelo LSTM-KNN presenta un mejor ajuste temporal del perfil de velocidad, según lo estudiado anteriormente. Por este motivo, se selecciona el modelo LSTM-KNN como el método más adecuado para la generación del comportamiento dinámico de velocidad en buques tipo *Fishing*, al reproducir de forma más fiel la evolución temporal de la velocidad a lo largo de la trayectoria.

4.3.4 Selección del método óptimo para buques tipo Fishing

Método	Mapa de calor+H3	LSTM-KNN
Distancia Fréchet		
Mapas de calor		
Análisis gráfico de la velocidad		
MAE		
DTW		
Cálculos estadísticos		

Tabla 4-15 Comparación modelos tipo pesqueros

MEJOR	
PEOR	

Tabla 4-16 Leyenda

A partir de los resultados recogidos en la Tabla 4-15, se observa que ambos modelos presentan comportamientos diferenciados en función del aspecto analizado. El método basado en mapas de calor con discretización H3 ofrece mejores resultados en la representación espacial de las trayectorias, mostrando una mayor similitud geométrica con las rutas reales. Sin embargo, su comportamiento en el

aspecto dinámico es más limitado, presentando mayores desviaciones en la reproducción del perfil de velocidad.

Por el contrario, el modelo LSTM-KNN muestra un comportamiento espacial ligeramente inferior, pero presenta un mejor rendimiento en la modelización de la velocidad, reproduciendo con mayor fidelidad su evolución temporal.

En términos generales, los resultados obtenidos para los buques tipo *Fishing* no alcanzan el mismo nivel de precisión observado previamente en los buques tipo *Cargo*, debido a la mayor complejidad y variabilidad del comportamiento de las trayectorias pesqueras. No obstante, los resultados pueden considerarse suficientemente representativos del comportamiento real. Dado que en la mayoría de los métodos de evaluación ambos modelos presentan resultados similares, pero el modelo LSTM-KNN muestra un rendimiento ligeramente superior en el conjunto global del análisis, se selecciona finalmente LSTM-KNN como el método más adecuado para representar el comportamiento de los buques pesqueros.

En la Figura 4-12 se muestra una comparación gráfica de los dos métodos para generar trayectorias sintéticas junto con una trayectoria real.

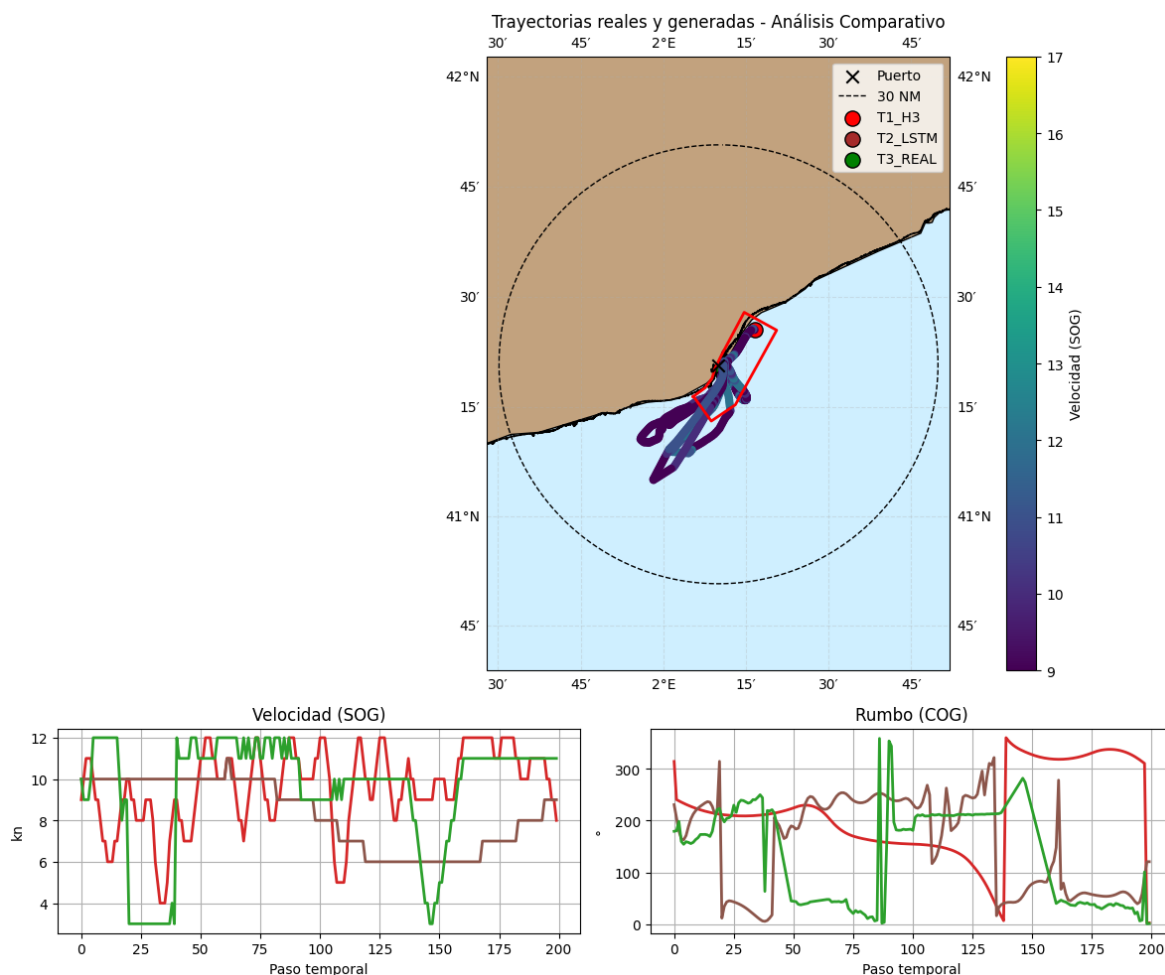


Figura 4-12 Trayectorias generadas con los distintos métodos Fishing (Fuente: propia)

4.4 Representación simulada del entorno operativo del Puerto de Barcelona

La Figura 4-13 presenta una representación conjunta de trayectorias sintéticas generadas para buques tipo *Cargo* y *Fishing* en el entorno del Puerto de Barcelona utilizando los modelos seleccionados en este trabajo. Para los buques de carga se emplea el modelo cGAN + KNN, mientras que para los buques pesqueros se utiliza el modelo LSTM-KNN y el modelo H3. La visualización incluye además la

evolución de variables cinemáticas relevantes, como la velocidad sobre el fondo (SOG) y el rumbo (COG), lo que permite observar de forma integrada el comportamiento espacial y dinámico de las trayectorias generadas.

