



ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

“DISEÑO DE UN ALGORITMO PREDICTIVO PARA EL ANÁLISIS DE
DISPONIBILIDAD DE CANALES Y EL USO EFICIENTE DE SISTEMAS
CON ACCESO OPORTUNISTA AL ESPECTRO BASADO EN EL
MODELO K-VECINOS MÁS CERCANOS”

INFORME DE MATERIA INTEGRADORA

Previo a la obtención del Título de:

**INGENIERO/A EN ELECTRÓNICA Y
TELECOMUNICACIONES**

ASTRID CAROLAY MEDINA VELASCO

MARCO ANDRÉS VILLALBA SOTOMAYOR

GUAYAQUIL – ECUADOR

AÑO: 2018

TRIBUNAL DE EVALUACIÓN

Ing. Washington Medina
PROFESOR DE MATERIA
INTEGRADORA

Ing. Miguel Molina
TUTOR ACADÉMICO

DECLARACIÓN EXPRESA

"La responsabilidad y la autoría del contenido de este Trabajo de Titulación, nos corresponde exclusivamente; y damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

Marco Andrés Villalba Sotomayor

Astrid Carolay Medina Velasco

RESUMEN

Para buscar formas de dar un uso eficiente del espectro radioeléctrico, como el desarrollo de dispositivos con acceso oportunista al espectro (OSA), con el fin de aprovechar los espacios libres encontrados en los canales sin causar interferencias a los usuarios licenciados, analizando en que tiempo y en que ancho de banda se encuentran disponibles para su uso, se realizó el presente proyecto que consiste en el diseño de un algoritmo predictivo basado en el modelo estadístico de aprendizaje K-vecinos más cercanos.

Este estudio se basó en un análisis en la banda wifi de 2.4GHz con datos reales tomados en un transcurso de 10 días, dentro de un edificio en tres diferentes pisos, 1, 9 y 16 respectivamente, ubicado en una zona urbana. Se convirtieron los datos de potencia a datos binarios, obteniendo los canales disponibles y no disponibles para así realizar la agrupación de n canales conjuntos.

El método K-vecinos más cercanos, considerado un método de aprendizaje que clasifica los datos recordando los datos anteriores llamados entrenamiento, cada vez que un nuevo dato se presenta al sistema, este lo clasifica de acuerdo al comportamiento del dato más cercano. La cercanía de los vecinos se define en base a los atributos del dato y de los de entrenamiento, el algoritmo calcula la distancia entre los atributos.

Los resultados del estudio se obtuvieron por medio de simulaciones en el software utilizado, el cual permitió validar el algoritmo de predicción diseñado y con el cual se realizó la comparación con los datos reales para determinar el porcentaje de error, obteniendo un valor bajo. Se representaron mediante gráficas el comportamiento en cada uno de los pisos para tres días diferentes respecto a los n canales conjuntos y la posibilidad de encontrar n canales conjuntos disponibles en determinado día, realizando también la comparación entre datos originales y datos predichos.

ÍNDICE GENERAL

RESUMEN	I
ÍNDICE GENERAL	II
ABREVIATURAS.....	IV
SIMBOLOGÍA.....	V
ÍNDICE DE FIGURAS	VI
ÍNDICE DE TABLAS	VIII
CAPÍTULO 1	1
1 DESCRIPCIÓN DEL PROYECTO	1
1.1 OBJETIVOS	1
1.1.1 Objetivo General.....	1
1.1.2 Objetivos Específicos	1
1.2 PROBLEMA A RESOLVER	2
1.3 JUSTIFICACIÓN	3
1.4 ESTADO DEL ARTE	5
CAPÍTULO 2	9
2 MARCO TEÓRICO	9
2.1 Espectro Radioeléctrico	9
2.2 Red de Área Local Inalámbrica (WLAN)	9
2.2.1 Estándares WLAN	10
2.2.2 Estándar IEEE 802.11	10
2.2.3 Estándar IEEE 802.11a	10
2.2.4 Estándar IEEE 802.11b	10
2.2.5 Estándar IEEE 802.11g	11
2.2.6 Estándar 802.11n	11
2.3 Acceso Dinámico al Espectro.....	11
2.3.1 Clasificación del Acceso Dinámico al Espectro	12
2.3.1.1 Modelo Dinámico de Uso Exclusivo	12
2.3.1.2 Modelo de Compartición Abierta	13
2.3.1.3 Modelo de Acceso Jerárquico	13
2.4 Acceso Oportunista al Espectro	13
2.4.1 Espacios en Blanco	14

2.4.2	Componentes Básico del Acceso Oportunista al Espectro	14
2.4.2.1	Identificación de Oportunidad del Espectro	15
2.4.2.2	Explotación de Oportunidad del Espectro	16
2.4.2.3	Política Regulatoria	16
2.5	Radio Cognitiva	16
2.5.1	Arquitectura de la Radio Cognitiva	17
2.5.2	Ciclo Cognitivo	18
2.5.2.1	Detección del espectro	18
2.5.2.2	Asignación del espectro	19
2.5.2.3	Movilidad del espectro	19
2.5.2.4	Compartir el espectro	19
2.6	Modelo K vecinos más cercanos (KNN)	19
2.6.1	Definición	19
2.6.2	Reglas de KNN	20
2.6.3	Medidas de Proximidad	20
2.6.3.1	Medidas de proximidad para datos homogéneos	22
CAPÍTULO 3		24
3	ANÁLISIS DE DATOS Y ALGORITMO DE PREDICCIÓN	24
3.1	Descripción del entorno de estudio	24
3.2	Localización del lugar de mediciones	25
3.3	Procesamiento de Datos	26
3.4	Modelamiento del Algoritmo	29
3.5	Calculando Distancia	30
3.6	Datos de entrenamiento	33
3.7	Algoritmo K-vecinos más cercanos en R	34
CAPÍTULO 4		36
4	RESULTADOS	36
CONCLUSIONES Y RECOMENDACIONES		57
Conclusiones		57
Recomendaciones		58
REFERENCIAS		59
ANEXOS		62

ABREVIATURAS

OSA Acceso Oportunista al Espectro

ARCOTEL Agencia de Regulación y Control de las Telecomunicaciones

FCC Federal Communications Commission

UIT-R Sector de Radiocomunicaciones de la Unión Internacional de Telecomunicaciones

SNR Relación Señal a Ruido

RN Redes Neuronales

SVM Máquinas de Soporte de Vectores

KNN K-vecinos más cercanos

WLAN Red de área local inalámbrica

IEEE Instituto de Ingeniería Eléctrica y Electrónica

ISM Industrial Scientific and Medical

ETSI Instituto Europeo de Normas de Telecomunicaciones

PSK Modulación por desplazamiento de fase

OFDM Multiplexación por división de frecuencia ortogonal

DSSS Espectro ensanchado por secuencia directa

CCK Codificación de código complementario

MIMO Múltiples entradas Múltiples salidas

DSA Acceso Dinámico al Espectro

UP usuarios primarios

US usuarios secundarios

SDR Radio definida por software

QoS Calidad de servicio

AP Access Point

CDF Función de distribución acumulada

SIMBOLOGÍA

MHz Megahercio

KHz Kilohercio

GHz Gigahercio

Mbps Megabit por segundo

dbm decibelio-milivatio

m metro

ÍNDICE DE FIGURAS

Figura 2.1 Clasificación del Acceso Dinámico al Espectro [15]	12
Figura 2.2 Identificación de los Espacios en Blanco [15]	14
Figura 2.3 Ilustración de la Oportunidad del Espectro [15].....	15
Figura 2.4 Arquitectura de la Radio Cognitiva [20].	17
Figura 2.5 Ciclo Cognitivo [20].	18
Figura 2.6 Clasificación de las técnicas de detección del espectro [21].....	18
Figura 3.1 Zona Urbana	24
Figura 3.2 WLAN en ambientes interiores.....	25
Figura 3.3 Torre A Edificio World Trade Center	26
Figura 3.4 Grafica de error en predicción del día 2 según K-vecinos.....	30
Figura 3.5 Gráfico de Error vs Datos de entrenamiento.	34
Figura 4.1 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 4	38
Figura 4.2 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 7	39
Figura 4.3 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 10	40
Figura 4.4 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 4	42
Figura 4.5 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 7.	43
Figura 4.6 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 10.	44
Figura 4.7 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 4.	46
Figura 4.8 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 4.	47
Figura 4.9 Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 4.	48

Figura 4.10 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 7.....	50
Figura 4.11 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 7.....	51
Figura 4.12 . Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 7.....	52
Figura 4.13 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 10.....	54
Figura 4.14 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 10.....	55
Figura 4.15 Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 10.....	56

ÍNDICE DE TABLAS

Tabla 2.1 Características de la Radio Cognitiva [18].....	17
Tabla 3.1 Tabla de frecuencia. [24]	31
Tabla 3.2 Matriz de datos binarios. [24]	32
Tabla 3.3 Tabla de frecuencia de la matriz binaria. [24].....	32
Tabla 3.4 Matriz de distancias. [24].....	33

CAPÍTULO 1

1 DESCRIPCIÓN DEL PROYECTO

1.1 OBJETIVOS

1.1.1 Objetivo General

Diseñar un algoritmo predictivo mediante el modelo k-vecinos más cercanos para canales en la banda WIFI de 2.4 GHz dentro de una edificación ubicado en una zona urbana para determinar el uso eficiente de sistemas y tecnologías con acceso oportunista al espectro (OSA).

1.1.2 Objetivos Específicos

- Encontrar la disponibilidad de los canales calculando la potencia de threshold.
- Identificar el modelo de predicción correspondiente en base a los datos proporcionados.
- Diseñar un algoritmo predictivo en base al modelo escogido para determinar la disponibilidad de los canales y su uso óptimo dentro del espectro, evitando así congestión en el mismo.
- Realizar pruebas y análisis de los resultados comparados con los datos reales.
- Validar el algoritmo predictivo mediante el uso de herramientas de simulación.

1.2 PROBLEMA A RESOLVER

El espectro radioeléctrico es un recurso que brinda varios servicios de telecomunicaciones, tales como realizar una llamada, sintonizar una emisora, ver la televisión, etc. Su uso y explotación durante el transcurso de los años ha sido objeto de una demanda creciente y casi exponencial, al punto de congestionarse y a que las bandas de frecuencia lleguen a saturarse.