La representación conjunta de ambas tipologías de buque permite ilustrar cómo la aplicación de los modelos desarrollados posibilita la construcción de un escenario marítimo simulado coherente con el comportamiento observado en los datos AIS reales. En este escenario se reproducen tanto las rutas de acceso características del tráfico mercante como las áreas de operación propias de los buques pesqueros, generando una aproximación realista al área operativa del Puerto de Barcelona. De este modo, se demuestra que los modelos permiten generar buenos escenarios sintéticos de tráfico marítimo, lo que abre la puerta a su aplicación en simulación, análisis del entorno marítimo y generación de escenarios para sistemas de apoyo a la vigilancia marítima.

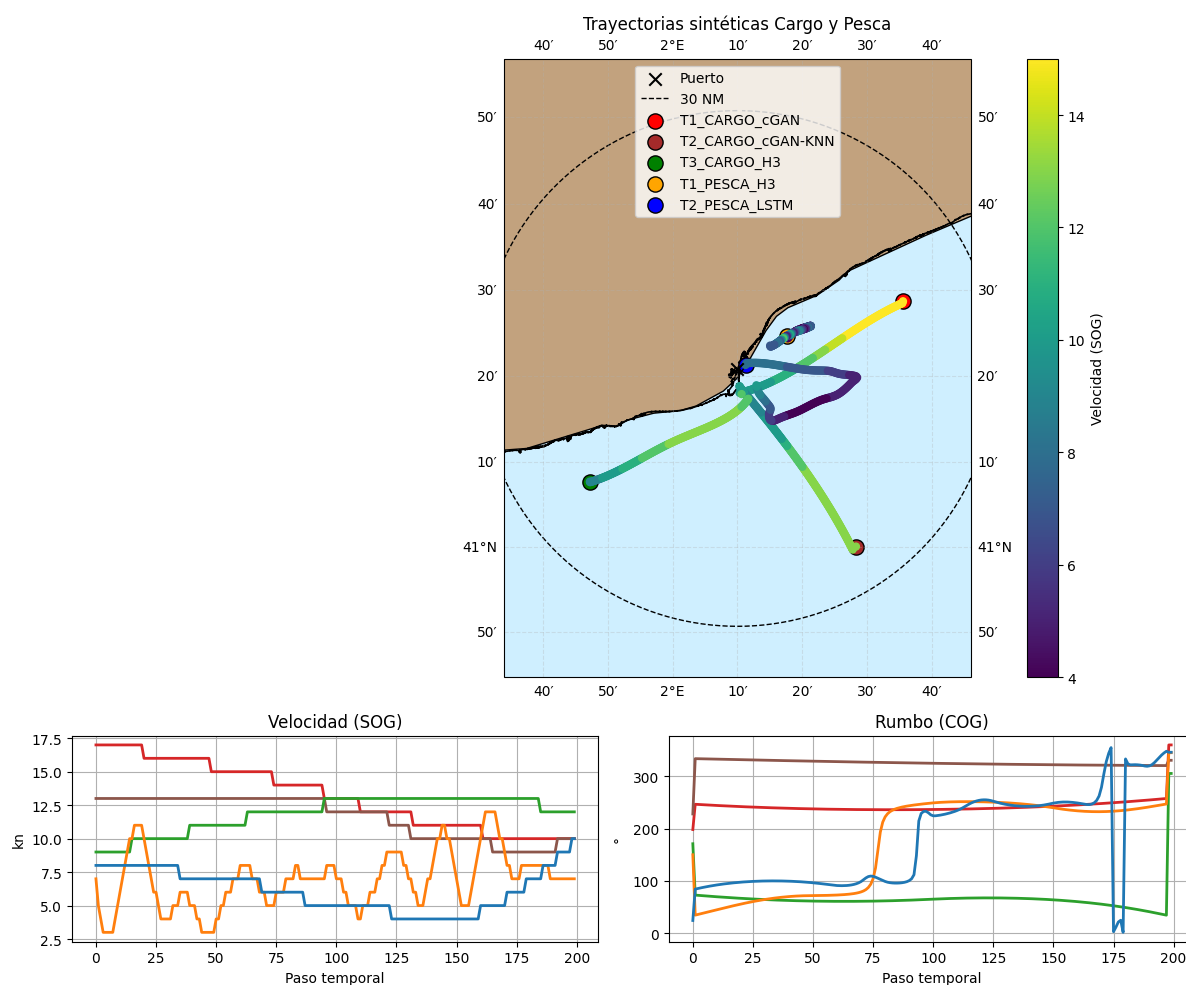


Figura 4-13 Ejemplo simulado de entorno operativo de Barcelona

4.5 Limitaciones del análisis

A pesar de los resultados obtenidos y del buen comportamiento general de los modelos desarrollados, es necesario señalar una serie de limitaciones inherentes al análisis realizado, que deben tenerse en cuenta al interpretar los resultados.

En primer lugar, el estudio se basa en un conjunto de datos AIS limitado a un área geográfica concreta, centrada en las inmediaciones del Puerto de Barcelona. Así mismo, corresponde a un periodo temporal determinado. Aunque el entorno seleccionado constituye un escenario representativo para analizar comportamientos de tráfico marítimo en zonas portuarias, el uso de estos resultados en otras áreas con características operativas distintas debe realizarse con precaución.

Por otra parte, la representación espacial del tráfico marítimo mediante cuadrículas hexagonales H3 introduce una dependencia directa respecto a la resolución espacial seleccionada. Una resolución demasiado baja puede simplificar en exceso la geometría de las trayectorias y ocultar detalles relevantes del movimiento de los buques. De igual forma, resoluciones más finas incrementan el número de celdas y pueden generar regiones con escasa densidad de datos. Esto puede afectar directamente a la estabilidad de los modelos basados en transiciones espaciales y al proceso de aprendizaje de los modelos generativos.

Así mismo, el rendimiento de los modelos implementados está condicionado por el tamaño y la diversidad del conjunto de trayectorias disponibles. Este aspecto resulta especialmente relevante en el caso de los buques tipo *Fishing*, cuyo comportamiento presenta una mayor variabilidad espacial y temporal asociada a la actividad de pesca. Un mayor volumen de trayectorias reales permitiría capturar con mayor precisión esta variabilidad y mejorar la capacidad de los modelos para representar patrones complejos de movimiento.

Finalmente, aunque los modelos desarrollados han demostrado ser capaces de generar trayectorias coherentes con los datos AIS originales, su rendimiento puede verse influido por la configuración de los parámetros del modelo y por las características específicas del conjunto de datos utilizado. En este sentido, futuros trabajos podrían ampliar el análisis incorporando conjuntos de datos AIS más extensos, diferentes áreas geográficas o evaluando de forma sistemática el efecto de distintas resoluciones espaciales en la discretización H3. Todo esto con el fin de mejorar la robustez y generalización de los modelos propuestos.

5 CONCLUSIONES Y LÍNEAS FUTURAS

5.1 Conclusiones

El presente Trabajo de Fin de Grado ha abordado el problema de la simulación del comportamiento de distintas clases de buques a partir de datos AIS reales, con el objetivo de generar trayectorias sintéticas capaces de reproducir de forma realista las dinámicas de navegación observadas en el entorno del Puerto de Barcelona, uno de los escenarios portuarios con mayor densidad y diversidad de tráfico del Mediterráneo. Para ello se ha desarrollado un proceso completo que abarca desde el tratamiento y filtrado de los datos AIS hasta la implementación y validación de distintos modelos de generación de trayectorias, integrando técnicas de análisis de datos, modelado probabilístico y aprendizaje automático.

El análisis realizado ha permitido caracterizar el comportamiento de dos tipologías de buque con dinámicas claramente distintas: buques tipo *Cargo* y buques tipo *Fishing*. Mientras que los buques mercantes presentan trayectorias más estructuradas y predecibles asociadas a corredores de acceso portuario, los buques pesqueros muestran patrones de movimiento más irregulares y variables, condicionados por la propia actividad de pesca. Estas diferencias estructurales han puesto de manifiesto que la idoneidad de los modelos depende en gran medida del tipo de tráfico analizado, evidenciando que no existe un único enfoque óptimo para representar todos los comportamientos en la mar.

En el caso de los buques de carga, los resultados obtenidos muestran que los modelos generativos basados en redes generativas antagónicas condicionales (cGAN) combinadas con estimación cinemática mediante KNN permiten reproducir con gran fidelidad tanto la geometría de las trayectorias como sus características dinámicas. Por otro lado, para los buques pesqueros, las arquitecturas recurrentes LSTM con KNN han demostrado una mayor capacidad para capturar la naturaleza secuencial y variable de sus movimientos. Aunque los resultados obtenidos con los pesqueros no alcanzan el mismo nivel de precisión que en el caso de los buques mercantes, los modelos desarrollados logran representar de forma suficientemente realista los patrones espaciales y dinámicos característicos de este tipo de tráfico.

De forma global, el trabajo demuestra que es posible generar escenarios sintéticos de tráfico marítimo coherentes con los datos AIS reales, capaces de reproducir tanto las rutas de acceso características del tráfico mercante como las áreas de operación propias de los buques pesqueros. Este tipo de representaciones puede ser una herramienta de gran valor para el análisis del entorno marítimo, ya que permite construir escenarios de navegación realistas que pueden emplearse en estudios de tráfico, en sistemas de apoyo al conocimiento del entorno marítimo o en entornos de simulación para entrenamiento y análisis operativo.

5.2 Líneas futuras

El desarrollo del presente Trabajo Fin de Grado ha permitido diseñar y comparar distintos modelos generativos orientados a la simulación del comportamiento de buques a partir de datos AIS reales. No obstante, la complejidad del entorno marítimo y el carácter emergente de las técnicas empleadas abren múltiples vías de ampliación y mejora que permitirían incrementar tanto el realismo de las simulaciones como su aplicabilidad operativa. Por ello, se proponen las siguientes líneas futuras:

- Integración del modelo generativo en simuladores de navegación con el fin de generar escenarios dinámicos basados en patrones reales de tráfico, aumentando el realismo de los ejercicios de adiestramiento en entornos portuarios complejos.
- Generación de escenarios simulados para un día del año concreto y una franja horaria determinada, partiendo del análisis temporal los datos AIS registrados en un entorno marítimo específico.
- Estudiar la incorporación de nuevas variables de entrada al modelo, como condiciones meteorológicas, restricciones normativas de tráfico marítimo o información batimétrica, así como información conocida del entorno a partir de publicaciones náuticas como derroteros, con el fin de aumentar el realismo de las trayectorias generadas.
- Analizar la aplicación de modelos generativos AIS en escenarios operativos, estudiando su posible empleo en la construcción de entornos marítimos AIS sintéticos orientados a la alteración deliberada de la percepción operativa adversaria, junto con el desarrollo de contramedidas que garanticen la resiliencia de los propios sistemas del COVAM.
- Ampliar el estudio incorporando nuevas tipologías de buques en el proceso de modelado y generación de trayectorias sintéticas, incluyendo embarcaciones como veleros, ferris, buques de pasajeros u otros tipos de tráfico marítimo. Esta ampliación permitiría analizar patrones de navegación más diversos y construir simulaciones del tráfico marítimo más completas y representativas del entorno operativo real.

En conjunto, estas líneas futuras consolidan el trabajo realizado como una base sólida para el desarrollo de herramientas avanzadas de simulación y análisis del tráfico marítimo, con potencial aplicación tanto en el ámbito académico como en el operativo, contribuyendo al fortalecimiento de las capacidades tecnológicas asociadas al entorno marítimo.

6 BIBLIOGRAFÍA

- [1] SVB Marine, «Todo lo que debe saber sobre AIS,» [En línea]. Available: <https://www.svb-marine.es/es/guia/todo-lo-que-debe-saber-sobre-ais.html>. [Último acceso: 15 enero 2026].
- [2] International Maritime Organization (IMO), «Resolution A.1106(29): Revised guidelines for the operational use of AIS,» 2015. [En línea]. Available: [https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution%20A.1106\(29\).pdf](https://wwwcdn.imo.org/localresources/en/OurWork/Safety/Documents/AIS/Resolution%20A.1106(29).pdf). [Último acceso: 15 enero 2026].
- [3] International Telecommunication Union (ITU), «Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band,» Recommendation ITU-R M.1371-5, 2014. [En línea]. Available: <https://www.itu.int/rec/R-REC-M.1371>. [Último acceso: 16 enero 2026].
- [4] R. Pallotta, M. Vespe y K. Bryan, «Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction,» *Entropy*, vol. 15, no. 6, pp. 2218-2245, 2013. [En línea]. Available: <https://www.mdpi.com/1099-4300/15/6/2218>. [Último acceso: 16 enero 2026].
- [5] Náutica Profesional, «Sistema AIS en la navegación», Náutica Profesional, 2023. [En línea]. Available: <https://nauticaprofesional.com/sistema-ais-en-la-navegacion/>. [Último acceso: 3 febrero 2026].
- [6] Wikiciencias, «Clasificación de buques: características clave y funciones marítimas,» [En línea]. Disponible en: <https://www.wikiciencias.net/clasificacion-de-buques-caracteristicas-clave-y-funciones-maritimas/>. [Último acceso: 16 marzo 2026].
- [7] United Nations Conference on Trade and Development (UNCTAD), «Review of Maritime Transport.» [En línea]. Available: <https://unctad.org/topic/transport-and-trade-logistics/review-of-maritime-transport>. [Último acceso: 16 enero 2026].
- [8] European Maritime Safety Agency (EMSA), «The 2025 European Maritime Safety Report,» 2025. [En línea]. Available: <https://www.emsa.europa.eu/publications/item/5598-emsafe-report-2025.html>. [Último acceso: 17 enero 2026].
- [9] European Maritime Safety Agency (EMSA), «The 2025 European Maritime Safety Report / Facts and Figures,» 2025. [En línea]. Available: <https://www.emsa.europa.eu/publications/download/8398/5596/23.html>. [Último acceso: 17 enero 2026].