Actualmente el rápido crecimiento y aumento de nuevas tecnologías e innovaciones que requieren transferencia de señales de datos, voz, audio y video en bandas de frecuencia específicas, han vuelto al espectro radioeléctrico un recurso esencial para proveer estos servicios tanto en entidades públicas como privadas, por lo que se ha creado una gran demanda del mismo, considerándolo así un recurso natural limitado. A pesar de que existen entidades regulatorias del espectro radioeléctrico, se ha tenido un uso ineficiente el cual conlleva a tener problemas de interferencias y deterioro de señal, obteniendo así una mala calidad del servicio.

La continua implementación de redes inalámbricas ha causado saturación y un uso excesivo de las bandas de frecuencias, una parte muy significativa de las bandas de frecuencias licenciadas tienen muy bajo tráfico; es decir, en ciertos momentos una parte del espectro no está siendo utilizado. Esto ha ocasionado que otros segmentos de espectro estén siendo sobreexplotados y por tanto congestionados [1]. Por esta razón usuarios no autorizados o sin licencia toman iniciativa al utilizar bandas ya licenciadas que tengan espacios disponibles, asumiendo que no provocarán interferencia. Por ejemplo, los radioaficionados no tenían el conocimiento de que las bandas de telefonía celular en su mayor parte están sobrecargadas, y por tanto estudios independientes concluyeron que la utilización del espectro depende fuertemente de factores como el tiempo, lugar, densidad poblacional y el nivel socioeconómico [2].

La Agencia de Regulación y Control de las Telecomunicaciones (ARCOTEL) que es la encargada de administrar, regular, y controlar las telecomunicaciones y la gestión del espectro radioeléctrico en el Ecuador, ha tomado acciones frente al problema del mal uso del espectro radioeléctrico, modificando y actualizando el Plan Nacional de Frecuencias y el Cuadro de Atribución de Bandas de Frecuencias.

Se encuentran tres factores importantes para obtener una eficiencia en el uso del espectro radioeléctrico que son: técnica, social y económica. La eficiencia técnica depende de la adecuada asignación de las bandas de frecuencias, cumpliendo las exigencias técnicas, que eviten o limiten cualquier tipo de interferencias perjudiciales entre servicios o estaciones que operan dentro de una misma banda de frecuencias o en otras bandas de frecuencias, la eficiencia social que busca promover el desarrollo social, a través de la prestación de servicios, que satisfacen necesidades de interés público y privado, y la eficiencia económica que debe garantizar un desarrollo económico mayor de acuerdo al uso que se le dé al espectro radioeléctrico, su disponibilidad es importante para el desarrollo de empleos, innovaciones, competitividad, etc. [3].

1.3 JUSTIFICACIÓN

Para aprovechar de manera eficiente el espectro, se han implementado diferentes arquitecturas de Acceso Dinámico al Espectro tales como el Acceso Oportunista al Espectro (OSA), que se basa en que usuarios no licenciados accedan y utilicen los llamados White Spaces (espacios en blanco) dentro del espectro de usuarios licenciados, siempre y cuando no causen interferencia con los mismos.

Una de las aplicaciones ya introducidas es la Radio Cognitiva, considerada “actualmente una de las mayores posibilidades existentes para hacer frente a la creciente escasez de espectro y está orientada a

aprovechar las frecuencias subutilizadas” [4], en donde un transceptor inteligente detecta de manera dinámica los canales ocupados y los disponibles, determinando si son reutilizables y no causen ningún tipo de interferencia en la transmisión de los operadores principales, siguiendo la Recomendación 802.22 [2].

El espectro compartido permite que se utilicen los espacios en blanco para diferentes aplicaciones o tecnologías, la FCC (Federal Communications Commission) ha regulado su uso para servicios de internet inalámbrico en banda ancha ya que podría llegar a varias decenas de kilómetros, teniendo un gran impacto social y económico, ya que permite que surjan nuevas industrias, proyectos y se potencie la inversión e innovación.

Estudios realizados por la FCC han permitido concluir tres cosas importantes:

- Sólo un porcentaje pequeño del espectro total disponible es utilizado para la operación de los sistemas de radiocomunicaciones.
- La congestión y escasez del espectro se debe a la administración y regulación, más no a una verdadera insuficiencia física del recurso.
- Es necesario diseñar e implementar tanto políticas como tecnologías que optimicen el uso general que se hace del recurso espectral [5].

Con el fin de que la disponibilidad del espectro crezca se han realizados métodos de fijación de precios, tales como subastas y mercado secundario, ya que así habría una liberación del espectro el cual puede ser usado de una manera eficiente. Las subastas generan ingresos que se pueden invertir en ámbitos sociales, desarrollo de infraestructuras y crecimiento de servicios; y el comercio secundario permite la

transferencia o intercambio de derechos de uso del espectro entre agentes proveedores de servicio [3].

En el Ecuador de acuerdo a la Ley Orgánica de Telecomunicaciones se ha generado un impacto con las regulaciones que se han dado por parte de ARCOTEL en los ámbitos técnicos, económicos y de acceso a insumos de infraestructura. Ya se cuenta con una regulación actualizada con acceso a servicios de calidad, fomentando el uso efectivo y eficiente del espectro, elaboración de bases de concurso público para la asignación adecuada de títulos habilitantes de uso y explotación del espectro, así como también la creación y actualización de canalizaciones de bandas de frecuencia [6].

Ya se han realizado estudios estadísticos y probabilísticos por problemas de emisiones no deseadas, afectando la eficiencia del espectro. La UIT-R realizó una modelación estadística de escenarios de interferencia por el Método Monte Carlo, utilizado para la simulación de procesos aleatorios, basándose en el principio de tomar muestras de variables aleatorias de la función densidad de probabilidad [7]. Dado esto, ya se ha dado rienda suelta para realizar análisis y diseños de algoritmos predictivos adaptados a un modelo estadístico, con el fin de determinar varios factores que se encuentran presentes dentro del espectro radioeléctrico y dar soluciones a una mejor gestión del mismo, obteniendo su eficiente uso.

1.4 ESTADO DEL ARTE

Artículos Relacionados:

[8]. Este artículo se enfoca en el diseño de un modelo dinámico para tomar decisiones en redes inalámbricas cognitivas.

El modelo permite a usuarios secundarios tener la oportunidad de usar el espectro y sus canales sin afectar el tráfico de los usuarios primarios. Este modelo incluye un algoritmo inicial de toma de decisiones con múltiples criterios que clasifican a los mejores canales acuerdo a la caracterización de su espectro y un algoritmo final para predecir el espectro que será ocupado por un canal primario y así el usuario secundario pueda cambiar de canal a tiempo.

El modelo fue entrenado usando tres técnicas: Redes Neuronales(RN), Maquinas de Soporte de Vectores(SVM) y K-vecinos más cercanos(KNN).

El proceso de predicción comienza con las estadísticas tomadas de las medidas con un alto y bajo tráfico. Esta predicción permitirá al usuario secundario reconocer los instantes de tiempo en los que el usuario primario puede ir a un canal, y así cambiarse a un segundo mejor canal y no interferir.

Para la técnica de KNN, el procesamiento de datos es realizado obteniendo datos normalizados entre 0 y 1, donde el 0 indica que un canal se encuentra libre y 1 indicando que se encuentra ocupado. El algoritmo obtiene un vector final para cada canal y con el mismo tamaño de los datos de entrada.

Los resultados fueron evaluados en dos escenarios, uno con alto tráfico y otro con bajo tráfico según un algoritmo que estimaba el total de handoffs.

Obteniendo para el escenario de bajo tráfico un excelente y notorio rendimiento y fue el más conveniente por el cambio de canales anticipados, con interferencia, perfectos y fallidos, mientras que para un alto tráfico tuvo un bajo rendimiento en comparación de los otros modelos.

[9]. Este artículo contiene un análisis de los algoritmos más usados y de técnicas de posicionamiento que podrían ser aplicadas en ambientes indoor y adaptadas a dispositivos móviles de última generación.

Nos plantea los pros y contras de cada algoritmo utilizado, tratando de que sean aplicables a redes móviles en los interiores de edificios.

Nuestro interés va a estar enfocado en el método K-nearest-neighbor, este artículo emplea las medidas de las distintas potencias de la señal recibida llevadas a cabo online para realizar una búsqueda y escoger las k equivalencias más cercanas de localizaciones conocidas, las cuales han sido previamente almacenadas en bases de datos.

Existen dos fases en el proceso de este estudio sobre KNN: en la primera fase el valor obtenido se compara con todos los valores ya almacenados previamente, y se selecciona los valores de k más cercanos. Tomando en cuenta las distancias entre las estaciones emisoras y usando como base el principio de la raíz del error cuadrático medio, se establece el criterio de proximidad para este método.

La segunda fase vendría dada por la determinación de la posición, la cual se obtiene cuando todos los valores de posición obtenidos en la primera fase del proceso son promediados.

El inconveniente que se presentó en este método es que existe el caso de poder encontrar un conjunto de datos equivalentes que estén un poco concentrados y un valor con alta coincidencia, pero disperso a nivel de posición física. Para este caso hay que realizar ponderaciones adicionales para descartar valores fuera de las áreas de concentración. Otro problema a considerar en las técnicas de posicionamiento con los métodos deterministas KNN, es que no considera los posibles fallos que se podrían dar durante la determinación de la posición, es por eso por lo que en este artículo se trata de investigar cómo mejorar la tolerancia a las posibles falencias que presente el sistema KNN ante el mal

funcionamiento de las estaciones emisoras o de ataques maliciosos que quieran corromper los valores de potencia ya medidos durante el proceso de posicionamiento.

En 2005 T. N. Lin [10] llevo a cabo por medio de la tecnología WLAN una comparación entre algoritmos en los que a pesar de una elevada densidad de tráfico y numerosos obstáculos como muros obtuvo resultados muy buenos. El algoritmo que más se destacó en aquel momento fue el método KNN. Se consiguió una exactitud de 1 m con el método KNN utilizando la comparación sobre 10 muestras. Además, el método acertó el 80% de las veces en posicionamientos menores a 1 m.

CAPÍTULO 2

2 MARCO TEÓRICO

2.1 Espectro Radioeléctrico

El Espectro Radioeléctrico es el conjunto de ondas electromagnéticas agrupadas en bandas de frecuencia desde los 10KHz a los 3000GHz, usadas para prestar servicios de comunicación, los cuales son asignados por instituciones gubernamentales de cada país para su correcta administración y regulación [11].