- [10] R. Pallotta, M. Vespe y K. Bryan, «Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction,» *Entropy*, vol. 15, no. 6, pp. 2218-2245, 2013. [En línea]. Available: <https://www.mdpi.com/1099-4300/15/6/2218>. [Último acceso: 17 enero 2026].
- [11] C. M. Bishop, *Pattern Recognition and Machine learning*, Springer, 2006.
- [12] Uber Technologies, «H3: A Hexagonal Hierarchical Spatial Index,» 2018. [En línea]. Available: <https://h3geo.org/docs/>. [Último acceso: 26 enero 2026].
- [13] S. Russell y P. Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson, 2021.
- [14] IBM Cloud Education, «Machine Learning,» IBM, 2020. [En línea]. Available: <https://www.ibm.com/es-es/cloud/learn/machine-learning#toc-cmo-funcio-2KNEOt4z>. [Último acceso: 28 enero 2022].
- [15] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*, MIT Press, 2016. [En línea]. Available: <https://www.deeplearningbook.org/>. [Último acceso: 26 enero 2026].
- [16] European Commission, Joint Research Centre, *Artificial Intelligence: A European Perspective*, Publications Office of the European Union, 2018. [En línea]. Available: <https://publications.jrc.ec.europa.eu/repository/handle/JRC113826>. [Último acceso: 25 enero 2026].
- [17] S. Hochreiter y J. Schmidhuber, «Long short-term memory,» *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997. [En línea]. Available: <https://www.bioinf.jku.at/publications/older/2604.pdf> [Último acceso: 20 enero 2026].
- [18] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink y J. Schmidhuber, «LSTM: A Search Space Odyssey,» *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, 2017. [En línea]. Available: <https://arxiv.org/abs/1503.04069>. [Último acceso: 25 enero 2026].
- [19] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer, 2009.
- [20] Scikit-learn Developers, «Nearest Neighbors» *scikit-learn documentation*. [En línea]. Available: <https://sklearn.org/stable/modules/neighbors.html>. [Último acceso: 25 enero 2026].
- [21] L. Frank y S. Mohamed, *Generative Adversarial Networks (GANs) for Synthetic Test Data*, Jun. 2024.
- [22] K. V. Keerthana y S. Boopathi Raja, «GANs for Synthetic Data Generation: Advancements and Challenges using Machine Learning,» *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 10, no. 1, 2024. [En línea]. Available: <https://www.ijariit.com/manuscripts/v10i1/V10I1-1234.pdf>. [Último acceso: 26 enero 2026].
- [23] T. König, A. Mangaluru Ramananda, F. Wagner y M. Kley, «A LSTM-GAN Algorithm for Synthetic Data Generation of Time Series Data for Condition Monitoring», *Procedia Computer Science*, vol. 246, pp. 1508–1517, 2024. [En línea]. Disponible en: <https://www.sciencedirect.com/science/article/pii/S187705092402653X>. [Último acceso: 24 marzo 2026].
- [24] A. Vaswani et al., «Attention Is All You Need,» in *Advances in Neural Information Processing Systems*, vol. 30, 2017, pp. 5998-6008. [En línea]. Available:

- <https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf>. [Último acceso: 26 enero 2026].
- [25] Departamento de Seguridad Nacional (DSN), «Estrategia Nacional de Seguridad Marítima 2024,» Gobierno de España, 2024. [En línea]. Available: <https://www.dsn.gob.es/sites/default/files/2025-03/ENSM%202024%20Castellano%20Accesible.pdf>. [Último acceso: 27 enero 2026].
- [26] Defensa.com, «Centro de Operaciones de Vigilancia y Acción Marítima de la Armada,» [En línea]. Available: <https://www.defensa.com/en-abierto/centro-operaciones-vigilancia-accion-maritima-armada>. [Último acceso: 27 enero 2026].
- [27] Armada Española, Revista General de Marina, «Artículos sobre inteligencia artificial y transformación digital en la Armada,» [En línea]. Available: <https://armada.defensa.gob.es/archivo/rgm/>. [Último acceso: 27 enero 2026].
- [28] Ministerio de Defensa - Dirección General de Armamento y Material (DGAM), «Programas de modernización de sistemas de mando y control,» [En línea]. Available: <https://www.defensa.gob.es/>. [Último acceso: 27 enero 2026].
- [29] Ministerio de Defensa, «Estrategias de innovación y transformación digital en el ámbito de la defensa,» [En línea]. Available: <https://www.defensa.gob.es/defensa/estrategias/>. [Último acceso: 27 enero 2026].
- [30] La Verdad, «La Armada invertirá millones en inteligencia artificial para el COVAM,» [En línea]. Available: <https://www.laverdad.es/murcia/cartagena/armada-invertira-millones-inteligencia-artificial-para-covam-20191011235656-ga.html>. [Último acceso: 27 enero 2026].
- [31] Acronáutica, «5 consejos para navegar a vela,» [En línea]. Available: <https://www.acronautica.com/5-consejos-para-navegar-a-vela/>. [Último acceso: 29 enero 2026].
- [32] A. Tripathi, «What is the main difference between RNN and LSTM?,» [En línea]. Available: <https://ashutoshtripathi.com/2021/07/02/what-is-the-main-difference-between-rnn-and-lstm-nlp-rnn-vs-lstm/>. [Último acceso: 29 enero 2026].
- [33] EA1HLH, «Sistema AIS,» [En línea]. Available: <https://ea1hlh.es/blog/sistema-ais/>. [Último acceso: 29 enero 2026].
- [34] PocketMariner, «Latest global AIS coverage map,» [En línea]. Available: <https://pocketmariner.com/2018/02/06/latest-global-ais-coverage-map/>. [Último acceso: 29 enero 2026].
- [35] Costa Cruceros, «Cruceros transatlánticos,» [En línea]. Available: <https://www.costacruceros.es/destinos/transatlanticos.html>. [Último acceso: 27 enero 2026].
- [36] Puertos de Canarias, «El puerto de Granadilla refuerza su actividad con reparaciones de buques,» [En línea]. Available: <https://www.puertosdec Canarias.es/2024/04/24/puerto-granadilla-reparaciones-buques-1777.html>. [Último acceso: 27 enero 2026].
- [37] DHL Global Forwarding, «Tipos y tamaños de buques de carga,» [En línea]. Available: <https://www.dhl.com/es-es/home/global-forwarding/centro-de-formacion-freight-forwarding/tipos-y-tamanos-de-buques-de-carga.html>. [Último acceso: 27 enero 2026].