Considerado un recurso natural limitado, con una creciente demanda en los últimos años tanto en entidades públicas como privadas, debido a nuevas tecnologías y servicios de comunicación móvil, televisión digital, etc.

2.2 Red de Área Local Inalámbrica (WLAN)

Las Redes de Área Local Inalámbricas es un sistema de comunicación flexible, estructurada para transmitir y recibir datos utilizando ondas electromagnéticas permitiendo mayor movilidad a los usuarios situados en una misma zona de cobertura (edificios, campus, hogares, etc.) con conexiones inalámbricas punto a punto, minimizando de esta manera las conexiones por cableado.

Las WLANs están regidas por los estándares establecidos por el IEEE para los distintos tipos de WLAN utilizados actualmente. Estos sistemas trabajan principalmente en las bandas ISM (Industrial Scientific and Medical), que son bandas de frecuencia libres, es decir no se requiere licencia para su uso, por lo que no están libres de interferencias [12].

2.2.1 Estándares WLAN

En la actualidad existen organizaciones internacionales tales como las ITU, IEEE, FCC, ETSI que han desarrollado la normativa y estandarización de WLAN.

2.2.2 Estándar IEEE 802.11

El estándar original desarrollado en el año 1997, trabaja en el rango de operación de 2.4 GHz, soportando velocidades de transmisión de 1 y 2 Mbps utilizando modulación por desplazamiento de fase (PSK) [13].

2.2.3 Estándar IEEE 802.11a

Siendo una extensión del estándar original, también llamado wifi5 fue publicado en 1999. Opera en el rango de 5 GHz que ofrece menor rango de cobertura, pero recibe menos interferencias. Soporta velocidades de transmisión en el rango de 6 Mbps, 12 Mbps o 24 Mbps; al utilizar multiplexación por división de frecuencia ortogonal (OFDM) la velocidad puede ser tan alta llegando hasta los 54 Mbps [13].

2.2.4 Estándar IEEE 802.11b

También conocido como Wi-Fi se desarrolló para conseguir una tasa de transmisión mayor en la banda de 2.4 GHz; trabajar en este rango de frecuencia es un problema, ya que lo hace vulnerable a interferencias debido a los otros estándares que operan en esta misma banda [13].

Soporta velocidades de transmisión de hasta 11 Mbps, siendo compatible con el estándar 802.11. Utiliza modulación DSSS por codificación de código complementario (CCK), que permite velocidades de datos más altas con menos posibilidades de interferencia de propagación multitrayectoria (las señales duplicadas rebotan en las paredes) [13] [14].

Este estándar tuvo una rápida aceptación el cual lo llevó a ser el estándar más extendido en el mercado, siendo este la tecnología WLAN definitiva.

2.2.5 Estándar IEEE 802.11g

Este estándar fue publicado en el 2013 ofreciendo las mejores características de los estándares 802.11a y 802.11b, opera en el rango de 2.4 GHz con velocidades de datos de hasta 54 Mbps en una distancia limitada [14].

2.2.6 Estándar 802.11n

El desarrollo de este estándar surgió para proporcionar un mejor rendimiento y ser capaz de obtener velocidades de rápido crecimiento. Los datos pueden alcanzar una velocidad de 600 Mbps, en las bandas de frecuencia tanto de 2.4 GHz como de 5 GHz. Utiliza modulación CCK, DSSS u OFDM, se introdujo MIMO para mejorar el desempeño del sistema, permitiendo manejar más información, el ancho de canal es más amplio entre 20 o 40 MHz [14].

2.3 Acceso Dinámico al Espectro

Debido al incremento y desarrollo de nuevas tecnologías en los últimos años, se ha requerido crear alternativas para abastecer y atender la demanda de los servicios de comunicación

El Acceso Dinámico al Espectro (DSA) con el fin de realizar un uso eficiente del espectro, permite que los usuarios accedan de manera dinámica al espectro, los usuarios no licenciados o también llamados usuarios secundarios puedan acceder a las bandas de frecuencia de usuario licenciados o llamados usuarios primarios que no estén siendo utilizadas.

2.3.1 Clasificación del Acceso Dinámico al Espectro

El Acceso Dinámico al Espectro (DSA) se ha categorizado en tres modelos estratégicos como se muestra en la Figura 2.1.

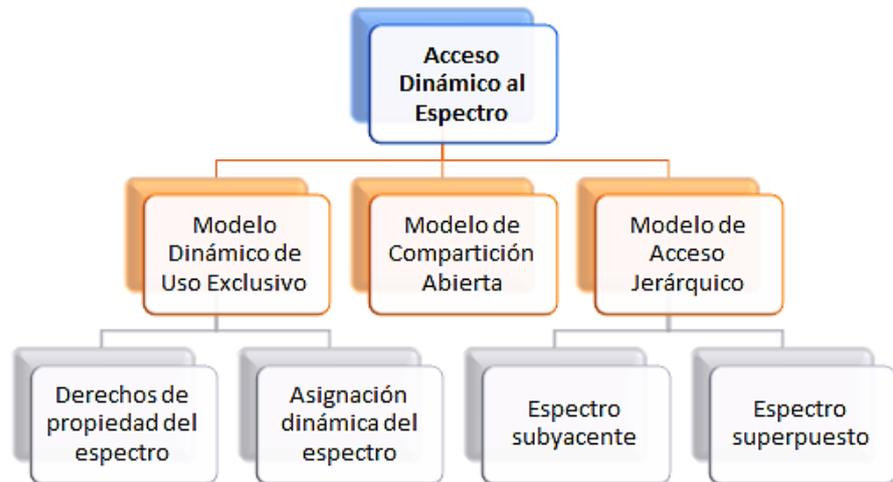


Figura 2.1 Clasificación del Acceso Dinámico al Espectro [15]

2.3.1.1 Modelo Dinámico de Uso Exclusivo

El Modelo Dinámico de Uso Exclusivo tiene como objetivo principal mejorar la eficiencia del espectro mediante medidas flexibles. Se basa en las políticas actuales de gestión y regulación del espectro, las bandas asignadas a servicios son otorgadas bajo licencia. En este modelo encontramos dos enfoques: derechos de propiedad del espectro, y asignación dinámica del espectro. El primer enfoque autoriza a los usuarios licenciados a comercializar (vender, arrendar o compartir) el espectro y elegir con que tecnología trabajar. El segundo enfoque permite asignar dinámicamente el espectro basándose en estadísticas del tráfico espacial y temporal de diferentes servicios [15].

2.3.1.2 Modelo de Compartición Abierta

El Modelo de Compartición Abierta o también llamada Modelo de Espectro Común, gestiona el espectro mediante el intercambio abierto entre usuarios iguales; el intercambio puede ser centralizado o distribuido [15].

2.3.1.3 Modelo de Acceso Jerárquico

El Modelo de Acceso Jerárquico da cabida en el espectro licenciado a los usuarios secundarios, limitando y controlando la interferencia que puedan percibir los usuarios primarios. Para compartir el espectro se han considerado dos enfoques: espectro subyacente y espectro superpuesto. El primer enfoque aplica restricciones a la potencia de transmisión de los usuarios secundarios, deben operar bajo el piso de ruido de los usuarios primarios, la velocidad de datos transmitida es alta y de corto alcance. El segundo enfoque no se aplican restricciones a la potencia de transmisión de los usuarios secundarios, pero si sobre dónde y cuándo pueden transmitir [15].

2.4 Acceso Oportunista al Espectro

El Acceso Oportunista al Espectro (OSA) basado en la arquitectura del acceso dinámico al espectro, busca minimizar los problemas de escasez y uso ineficiente del espectro radioeléctrico, debido a la mala gestión y administración de tal recurso. Siguiendo un enfoque jerárquico de usuarios secundarios (US) y usuarios primarios (UP), se proporcionan beneficios a los US y al mismo tiempo protege de interferencias a los UP; el espectro licenciado está siendo subutilizado por los US significativamente en los dominios de tiempo, espacio y frecuencia de los llamados espacios en blanco [16] [17].

2.4.1 Espacios en Blanco

Los Espacios en Blanco (White Spaces) son porciones de espectro radioeléctrico licenciado o no licenciado, que no están siendo utilizados en un determinado tiempo, el cual lo hace libre para su acceso.

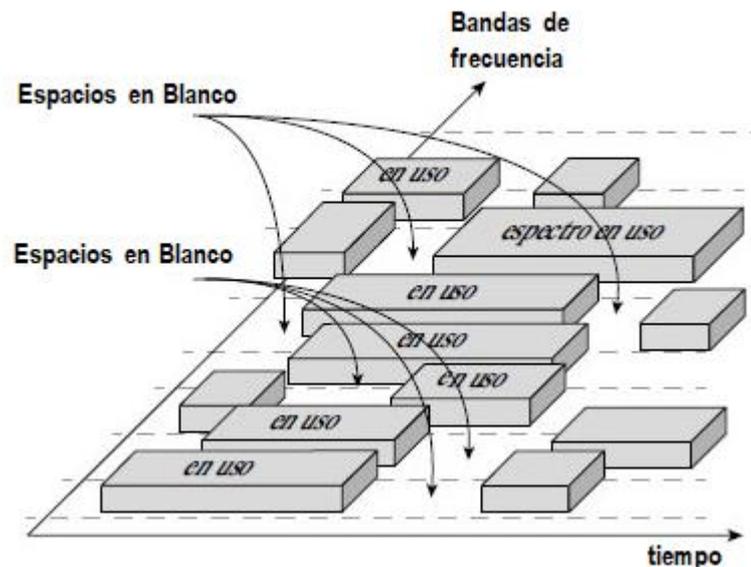


Figura 2.2 Identificación de los Espacios en Blanco [15]

2.4.2 Componentes Básico del Acceso Oportunista al Espectro

Se incluyen tres componentes básicos del OSA: identificación de oportunidad del espectro, explotación de oportunidad del espectro y política regulatoria.

Antes de explicar cada uno, se debe tener claro lo que es una oportunidad del espectro. Un canal se considera como una oportunidad si los usuarios primarios no lo están usando en un determinado momento.

La figura 2.3 ilustra cuando se tiene una oportunidad del espectro; el usuario secundario A desea transmitir al usuario secundario B, un canal es una oportunidad para ellos siempre y cuando no se encuentren usuarios primarios recibiendo la señal dentro de una distancia de r_{tx} de A y no se encuentren usuarios primarios transmitiendo la señal a una distancia de r_{rx} de B. En pocas palabras los usuarios secundarios A y B deben vigilar a los transmisores y receptores primarios para no causar interferencias [15].

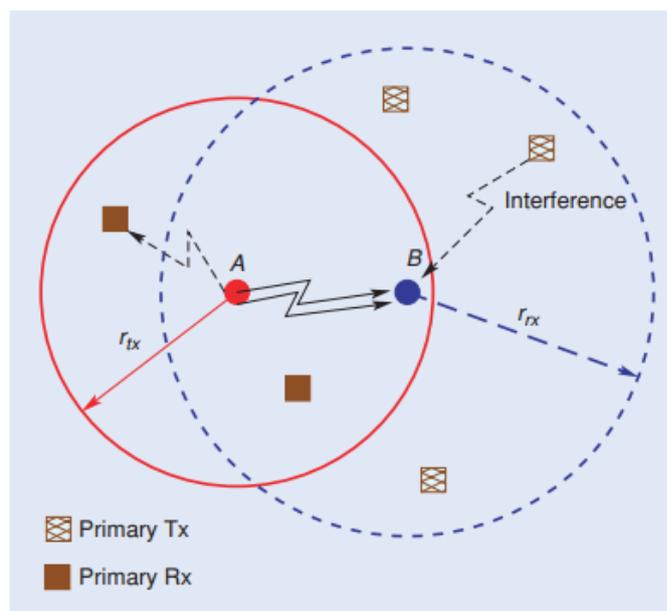


Figura 2.3 Ilustración de la Oportunidad del Espectro [15]

2.4.2.1 Identificación de Oportunidad del Espectro

Durante este proceso se identifica y rastrea las bandas de frecuencias que se encuentran inactivas y son dinámicas en tiempo y espacio [15].

2.4.2.2 Explotación de Oportunidad del Espectro

En este momento los usuarios secundarios deben decidir si explotar o no las oportunidades de espectro que se identificaron. Los eventos que se manifiestan incluyen si se debe transmitir (debido a eventuales errores en la detección de oportunidades), qué modulación usar, qué potencia de transmisión usar y cómo compartir oportunidades entre US para obtener resultados óptimos en la red [15].

2.4.2.3 Política Regulatoria

La política regulatoria atribuye un conjunto de reglas básicas que los US deben respetar para garantizar compatibilidad con los sistemas heredados [15].

2.5 Radio Cognitiva

La Radio Cognitiva se basa en la radio definida por software (SDR), es considerado una radio inteligente, proporcionando mecanismos de aprendizaje para reconocimiento y adaptación al medio que lo rodea (espectro radioeléctrico), con el fin de mejorar y optimizar el uso del espectro [2] [4].

Como forma de comunicación inalámbrica, la radio cognitiva permite detectar porciones de espectro que no están siendo utilizados determinando si son reutilizables para crear canales de comunicación, evitando los que se encuentran ocupados de manera que no haya interferencia en la transmisión de los usuarios licenciados [19].

De acuerdo a su funcionalidad, capacidad, configuración y entre otras características, la radio cognitiva ha tenido diferentes definiciones dadas por diferentes autores observadas en la Tabla 2.1 [18].

Autor	Adaptable	Autónomo entorno	Detectar el entorno	Transmisor	Receptor	"Consistente" del entorno	Metas	Aprendizaje del entorno	"Consistente" a capacidades	Negociar forma de onda	Sin Interferencia
FCC	✓	✓	✓	✓							
Haykin	✓	✓	✓	✓	✓	✓	✓	✓			
IEEE 1900.1	✓	✓	✓	✓	✓						
IEEE USA	✓	✓	✓	✓	✓	✓					✓
ITU-R	✓	✓	✓	✓	✓	✓					
Mitola	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
NTIA	✓	✓	✓	✓	✓	✓	✓				
SDRF CRWG	✓	✓	✓	✓	✓		✓				
VT CRWG	✓	✓	✓	✓	✓	✓	✓	✓	✓		

Tabla 2.1 Características de la Radio Cognitiva [18]

2.5.1 Arquitectura de la Radio Cognitiva

Para que se realice el intercambio de señales y datos que llevan a cabo el funcionamiento de la radio cognitiva, se necesitan componentes funcionales que forman lo que es la arquitectura de la radio cognitiva como se muestra en la Figura 2.4, cada una de esas funciones establecen varias interfaces en donde se realiza la transferencia de datos y señales de control que determinan el comportamiento del sistema. [20]



Figura 2.4 Arquitectura de la Radio Cognitiva [20].

2.5.2 Ciclo Cognitivo

El ciclo cognitivo representa una secuencia de estados que describen el proceso por etapas que realiza la radio cognitiva.

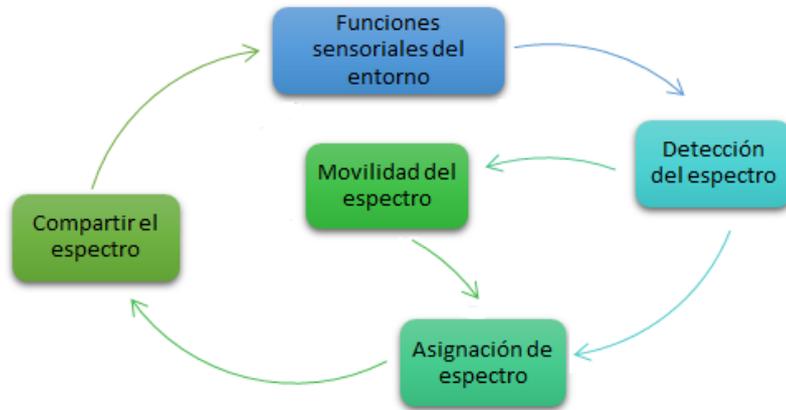


Figura 2.5 Ciclo Cognitivo [20].

2.5.2.1 Detección del espectro

Detectar qué fracción del espectro o espacios en blanco se encuentran en las bandas licenciadas de los usuarios principales, para que los usuarios secundarios sean asignados a estos sin causar interferencias; es por esto que los US deben siempre monitorear las bandas disponibles para obtener información y detectar dichos espacios en blanco. Existen técnicas de detección del espectro que se pueden clasificar en tres como se muestra en la Figura 2.6. [2] [21]



Figura 2.6 Clasificación de las técnicas de detección del espectro [21]

2.5.2.2 Asignación del espectro

Después de haber detectado los espacios en blanco, los US deben elegir la banda de espectro adecuada, teniendo en cuenta las características, sus ventajas y desventajas, así como también la calidad de servicio (QoS) y requerimientos del usuario para valorar cuál es su mejor opción. [2] [21]

2.5.2.3 Movilidad del espectro

Cuando los UP necesitan utilizar una porción específica del espectro, los US deben desocupar la banda que se encuentran utilizando y cambiar a otro espacio en blanco para continuar su comunicación. Esto debe realizarse de manera rápida e imperceptible para el usuario. [2] [21]

2.5.2.4 Compartir el espectro

Los US deben coordinar el acceso al espectro para poder compartirlo y así evitar colisiones con el sistema primario. [2] [21]

2.6 Modelo K vecinos más cercanos (KNN)

2.6.1 Definición

Para la clasificación de datos es muy importante conocer la naturaleza de los datos, es decir si es que estos son paramétricos o no paramétricos. Los datos paramétricos se refieren a los que tienen una distribución estadística durante los instantes de tiempo, saber qué tipo de datos son es esencial porque existen diferentes técnicas y algoritmos para los diferentes tipos de datos.

Los k vecinos más cercanos es una técnica de clasificación que asume la clase de un instante a ser la misma clase del instante más cercano. Este es uno de los métodos no paramétricos más simples para tomar decisiones. El termino no paramétrico se refiere al hecho de que no hay necesidad de conocer la distribución estadística de los datos, el método asume que los datos son independientes e idénticamente distribuidos y que los datos que están próximos a un instante de tiempo tendrán la misma clasificación.

K-vecinos más cercanos es un tipo de método de aprendizaje basado en instantes los cuales son llamados métodos de aprendizaje vagos. Este tipo de aprendizaje se limita a almacenar todos los datos de entrenamiento hasta el momento en que se decida realizar la predicción.

2.6.2 Reglas de KNN

La regla 1-NN asume que el valor del vecino inmediato será el mismo para la clase del nuevo instante. K puede ser cualquier número de sus vecinos, $K=1,2,3,4, n$, donde n es el número de casos.

La cercanía de los vecinos es definida en base a los atributos que definen el nuevo instante y los instantes de entrenamiento. Los instantes de entrenamiento los cuales sus valores de atributo son similares a los del nuevo instante se consideran los más cercanos, pero muchas veces el instante exacto similar no es encontrado, entonces el instante más cercano es el que sea más similar.

2.6.3 Medidas de Proximidad

El algoritmo del k- vecino más cercano calcula la distancia entre los atributos de un nuevo instante y de instantes previos para determinar la clase. El término “distancia” depende totalmente de los datos, diferentes tipos de datos tendrán diferentes caminos para encontrar su distancia.

Existen muchas maneras diferentes de categorizar los atributos de los datos:

- Atributos nominales son categóricamente datos discretos que consiste en categorizar nombres solamente y no hay distancia entre sus valores, solo difieren en ser similares o no. Atributos binarios son un caso especial de atributos nominales.
- Atributos ordinales tienen un ordenamiento natural de sus valores, pero no podemos definir la distancia entre ellos, por ejemplo, alto, medio y corto son tres atributos que definen el tamaño de una persona, pero su diferencia no puede ser dada por números, solo podemos aplicar mayor, menor, igual o diferente a ellos.
- Atributos de intervalo son los que tienen valores numéricos y la diferencia entre ellos es significativa. Por ejemplo, cuando se mide la temperatura en grados Celsius se pueden llevar a cabo operaciones como $>$, $<$, $=$, \neq , $+$, $-$.
- Atributos de relación son como los atributos de intervalo, pero la diferencia es que en los de relación el valor cero es un cero absoluto.

Otras formas de categorizar los atributos de los datos son entre cualitativos y cuantitativos:

- Los atributos cualitativos son atributos categóricos generalmente expresados como nombres de categorías. Pueden tener orden o no entre sus valores. Atributos nominales y ordinales son dos de sus tipos dependiendo del orden de sus valores.
- Los atributos cuantitativos son expresados como valores numéricos. Describen el valor como una medida cuantitativa, este valor puede ser exacto en términos numéricos, pero solo los atributos que pueden ser medidos son llamados atributos cuantitativos. Entre ellos tenemos a los atributos de intervalo y relación.