- [38] Science Learning Hub, «Neural network diagram,» [En línea]. Available: <https://www.sciencelearn.org.nz/images/5156-neural-network-diagram>. [Último acceso: 27 enero 2026].
- [39] ResearchGate, «Figura 26: Diagrama de una neurona artificial,» [En línea]. Available: https://www.researchgate.net/figure/Figura-26-Diagrama-de-una-neurona-artificial_fig16_317278546. [Último acceso: 27 enero 2026].
- [40] De la Torre AI, «Los transformers en la inteligencia artificial: una explicación sencilla,» [En línea]. Available: <https://delatorre.ai/los-transformers-en-la-inteligencia-artificial-una-explicacion-sencilla/>. [Último acceso: 27 enero 2026].
- [41] Blogger, «F-104 SPS Méndez Núñez,» [En línea]. Available: <https://1.bp.blogspot.com/-W2-jZ9GGzGo/YKAXoUohkCI/AAAAAABkKg/b96frlt-J9cJsotcdBBli6yPXYxUZrHQCLcBGAsYHQ/w640-h632/F-104-SPS-Mendez-Nunez-002.jpeg>. [Último acceso: 27 enero 2026].
- [42] Jupyter, [En línea]. Available: <https://jupyter.org/>. [Último acceso: 27 enero 2026].
- [43] EBISEducation, «JupyterLab: qué es y cómo funciona,» [En línea]. Available: <https://www.ebiseducation.com/jupyterlab-que-es-y-como-funciona>. [Último acceso: 27 enero 2026].
- [44] Cloudflare, «¿Cómo funcionan los CAPTCHA?,» [En línea]. Available: <https://www.cloudflare.com/es-es/learning/bots/how-captchas-work/>. [Último acceso: 27 enero 2026].
- [45] G. Romero Jiménez, «El CAPTCHA y tu contribución (anónima) en el entrenamiento sistemas de learning» LegalToday, 2021. [En línea]. Available: <https://www.legaltoday.com/legaltech/novedades-legaltech/el-captcha-y-tu-contribucion-anonima-en-el-entrenamiento-sistemas-de-machine-learning-2021-02-11/>. [Último acceso: 27 enero 2026].
- [46] Education Wiki, «What is supervised learning?,» [En línea]. Available: <https://es.education-wiki.com/5566119-what-is-supervised-learning>. [Último acceso: 27 enero 2026].
- [47] Google Cloud, «¿Qué es el aprendizaje no supervisado?,» [En línea]. Available: <https://cloud.google.com/discover/what-is-unsupervised-learning?hl=es-419>. [Último acceso: 27 enero 2026].
- [48] AprendeIA, «Diferencia entre aprendizaje supervisado y no supervisado,» [En línea]. Available: <https://aprendeia.com/2018/06/15/diferencia-entre-aprendizaje-supervisado-y-no-supervisado/>. [Último acceso: 29 enero 2026].
- [49] Oracle, «Aprendizaje semisupervisado,» [En línea]. Available: <https://www.oracle.com/latam/artificial-intelligence/machine-learning/semi-supervised-learning/>. [Último acceso: 07 febrero 2026].
- [50] Animalia Life Club, «Aprendizaje semi-supervisado,» [En línea]. Available: <https://esa.animalia-life.club/aprendizaje-semi-supervisado>. [Último acceso: 07 febrero 2026].
- [51] De la Torre AI, «Qué es el aprendizaje por refuerzo y cómo funciona,» [En línea]. Available: <https://delatorre.ai/que-es-el-aprendizaje-por-refuerzo-y-como-funciona/>. [Último acceso: 08 febrero 2026].

- [52] IBM, «Historia de la inteligencia artificial,» [En línea]. Available: <https://www.ibm.com/es-es/think/topics/history-of-artificial-intelligence>. [Último acceso: 11 febrero 2026].
- [53] G. Rodríguez Casajús, «Análisis y aplicación de técnicas de inteligencia artificial en sistemas marítimos,» 2022. [En línea]. Available: <https://calderon.cud.uvigo.es/server/api/core/bitstreams/6f896f73-c9ad-41b7-8029-f5fc290651b1/content>. [Último acceso: 17 febrero 2026].
- [54] W. X. Zhao, J. Zhou, Z. Li, X. Tang, X. Wang, Y. Hou et al., «Large Language Models: A Survey,» ACM Computing Surveys, 2024. [En línea]. Available: <https://dl.acm.org/doi/pdf/10.1145/3777411>. [Último acceso: 03 febrero 2026].
- [55] Space-AIS, «Is AIS mandatory?,» [En línea]. Available: <https://space-ais.com/faq-glossary/ais-mandatory/>. [Último acceso: 03 febrero 2026].
- [56] U.S. Coast Guard Navigation Center, «AIS Class A Reports,» [En línea]. Available: <https://www.navcen.uscg.gov/ais-class-a-reports>. [Último acceso: 03 febrero 2026].
- [57] S. Universidades, «Python: qué es y por qué deberías aprender a utilizarlo,» Santander, 2021. [En línea]. Available: <https://www.becas-santander.com/es/blog/python-que-es.html>. [Último acceso: 03 febrero 2026].
- [58] Anaconda, Inc., «Miniconda,» [En línea]. Available: <https://www.anaconda.com/docs/getting-started/miniconda>. [Último acceso: 05 febrero 2026].
- [59] M. Hands, «Introducción a Conda, Anaconda y Miniconda,» MikiHands Blog, 2025. [En línea]. Available: <https://blog.mikihands.com/es/whitedec/2025/4/25/intro-to-conda-anaconda-miniconda/>. [Último acceso: 03 de febrero 2026].
- [60] Autoridad Portuaria de Barcelona, «Memoria estadística 2024,» [En línea]. Available: https://www.portdebarcelona.cat/sites/default/files/content/file/2025/08/01/98/2024_mem_estadistica_es.pdf. [Último acceso: 10 febrero 2026].
- [61] Institut Cartogràfic i Geològic de Catalunya (ICGC), «Línea de costa,» [En línea]. Available: <https://www.icgc.cat/es/Geoinformacion-y-mapas/Datos-y-productos/Geoinformacion-cartografica/Linea-de-costa>. [Último acceso: 12 febrero 2026].
- [62] Google, «Google Maps,» [En línea]. Available: <https://www.google.es/maps>. [Último acceso: 27 febrero 2026].
- [63] Generalitat de Catalunya, Ports de la Generalitat, «Mapa y sistema portuario catalán,» [En línea]. Available: <https://ports.gencat.cat/es/mapa-y-sistema-portuario-catalan/#puntZonaCentre>. [Último acceso: 12 febrero 2026].
- [64] Keene State College, «Campus Map Tool,» [En línea]. Available: <https://www.keene.edu/campus/maps/tool/>. [Último acceso: 14 febrero 2026].
- [65] CSUC Dataverse, «Dataset with persistent identifier doi:10.34810/data2026,» [En línea]. Available: <https://dataverse.csuc.cat/dataset.xhtml?persistentId=doi:10.34810/data2026>. [Último acceso: 28 febrero 2026].
- [66] L. Zhao, G. Shi y J. Yang, “Ship Trajectories Pre-processing Based on AIS Data,” Journal of Navigation, vol. 71, pp. 1–21, 2018. [En línea]. Available: https://www.researchgate.net/publication/324689329_Ship_Trajectories_Pre-processing_Based_on_AIS_Data. [Último acceso: 28 febrero 2026].