También existe otra forma de categorizar los datos:

Datos homogéneos donde todos sus datos son del mismo tipo y los datos heterogéneos donde sus atributos son de diferentes tipos, por ejemplo, un atributo es nominal mientras el otro es de relación.

2.6.3.1 Medidas de proximidad para datos homogéneos

Las medidas de distancia son definidas por medidas de proximidad. La medida de similitud significa que mientras más similares sean los casos, más cerca están los unos de los otros y una medida de no similitud indicara que están distantes.

Una métrica es una función de no similitud que cumple cuatro condiciones [22]:

$d(x, y) \geq 0$ (para cada x & y las distancias son números no negativos).

$d(x, x) = 0$ (la distancia de un objeto a sí mismo es 0) (también se llama reflexividad).

$d(x, y) = d(y, x)$ (simétrica).

$d(x, y) \leq d(x, z) + d(z, y)$ (desigualdad triangular ir directamente de x a y es más corto que hacer un desvío sobre el objeto z).

Para un buen sistema de aprendizaje, la función de distancia es importante. Aquí hay una variedad de funciones de distancia. Donde x & y son dos instancias de entrada y m es la cantidad de atributos [23].

- Función de distancia Euclidiana

$$d(x, y) = \sqrt{\sum_{a=1}^m (X_a - Y_a)^2} \quad (2.1)$$

- Función de distancia Manhattan

$$d(x, y) = \sum_{a=1}^m |X_a - Y_a| \quad (2.2)$$

- Función de distancia Minkowskian

$$d(x, y) = \sqrt[r]{\sum_{a=1}^m (X_a - Y_a)^r} \quad (2.3)$$

- Función de distancia Mahalanobis

$$d(x, y) = [\det V]^{1/m} (x - y)^T V^{-1} (x - y) \quad (2.4)$$

- Función de distancia Canberra

$$d(x, y) = \sum_{a=1}^m \frac{|X_a - Y_a|}{|X_a + Y_a|} \quad (2.5)$$

- Función de distancia Chebychev

$$d(x, y) = \max_{a=1}^m |x_a - y_a| \quad (2.6)$$

- Función de distancia de correlación

$$d(x, y) = \frac{\sum_{a=1}^m (X_a - \bar{X}_a)(Y_a - \bar{Y}_a)}{\sqrt{\sum_{a=1}^m (X_a - \bar{X}_a)^2 \sum_{a=1}^m (Y_a - \bar{Y}_a)^2}} \quad (2.7)$$

- Función de distancia chi-cuadrada

$$d(x, y) = \sum_{a=1}^m \frac{1}{\text{suma}_a} \left(\frac{x_a}{\text{tam}_x} - \frac{y_a}{\text{tam}_y} \right) \quad (2.8)$$

Donde suma es la suma de todos los valores para el atributo a ocurridos durante el entrenamiento y tam_x y tam_y son la suma de todos los valores en los casos x y y respectivamente.

CAPÍTULO 3

3 ANÁLISIS DE DATOS Y ALGORITMO DE PREDICCIÓN

En este capítulo se detallarán todos los procedimientos que se realizaron para la implementación del algoritmo predictivo y para el análisis de la disponibilidad de canales conjuntos en base a los datos de potencia obtenidos en el entorno de estudio.

En esta sección se dará a conocer una descripción del entorno donde los datos fueron obtenidos, además de especificar la manera en que estas mediciones fueron realizadas para posteriormente ser proporcionadas a nosotros para su respectivo análisis.

3.1 Descripción del entorno de estudio

Las mediciones fueron realizadas en la torre A del World Trade Center, un edificio que tiene 15 pisos y una terraza, se encuentra ubicado en el sector norte de la ciudad de Guayaquil, específicamente en un ambiente considerado como zona urbana con una gran cantidad de edificios.

La presencia de distintos tipos de interferencia en el entorno donde las mediciones fueron realizadas es algo que hay que considerar, existe interferencia tanto en el interior de los edificios, así como en el exterior, por ejemplo, la reflexión, refracción, dispersión, difracción y absorción que pueden ser generadas por paredes, ventanas y los distintos objetos de su alrededor.



Figura 3.1 Zona Urbana

Debido a las distintas áreas de oficinas que posee este tipo de edificios es común encontrar una gran variedad de APs localizados en un mismo piso para poder dar una buena cobertura al lugar, pero así mismo esto va ligado con la saturación de los canales por la cantidad de usuarios.

Esta saturación del espectro radioeléctrico es lo que obliga a buscar soluciones como el desarrollo de dispositivos con acceso oportunista al espectro (OSA) para tratar de buscar y usar esos espacios libres de espectro que hay en los canales.



Figura 3.2 WLAN en ambientes interiores

En este estudio se procederá a verificar la disponibilidad de los canales conjuntos en la banda de 2.4 GHz que consta de 14 canales, separados entre sí por 5 MHz. Cada uno de estos canales opera en un ancho de banda de 22 MHz. Las mediciones que nos fueron proporcionadas se realizaron mediante equipos que estuvieron instalados durante diez días en distintos pisos de la Torre A del Edificio World Trade Center.

3.2 Localización del lugar de mediciones

Las mediciones que fueron proporcionadas para el estudio tanto de disponibilidad de canales conjuntos como para la comprobación y validación del algoritmo predictivo se realizaron en la Torre A del edificio World Trade Center ubicado al norte de la ciudad de Guayaquil en un

sector considerado como zona urbana con una alta densidad de edificios, el edificio consta con 15 pisos y una terraza del cual se seleccionaron tres pisos para realizar dichas mediciones, estos fueron:

El piso 1, conformado por oficinas de ACE Seguros.

El piso 9, donde se encuentran las oficinas de la Notaria del cantón Guayaquil.

El piso 16, la Terraza de ACE Seguros.

Las mediciones se tomaron durante diez días consecutivos para cada uno de estos pisos.

La torre A del World Trade Center se encuentra ubicada a Latitud: $2^{\circ}9'49''S$ y Longitud: $79^{\circ}53'52''W$ de acuerdo a la figura 3.3.



Figura 3.3 Torre A Edificio World Trade Center

3.3 Procesamiento de Datos

Las mediciones que fueron asignadas se encontraban en libros de Excel, cada uno de estos libros contenían los datos de los 10 días de un solo canal.

El primer paso fue organizar estos datos, de manera que se separaron las mediciones desde el día 1 hasta el día 10, obteniendo 5760 mediciones por día en cada canal. Se observó que se obtenían 4 mediciones por minuto por ende se llegó a la decisión de manejar una sola medición por minuto para un mejor manejo posterior de los datos. Para esto mediante el modo programador de Excel se realizó un macro para sacar el promedio cada 4 mediciones y asignarla a su minuto correspondiente.

A partir de esto se obtuvo una matriz de potencias por cada canal donde las filas eran representadas por los minutos y las columnas por los 10 días, a la cual decidimos transformarla a la matriz de disponibilidad para solo manejar datos binarios, para esto se tuvo que fijar una potencia de umbral para ser comparada con cada minuto de la matriz de potencias y decidir si un canal se encontraba libre u ocupado en determinado tiempo.

Para escoger un umbral adecuado de acuerdo con las mediciones se realizaron gráficos que mostraban el comportamiento de la señal de potencia en el transcurso del día, de esta manera se pudo observar que el piso de ruido se encontraba por los -90 dbm, por lo tanto, para descartar que mínimos ruidos afecten a la decisión de disponibilidad se escogió un umbral de -85 dbm. Con esto todas las potencias que sean menores que -85 dbm serían considerados como un espacio disponible, mientras que las potencias mayores del umbral estarían consideradas como espacios ocupados.

La disponibilidad del canal en determinado tiempo fue representada por datos binarios, donde el 1 significaba que el canal se encontraba disponible y el 0 indicaba que se encontraba ocupado.

Una vez que realizamos el proceso de ver la disponibilidad de los canales de todos los días y de los 3 pisos, se procedió a organizar los

datos de tal manera que se obtuvieron libros de Excel de cada día con sus respectivos 14 canales.

Ya con dichos archivos se pudo ya orientarse a la parte del estudio de los canales conjuntos, que consiste en comprobar si existen de dos hasta catorce canales disponibles al mismo tiempo.

Nuevamente cuando ya se obtuvieron todos los archivos con la disponibilidad de los canales conjuntos, se procedió a ordenarlos de manera que se quedó con tres archivos, correspondientes al piso 1, piso 9 y piso 16 respectivamente.

Cada uno de estos archivos contenía hojas con las matrices de disponibilidad para 2 canales conjuntos hasta para los 14 canales, cada hoja tenía a su vez los diez días.

Se ordenó de tal manera estos archivos por dos motivos muy importantes, el primero de estos motivos fue para poder analizar la disponibilidad que existe para cada uno de los posibles grupos canales conjuntos y comprobar si existe disponibilidad por minuto de estos mismos grupos de canales conjuntos durante los diez días en los tres diferentes pisos.

Con estos resultados finales sobre la disponibilidad de los canales conjuntos de cada uno de los diez días con sus pisos respectivos se crearon unas tablas en Excel con las cuales se procedió a obtener de los datos, la frecuencia absoluta que es el número de veces que un valor está dentro de un conjunto, la frecuencia relativa que es la división de cada frecuencia absoluta con la suma de todas las frecuencias absolutas y la frecuencia acumulada que es la suma sucesiva de las frecuencias relativas, desde el menor al mayor de sus valores para generar las gráficas de función de distribución acumulada.

El segundo motivo también importante va más orientado al algoritmo de predicción, ya que de esta manera se podrá importar un archivo completo al programa de análisis y ordenado para así manejarlo de una manera comprensible.

3.4 Modelamiento del Algoritmo

El modelamiento de este algoritmo en el programa R parte del uso del método de los k vecinos más cercanos considerado como un algoritmo de aprendizaje supervisado, la forma en que la función de este algoritmo mostraba los resultados era de una forma reducida en comparación a la cantidad de resultados que se necesitaban para predecir una gran cantidad de datos al mismo tiempo. Mediante programación y variación de ciertos parámetros se logró construir una función que nos prediga la cantidad de datos que necesitemos manteniendo un error aceptable. [25]

K-vecinos más cercanos es un algoritmo simple que almacena todos los casos posibles y clasifica los nuevos casos por mayoría de votos de sus K vecinos.

Escoger el número de vecinos cercanos juega un rol importante para definir la eficacia del modelo. Este K determinara que tan bien son usados los datos que se le proporcionen para que se generen buenas predicciones. Generalmente un valor de K alto tiene beneficios como reducir la varianza que generan datos insignificantes, por otro lado, con K pequeños se suele ignorar patrones que pueden tener un gran significado al momento de predecir. [25] En este caso en específico debido a que se trabajó con los datos binarios en vez de potencias para predecir la disponibilidad de canales conjuntos directamente, se decidió predecir un día completo proporcionándole al algoritmo los datos del día anterior, pero variando el valor de K desde uno hasta diez, se enfocó en el error total de la predicción comparado con los datos originales y se graficó estos valores, obteniendo la siguiente gráfica:

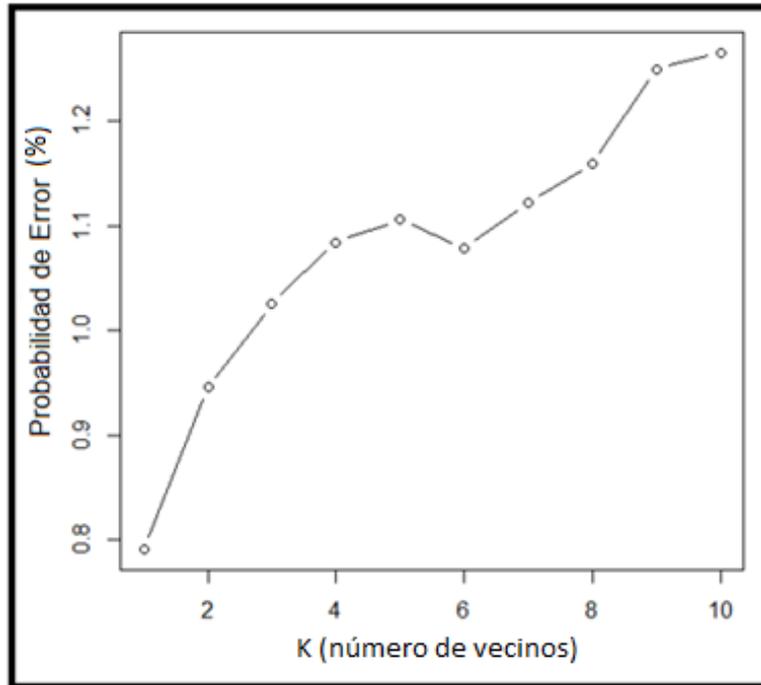


Figura 3.4 Grafica de error en predicción del día 2 según K-vecinos

Observando la figura 3.4 se puede observar que para $K=1$ se obtuvo un error de 0.8% aproximadamente en la predicción del día 2 y para $K = 10$ tenemos un error aproximado de 1.3%, concluyendo así que a medida que K va aumentando su error por lo general también aumenta a excepción de $K = 6$ aunque este sigue siendo un error mayor que $K=1$.

Este comportamiento de error vs K que se observa en la figura 3.4 se sigue repitiendo en las predicciones de los demás días, por lo que se decidió trabajar con $K = 1$ para la predicción de todos los días junto con sus respectivos canales conjuntos.

3.5 Calculando Distancia

El cálculo de la distancia entre un dato y sus vecinos más cercanos requieren el uso de una fórmula de distancia, la más popular es la distancia Euclidiana que además es la que usa nuestra función en el

programa R, esta distancia euclidiana vendría a ser la distancia más corta entre dos puntos la cual podría ser obtenida usando una regla. [24]

La distancia euclidiana para variables binarias tiene una versión un poco distinta a la estándar, se calcula con una tabla de frecuencia para cada par de datos a clasificar. Se tiene como ejemplo esta tabla:

S_i	$S_j \rightarrow$	1	0
1		X_{11}	X_{12}
0		X_{21}	X_{22}

Tabla 3.1 Tabla de frecuencia.
[24]

En la tabla 3.1 se puede describir a los X_{ij} como:

X_{11} : Frecuencia de acuerdos en que ambos datos tienen valor 1.

X_{22} : Frecuencia de acuerdos en que ambos datos tienen valor 0.

X_{12} : Frecuencia de desacuerdos en que el dato i tiene valor 1 y el dato j tiene valor 0.

X_{21} : Frecuencia de desacuerdos en que el dato i tiene valor 0 y el dato j tiene valor 1.

De acuerdo con la tabla de frecuencia se puede obtener la distancia euclidiana binaria con la siguiente ecuación:

$$d_{ij} = \sqrt{X_{12} + X_{21}} = \sqrt{\text{desacuerdos}} \quad (3.2)$$

Esta distancia que se obtiene es un índice de NO similitud entre los datos. Siendo así el valor mínimo un 0 cuando todos son acuerdos y el máximo valor cuando no hay acuerdos entre los datos comparados.

Se puede hacer semejanza los datos de estudio con el siguiente ejemplo que se muestra, donde las X_j vendrían a ser los grupos de canales conjuntos y las S_i los distintos tiempos.

	X_1	X_2	X_3	X_4	X_5
S_1	1	1	0	1	0
S_2	1	1	1	0	0
S_3	0	0	0	1	1
S_4	0	0	0	0	1
S_5	1	0	0	1	0

Tabla 3.2 Matriz de datos binarios. [24]

De aquí se procede a realizar la tabla de frecuencia para cada par posible cercano, generando lo siguiente:

		S_2		S_3		S_4		S_5	
		1	0	1	0	1	0	1	0
S_1	1	2	1	1	2	0	3	2	1
	0	1	1	1	1	1	1	0	2
S_2	1			0	3	0	3	1	2
	0			2	0	1	1	1	1
S_3	1					1	1	1	1
	0					0	3	1	2
S_4	1							0	1
	0							2	2

Tabla 3.3 Tabla de frecuencia de la matriz binaria. [24]

Una vez obtenida la tabla de frecuencia se usa la ecuación euclidiana para obtener la distancia entre los vecinos.

	S ₂	S ₃	S ₄	S ₅
S ₁	$\sqrt{2} = 1.4142$	$\sqrt{3} = 1.7321$	$\sqrt{4} = 2$	$\sqrt{1} = 1$
S ₂		$\sqrt{5} = 2.2361$	$\sqrt{4} = 2$	$\sqrt{3} = 1.7321$
S ₃			$\sqrt{1} = 1$	$\sqrt{2} = 1.4142$
S ₄				$\sqrt{3} = 1.7321$

Tabla 3.4 Matriz de distancias. [24]

De acuerdo con estas distancias es que el algoritmo toma en cuenta los K datos más cercanos para generar las predicciones. El algoritmo realiza esto con cada uno de los datos de entrenamiento para luego devolver una matriz del tamaño que se asigne con las predicciones correspondientes.

3.6 Datos de entrenamiento

Durante el proceso de elaboración de la función para el algoritmo predictivo también se tuvo que escoger la cantidad de datos de entrenamiento que se le iban a proporcionar para obtener una predicción con bajo porcentaje de error. Por lo tanto, de la misma forma que se hizo para obtener el mejor valor de K, se corrió el algoritmo con diferentes datos de entrenamiento desde 40 hasta 1940 datos y que prediga el día siguiente. Obteniendo el siguiente gráfico:

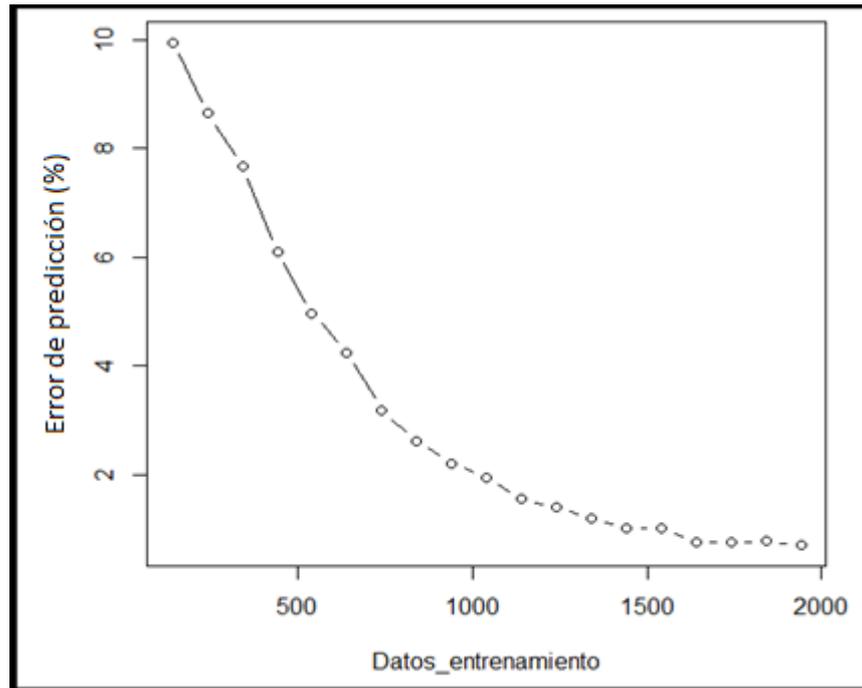


Figura 3.5 Gráfico de Error vs Datos de entrenamiento.

Del gráfico se observó que a medida que se le daban más datos de entrenamiento el error iba decreciendo desde el 10% hasta un poco menos del 1%. Pero por comodidad al momento de predecir de un día al siguiente día, se escogió 1440 como el número de datos de entrenamiento con un porcentaje promedio de error del 1% para realizar todas las predicciones de los canales conjuntos.

3.7 Algoritmo K-vecinos más cercanos en R

Para el uso de la base de este algoritmo es necesario llamar a la librería `class` en la consola de R con el comando:

`Library(class).`

También uno de los primeros pasos es establecer el directorio donde se encontrarán todos los archivos de canales conjuntos que ordenamos previamente con el siguiente comando:

```
setwd("C:/Users/Espol/Desktop/KNN")
```

De este comando se puede señalar algo muy importante, evitar el error de usar “\” en lugar de “/”.

Una vez establecido el directorio se puede importar cualquier archivo hacia un arreglo de datos con el comando read y su respectiva librería xlsx.

```
Data <- read.xlsx("Canales_Conjuntos.xlsx",1)
```

De aquí en adelante se irá ordenando y creando arreglos según su interés de uso.

La siguiente función de suma importancia es la del algoritmo base:

```
knn (data_train, data_test, data_train_labels, k)
```

Esta función requiere la data_train que son los datos que se le va a dar como entrenamiento, la data_test que es el conjunto de datos que se quiere predecir, la data_train_labels que es una porción de la data_train de la cual la función devolverá la predicción de lo que falta de esa porción y el valor de K que previamente ya se ha seleccionado.