- [67] ResearchGate, «A 5000×4000 m² view of a wrap-around model of honeycomb (regular hexagon),» [En línea]. Available: https://www.researchgate.net/figure/A-5000-4000-m-2-view-of-a-Wrap-Around-model-of-honeycomb-the-red-line-regular-hexagon_fig2_327214690. [Último acceso: 28 febrero 2026].
- [68] Liora, «Cadena de Markov: ¿qué es y cómo funciona?», Liora.io, 2024. [En línea]. Available: <https://liora.io/es/cadena-de-markov-que-es>. [Último acceso: 03 marzo 2026].
- [69] Economipedia, «Cadena de Markov», Economipedia, 2024. [En línea]. Available: <https://economipedia.com/definiciones/cadena-de-markov.html>. [Último acceso: 04 marzo 2026].
- [70] H. Anaya Sánchez, “Aumento de imágenes usando una GAN condicional para clasificación de imágenes médicas” Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), [En línea]. Disponible en: https://inaoe.repositorioinstitucional.mx/jspui/bitstream/1009/2646/1/ANAYASH_MCC.pdf. [Último acceso: 27 enero 2026].
- [71] Statistics Easily, “¿Qué es la distancia de inicio de FID (Fréchet)?”, [En línea]. Disponible en: <https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-es-la-distancia-de-inicio-de-Fid-Frechet%3F/>. [Último acceso: 27 enero 2026].
- [72] Zhihu, “知乎专栏文章,” [En línea]. Disponible en: <https://zhuanlan.zhihu.com/p/20159963>. [Último acceso: 28 febrero 2026].
- [73] Statistics Easily, “¿Qué es la guía completa de visualización de mapas de calor?” [En línea]. Disponible en: <https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-es-la-gu%C3%ADa-completa-de-visualizaci%C3%B3n-de-mapas-de-calor%3F/>. [Último acceso: 6 marzo 2026].
- [74] GeeksforGeeks, “Dynamic Time Warping (DTW) in Time Series,” [En línea]. Disponible en: <https://www.geeksforgeeks.org/machine-learning/dynamic-time-warping-dtw-in-time-series/>. [Último acceso: 6 marzo 2026].
- [75] MathWorks, “Dynamic Time Warping (dtw),” [En línea]. Disponible en: <https://es.mathworks.com/help/signal/ref/dtw.html>. [Último acceso: 6 marzo 2026].
- [76] R. Tavenard, “Dynamic Time Warping,” [En línea]. Disponible en: <https://rtavenar.github.io/blog/dtw.html> [Último acceso: 6 marzo 2026].
- [77] Statistics Easily, “¿Qué significa error absoluto medio (MAE)?,” [En línea]. Disponible en: <https://es.statisticseasily.com/glosario/%C2%BFQu%C3%A9-significa-error-absoluto-mae%3F/>. [Último acceso: 6 marzo 2026].
- [78] DataCamp, “Mean Absolute Error (MAE),” [En línea]. Disponible en: <https://www.datacamp.com/es/tutorial/mean-absolute-error>. [Último acceso: 6 marzo 2026].
- [79] A. Radford, L. Metz, y S. Chintala, «Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks,» arXiv preprint arXiv:1511.06434, 2015. [En línea]. Disponible en: <https://arxiv.org/abs/1511.06434>. [Último acceso: 10 marzo 2026].
- [80] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization,» 3rd International Conference on Learning Representations (ICLR), 2015. [En línea]. Disponible en: <https://arxiv.org/abs/1412.6980>. [Último acceso: 10 marzo 2026].

- [81] W. Zhang, W. Jiang, Q. Liu, y W. Wang, «AIS data repair model based on generative adversarial network,» Reliability Engineering & System Safety, vol. 240, p. 109572, Dic. 2023.
- [82] OpenAI, “ChatGPT,” herramienta de inteligencia artificial, [En línea]. Disponible en: <https://chatgpt.com/>. [Último acceso: 16 marzo 2026].
- [83] Google, “Gemini,” herramienta de inteligencia artificial, [En línea]. Disponible en: <https://gemini.google.com/app>. [Último acceso: 16 marzo 2026].
- [84] Google, “NotebookLM,” herramienta de inteligencia artificial, [En línea]. Disponible en: <https://notebooklm.google.com/>. [Último acceso: 16 marzo 2026].
- [85] MathWorks, “smooth,” [En línea]. Disponible en: <https://es.mathworks.com/help/driving/ref/scenariobuilder.trajectory.smooth.html>. [Último acceso: 24 marzo 2026].

ANEXO I: IMPLICACIONES SOCIALES, Y/O ECONÓMICAS, Y/O AMBIENTALES

El presente trabajo desarrolla un sistema de simulación de trayectorias de buques a partir de datos AIS reales, combinando modelos probabilísticos de transición espacial y modelos generativos basados en aprendizaje profundo. Aunque su objetivo principal es metodológico y técnico, el tipo de herramientas desarrolladas presenta implicaciones relevantes desde el punto de vista social, económico y ambiental.

En el **ámbito social**, la modelización del tráfico marítimo permite analizar patrones de navegación y reproducir escenarios de comportamiento sin intervenir sobre el tráfico real. Este tipo de simulación puede contribuir a mejorar la comprensión del funcionamiento de zonas portuarias complejas y facilitar estudios de planificación y análisis de riesgos. La posibilidad de generar trayectorias sintéticas coherentes con los patrones observados permite evaluar escenarios hipotéticos (por ejemplo, incrementos de tráfico o cambios en rutas habituales) sin comprometer la seguridad operacional. No obstante, la incorporación de modelos basados en inteligencia artificial en entornos de análisis marítimo exige cautela en su interpretación, ya que los resultados generados deben entenderse como aproximaciones estadísticas y no como predicciones deterministas. La validación rigurosa y la supervisión técnica resultan esenciales para evitar una dependencia excesiva de modelos automatizados.

Desde la **perspectiva económica**, el transporte marítimo constituye un pilar fundamental del comercio internacional. La capacidad de modelizar trayectorias y simular comportamientos de distintas clases de buques puede servir como herramienta de apoyo en estudios de optimización portuaria, planificación de infraestructuras o análisis de eficiencia operativa. La generación de datos sintéticos permite realizar pruebas de planificación sin necesidad de esperar a que se produzcan determinadas configuraciones reales de tráfico, lo que puede reducir costes asociados a estudios experimentales y facilitar la toma de decisiones estratégicas. Asimismo, una mejor comprensión de los patrones espaciales de navegación puede contribuir indirectamente a mejorar la gestión de flujos y reducir situaciones de congestión.