Y para comparar la predicción con los datos originales se tiene la siguiente función:

```
mean (knn (prc_train, prc_test, prc_train_labels, k) == prc_test_labels)
```

Esta función devolverá la confiabilidad que existe entre un vector y el otro, es decir cuan semejante es uno del otro, teniendo un rango de 0 a 1. Con esto se puede generar un vector de errores por canal y el resultado de error total mediante simple manipulación de fórmulas y comandos.

CAPÍTULO 4

4 RESULTADOS

En este capítulo se mostrarán las gráficas más relevantes en el análisis de nuestro estudio sobre los n-canales conjuntos según las respectivas mediciones que nos fueron proporcionadas, las mismas que fueron capturadas durante un periodo de 10 días en tres pisos diferentes de la torre A del edificio World Trade Center, ubicado en el sector norte de la ciudad de Guayaquil.

De igual manera se representará en las mismas gráficas el comportamiento del algoritmo predictivo basado en el método de los K-vecinos más cercanos obtenido del software R, de esta forma se podrá analizar y comparar la precisión de las predicciones generadas por el algoritmo con los resultados obtenidos del análisis con las mediciones reales, todo esto respecto a los n-canales conjuntos y la posibilidad tanto de encontrar n-canales conjuntos disponibles en determinado día de la semana como la posibilidad de encontrar por lo menos un grupo de canales conjuntos en el transcurso de dicho día.

Los resultados expuestos son del comportamiento de los días 4, 7 y 10 los cuales fueron seleccionados para tratar de observar un comportamiento más variado. Así mismo de cada uno de estos tres días se seleccionaron para graficar el comportamiento de los 3, 6 y 9 canales conjuntos, analizando la probabilidad de disponibilidad de las posibles agrupaciones de canales conjuntos mostrando una comparativa entre estos y los tres diferentes pisos, además de poder comparar gráficamente todos estos resultados con el algoritmo predictivo.

En la figura 4.1, figura 4.2, figura 4.3 se observan las gráficas de la función de distribución acumulada (CDF) del Día 4, Día 7 y Día 10 describiendo la probabilidad de disponibilidad por canales conjuntos de los datos originales y los datos predichos, realizando también una comparación entre los diferentes pisos. Se puede apreciar que la diferencia entre originales y predichos es mínima, es por esto que las líneas están sobrepuestas.

En la figura 4.1, para los diferentes pisos se puede observar que hay una alta probabilidad de que existan hasta 14 canales conjuntos disponibles (100%), el comportamiento en el Piso 1 y el Piso 9 tienden a ser la misma con la diferencia de que existe una variación en el rango entre 2 y 7 canales conjuntos. De entre los tres pisos se puede distinguir que el Piso 1 en todo el transcurso del Día 4 tiene mejor disponibilidad de canales conjuntos.

En la figura 4.2, se puede observar que el Piso 1 de este día al igual que el Día 4 tiene la mejor disponibilidad de canales conjuntos. El comportamiento del Piso 9 y el Piso 16 tienden a ser la misma con una pequeña variación en el rango entre 2 y 10 canales conjuntos. Para los tres pisos se puede decir que existen hasta 7 canales conjuntos en adelante con probabilidad de disponibilidad muy buena (entre 80% y 100%).

En la figura 4.3, se distingue que para este día los tres pisos tienen un comportamiento diferente. Se aprecia que el Piso 1 sigue siendo el que tiene mejor disponibilidad de canales conjuntos durante todo el día. Para los tres pisos existen hasta 8 canales conjuntos en adelante con una muy buena probabilidad (entre 83% y 100%).