En el **ámbito ambiental**, la optimización y planificación del tráfico marítimo están estrechamente relacionadas con el consumo energético y las emisiones asociadas. Aunque el presente trabajo no aborda directamente la reducción de emisiones, la capacidad de simular trayectorias y analizar patrones espaciales puede constituir una herramienta útil para evaluar configuraciones que minimicen desplazamientos innecesarios o tiempos de espera prolongados en zonas portuarias. De forma indirecta, una mejor planificación del tráfico puede contribuir a una gestión más eficiente del combustible y a la reducción del impacto ambiental asociado a la actividad marítima.

En conjunto, el desarrollo de modelos de simulación basados en datos AIS representa una herramienta con potencial impacto positivo en términos de seguridad, eficiencia y sostenibilidad, siempre que su aplicación práctica se realice bajo criterios de validación técnica rigurosa y uso responsable de la información generada.

ANEXO II: REFLEXIONES ÉTICAS Y SOCIALES

El desarrollo de modelos de simulación de trayectorias a partir de datos AIS mediante técnicas probabilísticas y redes generativas implica una serie de consideraciones éticas que deben ser analizadas con rigor. Aunque el presente trabajo tiene un carácter académico y metodológico, la naturaleza de la tecnología empleada exige una reflexión responsable sobre su posible aplicación en entornos reales.

En primer lugar, el uso de datos AIS, aun siendo información pública, requiere un tratamiento responsable. Estos datos reflejan patrones operativos reales de navegación y actividad económica, por lo que su análisis debe orientarse a fines legítimos de estudio, planificación o mejora operativa. La modelización desarrollada no pretende replicar comportamientos individuales concretos, sino capturar patrones agregados, minimizando así riesgos asociados a la exposición de información sensible.

En segundo lugar, los modelos generativos implementados producen trayectorias sintéticas coherentes desde el punto de vista estadístico, pero no constituyen predicciones deterministas ni sustituyen el análisis experto. Desde una perspectiva ética, es fundamental evitar la sobre confianza en sistemas automatizados y mantener siempre la supervisión humana en cualquier posible aplicación práctica. La inteligencia artificial debe entenderse como herramienta de apoyo a la decisión, no como reemplazo del criterio profesional.

Asimismo, toda tecnología generativa posee un potencial dual: puede emplearse para mejorar la planificación, la formación o el análisis estratégico, pero también podría utilizarse de manera inadecuada si no existe un marco institucional y normativo claro. Por ello, su implementación debe ir acompañada de principios de responsabilidad, validación técnica y control de uso.

En conclusión, la aplicación de modelos de inteligencia artificial en el ámbito marítimo debe sustentarse en criterios de transparencia metodológica, comprensión de sus limitaciones y supervisión técnica continua, garantizando que su utilización contribuya al interés general y no genere efectos adversos derivados de una confianza indebida o un uso inapropiado.

ANEXO III: SIGLAS, ACRÓNIMOS Y ABREVIATURAS

- AIS** - Automatic Identification System (Sistema de Identificación Automática)
- ALMART** - Almirante de Acción Marítima
- ANN** - Artificial Neural Network (Red neuronal artificial)
- API** - Application Programming Interface
- CAPTCHA** - Completely Automated Public Turing test to tell Computers and Humans Apart
- CEM** - Conocimiento del Entorno Marítimo
- cGAN** - Conditional Generative Adversarial Network
- CMOM** - Comandante del Mando Operativo Marítimo
- COG** - Course Over Ground (Rumbo sobre el Fondo)
- COVAM** - Centro de Operaciones y Vigilancia de la Acción Marítima
- CPA** - Closest Point of Approach
- CSV** - Comma-Separated Values
- DGAM** - Dirección General de Armamento y Material
- DL** - Deep Learning (Aprendizaje profundo)
- DTW** - Dynamic Time Warping (alineamiento dinámico del tiempo)
- EMSA** - European Maritime Safety Agency
- EOS** - End Of Sequence (Fin de Secuencia)
- ETA** - Estimated Time of Arrival
- FAM** - Fuerza de Acción Marítima
- GAN** - Generative Adversarial Network
- GPS** - Global Positioning System
- GT** - Gross Tonnage (Arqueo bruto)
- H3** - Hexagonal Hierarchical Spatial Index
- HSC** - High Speed Craft (Buque de Alta Velocidad)
- IA** - Inteligencia Artificial
- IBM** - International Business Machines
- IMO / OMI** - International Maritime Organization / Organización Marítima Internacional
- ITU** - International Telecommunication Union
- KNN** - K-Nearest Neighbors
- LAT** - Latitude (Latitud)
- LSTM** - Long Short-Term Memory
- LON** - Longitude (Longitud)
- MAE** - Mean Absolute Error (error absoluto medio)

- ML** - Machine Learning
- MMSI** - Maritime Mobile Service Identity
- OTAN** - Organización del Tratado del Atlántico Norte
- PCA** - Principal Component Analysis
- RMP** - Recognized Maritime Picture
- RNN** - Recurrent Neural Network
- ROT** - Rate of Turn
- SOLAS** - Safety of Life at Sea
- TDMA** - Time Division Multiple Access
- TEU** - Twenty-foot Equivalent Unit
- TFG** - Trabajo Fin de Grado
- UBER** - Plataforma digital (referida al sistema de cuadrículas espaciales)
- UE** - Unión Europea
- UTC** - Coordinated Universal Time
- UNCTAD** - United Nations Conference on Trade and Development
- UTC** - Coordinated Universal Time
- VHF** - Very High Frequency
- WIG** - Wing In Ground
- ZEE** - Zona Económica Exclusiva

ANEXO IV: CÓDIGO NUMÉRICO DE LA TIPIFICACIÓN DE BUQUES EN LOS DATOS AIS

Código numérico	Clase de buque	Código numérico	Clase de buque
20-29	Wing in ground	52	Tug
30	<i>Fishing</i>	53	Port Tender
31	Towing	54	Anti-pollution equipment
32	Towing: length exceeds 200m or breadth exceeds 25m	55	Law Enforcement
33	Dredging or underwater ops	56	Spare - Local Vessel
34	Diving ops	57	Spare - Local Vessel
35	Military ops	58	Medical Transport
36	Sailing	59	Noncombatant ship according to RR Resolution No. 18
37	Pleasure Craft	60-69	Passenger
40-49	High speed craft	70-79	<i>Cargo</i>
50	Pilot Vessel	80-89	Tanker
51	Search and Rescue vessel	90-99	Other type

Tabla 0-1 Código numérico según tipo de buque en los mensajes AIS [53]

ANEXO V: ESTRUCTURA DE UN MENSAJE AIS

32 columnas:

mmsi,lat,lon,accuracy,receiver_timestamp,device,msg_type,eta,course,speed,heading,alt,turn,status,ais_version,ship_type,name,callsign,shipname,destination,imo,draught,to_bow,to_stern,to_port,to_starboard,year,month,day,hour,minute,second

CAMPO	DESCRIPCIÓN	EJEMPLO
MMSI	El identificador del Servicio Móvil Marítimo consiste en un código numérico de nueve cifras que permite distinguir de forma única a cada estación encargada de la transmisión.	357444000
LATITUDE	Latitud	42.23282999999999
LATITUDE	Latitud con rango de -90 a +90	42.23282999999999
LONGITUDE	Longitud	-8.722649999999999
ACCURACY	Indicador de precisión de la posición AIS: 0 : low accuracy (+-10m) 1 : high accuracy (<=10m)	1
TIMESTAMP	Corresponde al instante en el que el sistema receptor recibe el mensaje AIS, expresado conforme al formato temporal definido por el propio sistema AIS. Date-time format (e.g., YYYY-MM-DD HH:MM:SS UTC+0100)	2024-11-22 00:23:33 UTC +0100
DEVICE	Tipo de transmisor AIS	Class A, Base Station, SAR Aircraft, AtoN, Class B
MSG TYPE	Tipo de mensaje AIS	18
ETA	Estimated Time of Arrival(Hora estimado de llegada) Formato en MM-DD HH:mm	07-21 14:00

COURSE	COG(Course over ground) Rumbo efectivo del buque	170.2
SPEED	SOG(Speed over ground) Velocidad efectiva del buque en nudos	72
HEADING	Rumbo de verdadero de un buque	173
ALT	Altitud en metros	0
TURN	ROT (Rate of Turn): Como de rápido cae de rumbo un barco	010
STATUS	Identificador que señala la situación o maniobra que el buque se encuentra realizando en ese momento: • 0 = Under way using engine (Navegando a motor) • 1 = At anchor (Fondeado) • 2 = Not under command (Buque sin gobierno) • 3 = Restricted manoeuvrability (Maniobra restringida) • 4 = Constrained by her draught (Restringido por su calado) • 5 = Moored (Atracado) • 6 = Aground (Varado) • 7 = Engaged in <i>Fishing</i> (Faenando) • 8 = Under way sailing (Navegando a vela) • 9 = Reserved for future amendment of Navigational Status for HSC (Reservado para futuras modificaciones del estado de navegación del HSC(High-Speed-Craft)) • 10 = Reserved for future amendment of Navigational Status for WIG (Reservado para futuras modificaciones del estado de navegación del WIG(Wing-In-Ground)) • 11 = Reserved for future use (Reservado para uso futuro)	1

	• 12 = Reserved for future use (Reservado para uso futuro) • 13 = Reserved for future use (Reservado para uso futuro) • 14 = AIS-SART is active (AIS-SART(Transmisor AIS de Búsqueda y Rescate) activado) • 15 = Undefined (per AIS standards) (Sin definir)	
AIS VERSION	Versión del AIS del transmisor	2
SHIP_TYPE	Tipo de buque (0-99)	70
NAME	Nombre AIS que transmite buque	MEDITERRANEAN
CALLSIGN	CALL-SIGN radio internacional del barco	9981M71
SHIPNAME	Nombre del buque tal como se transmite en los mensajes AIS de tipo 5, 19 y 24	MEDITERRANEAN
DESTINATION	Destino del buque	PUERTO DE BARCELONA
IMO	Código numérico compuesto por siete cifras, asignado de manera individual a cada buque, cuyo objetivo es permitir su identificación y contribuir a la seguridad de la navegación marítima.	9102710
DRAUGHT	Calado estático actual del buque en decímetros	200
TO_BOW	Distancia del dispositivo AIS hasta la proa en metros	22
TO_STERN	Distancia del dispositivo AIS hasta la popa en metros	10
TO_PORT	Distancia desde el dispositivo AIS a babor en metros	3
TO_STARBOARD	Distancia desde el dispositivo AIS a estribor en metros	3
YEAR	Año de transmisión	2024
MONTH	Mes de transmisión	02

DAY	Día de transmisión	17
HOURL	Hora de transmisión	20
MINUTE	Minuto de transmisión	22
SECOND	Segundo de transmisión	33
EMPTY VALUE	En caso de que no haya datos disponibles en uno de los valores.	EMPTY VALUE

Tabla 0-1 Estructura de un mensaje AIS

ANEXO VI: ACCESO AL CÓDIGO DESARROLLADO DE LOS MODELOS GENERATIVOS

CARGO:

Notebook generación con cGAN y cGAN+KNN:

[NOTEBOOK_GENERACION_CARGO_cGAN-KNN_Y_CGAN-.ipynb - Colab](#)

Notebook generación con cuadrículas H3:

[NOTEBOOK_GENERACION_CARGO_H3.ipynb - Colab](#)

FISHING:

Notebook generación con cuadrículas H3:

[NOTEBOOK_GENERACION_PESCA_H3.ipynb - Colab](#)

Notebook generación con LSTM:

[NOTEBOOK_GENERACION_PESCA_LSTM.ipynb - Colab](#)

COMPARACIÓN Y RESULTADOS:

Notebook de comparación y resultados:

[NOTEBOOK_COMPARACIÓN_RESULTADOS.ipynb - Colab](#)

ANEXO VII: DECLARACIÓN DE USO DE INTELIGENCIA ARTIFICIAL

Durante la realización del presente Trabajo Fin de Grado se han utilizado herramientas de inteligencia artificial como apoyo en distintas fases del desarrollo. Dichas herramientas han sido empleadas exclusivamente con fines auxiliares, manteniendo en todo momento la autoría, el criterio técnico y la responsabilidad sobre el contenido por parte del autor.

En el desarrollo del proyecto se han utilizado herramientas de asistencia basadas en modelos de lenguaje de gran tamaño (LLM), entre ellas *ChatGPT* [82], *Gemini* [83] y *NotebookLM* [84] como apoyo en tareas de programación y en la exploración de posibles soluciones. Estas herramientas se emplearon principalmente para la generación de ejemplos de código, la propuesta de estrategias de implementación y la elaboración de prototipos preliminares durante el desarrollo del sistema. En todos los casos, las propuestas obtenidas fueron posteriormente analizadas, adaptadas y validadas antes de su integración en la solución final.

Adicionalmente, se han empleado herramientas de inteligencia artificial para la mejora de la redacción y el estilo del documento, contribuyendo a optimizar la claridad, coherencia y corrección gramatical de determinados fragmentos, sin alterar en ningún caso el contenido técnico ni las conclusiones del trabajo. Asimismo, se han utilizado como apoyo en la generación de imágenes, con el objetivo de facilitar la comprensión de los conceptos expuestos. Estas representaciones han sido posteriormente revisadas y ajustadas conforme a las necesidades específicas del trabajo.

Se declara que el uso de estas herramientas se ha realizado de acuerdo con principios éticos y académicos, garantizando en todo momento la originalidad del trabajo y evitando cualquier forma de plagio. Todas las decisiones metodológicas, enfoque adoptado, análisis de los resultados y las conclusiones presentadas corresponden íntegramente al autor.