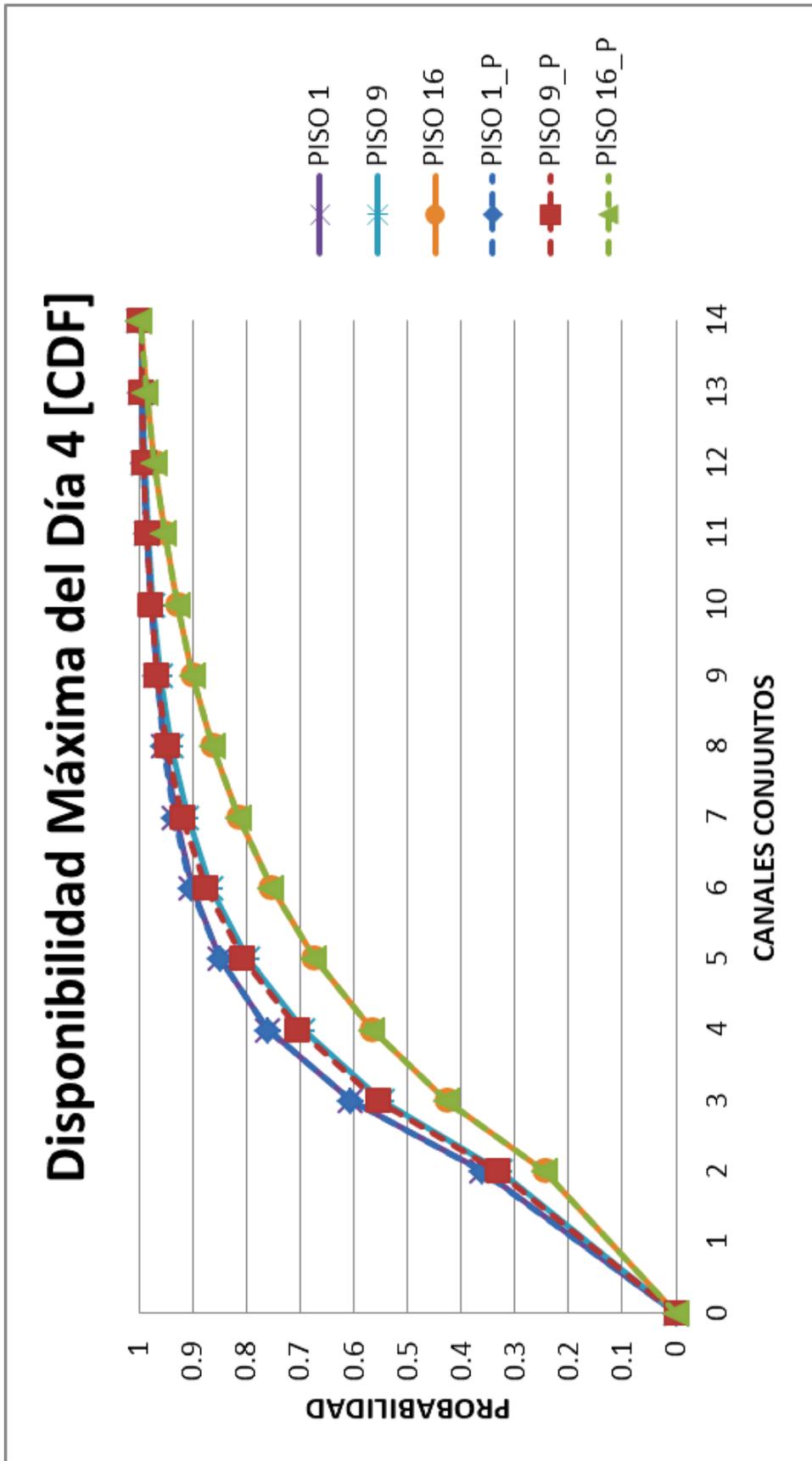


Figura 4.1 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 4

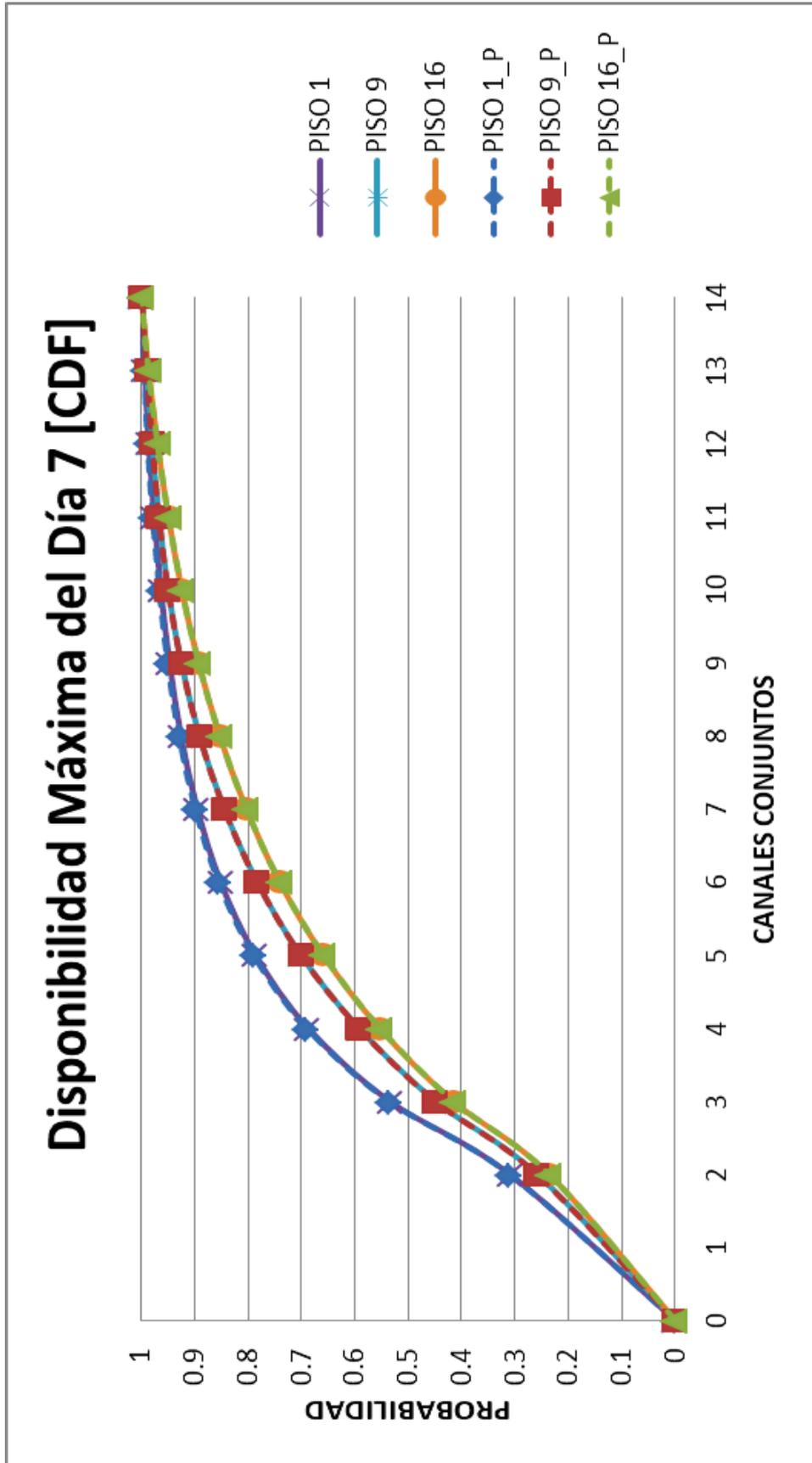


Figura 4.2 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 7

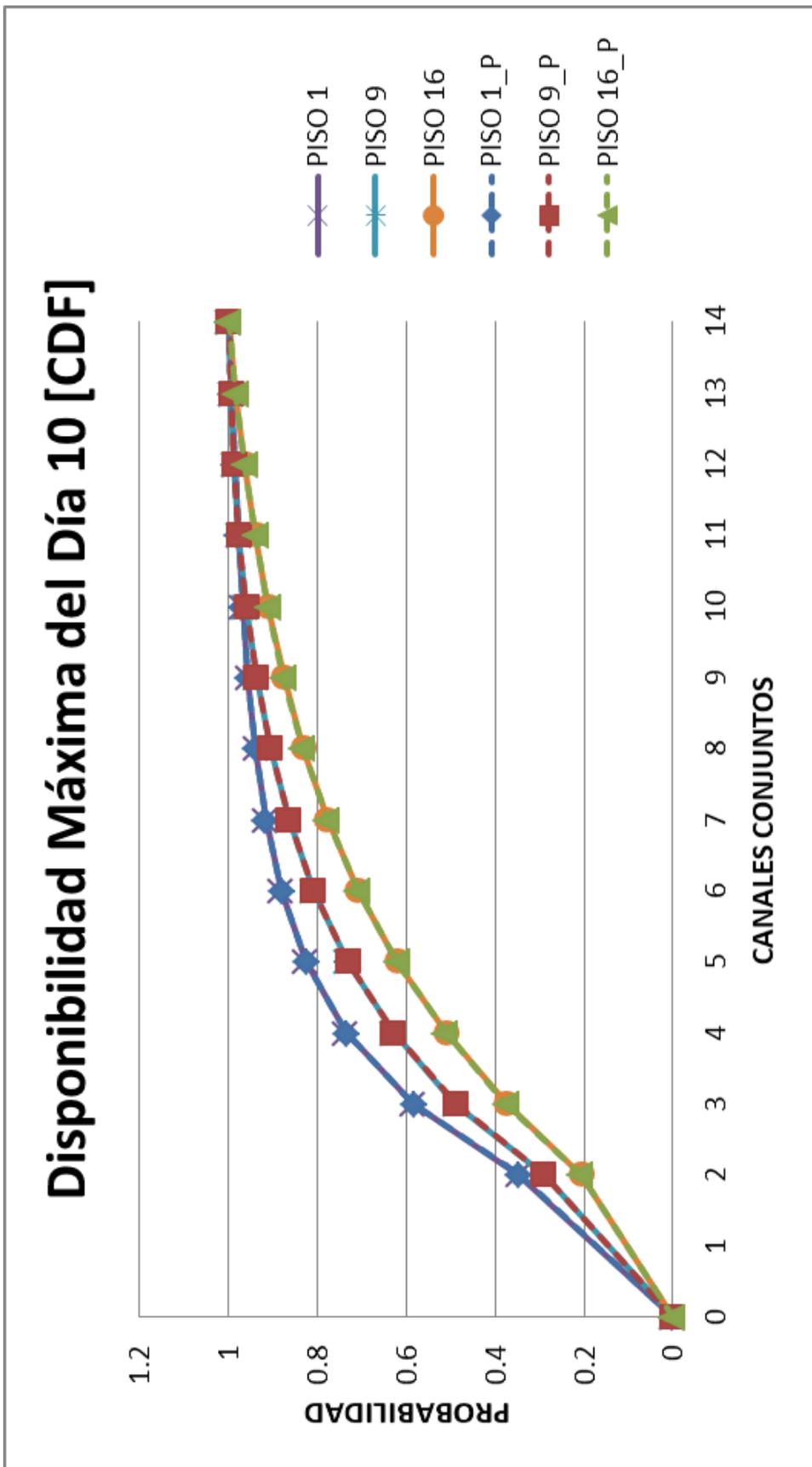


Figura 4.3 Distribución acumulada sobre la disponibilidad máxima de canales conjuntos en el día 10

En la figura 4.4, figura 4.5 y figura 4.6 se observan las gráficas de la función de distribución acumulada (CDF) del Día 4, Día 7 y Día 10 describiendo la probabilidad de disponibilidad en el tiempo por canales conjuntos de los datos originales y los datos predichos, realizando también una comparación entre los diferentes pisos. Se puede apreciar que la diferencia entre originales y predichos es mínima, es por esto que las líneas están sobrepuestas.

En la figura 4.4, para los diferentes pisos se puede observar que hay una alta probabilidad de que existan hasta 7 canales conjuntos en adelante disponibles en cierto instante de tiempo (entre 85% y 100%), el comportamiento en el Piso 1 y el Piso 9 tienden a ser la misma con la diferencia de que existe una variación en el rango entre 3 y 7 canales conjuntos. De entre los tres pisos se puede distinguir que el Piso 1 en todo el transcurso del Día 4 tiene mejor disponibilidad en el tiempo de canales conjuntos.

En la figura 4.5, se puede observar que el Piso 1 de este día al igual que el Día 4 tiene la mejor disponibilidad en el tiempo de canales conjuntos. El comportamiento del Piso 9 y el Piso 16 tienden a ser la misma con variación casi imperceptible en el rango entre 5 y 10 canales conjuntos. Para los tres pisos se puede decir que existen hasta 7 canales conjuntos en adelante con probabilidad de disponibilidad muy buena en cierto instante de tiempo (entre 84% y 100%).

En la figura 4.6, se distingue que para este día los tres pisos tienen un comportamiento diferente. Se aprecia que el Piso 1 sigue siendo el que tiene mejor disponibilidad en el tiempo de canales conjuntos durante todo el día. Para los tres pisos existen hasta 7 canales conjuntos en adelante con una muy buena probabilidad de disponibilidad en cierto instante de tiempo (entre 80% y 100%).

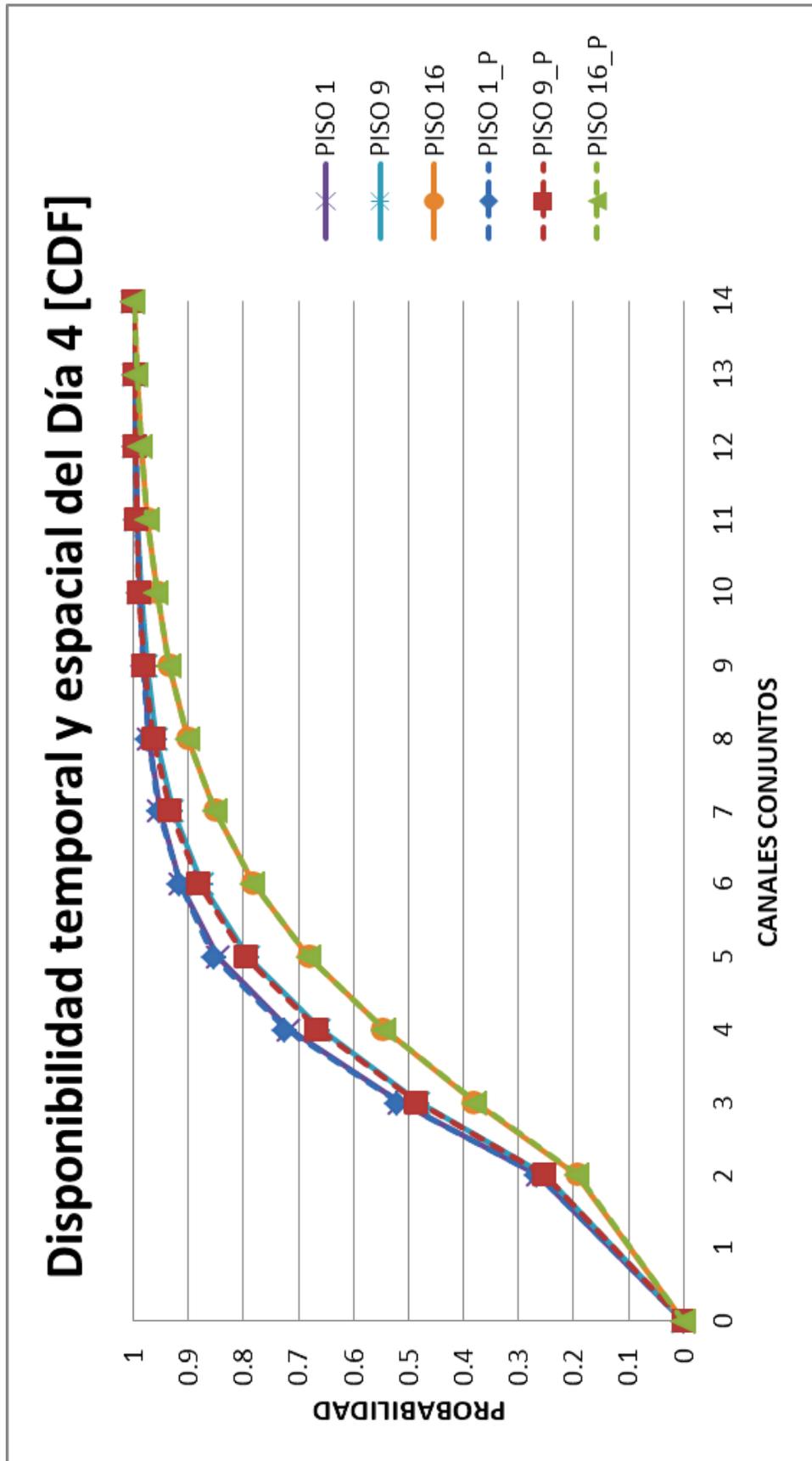


Figura 4.4 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 4

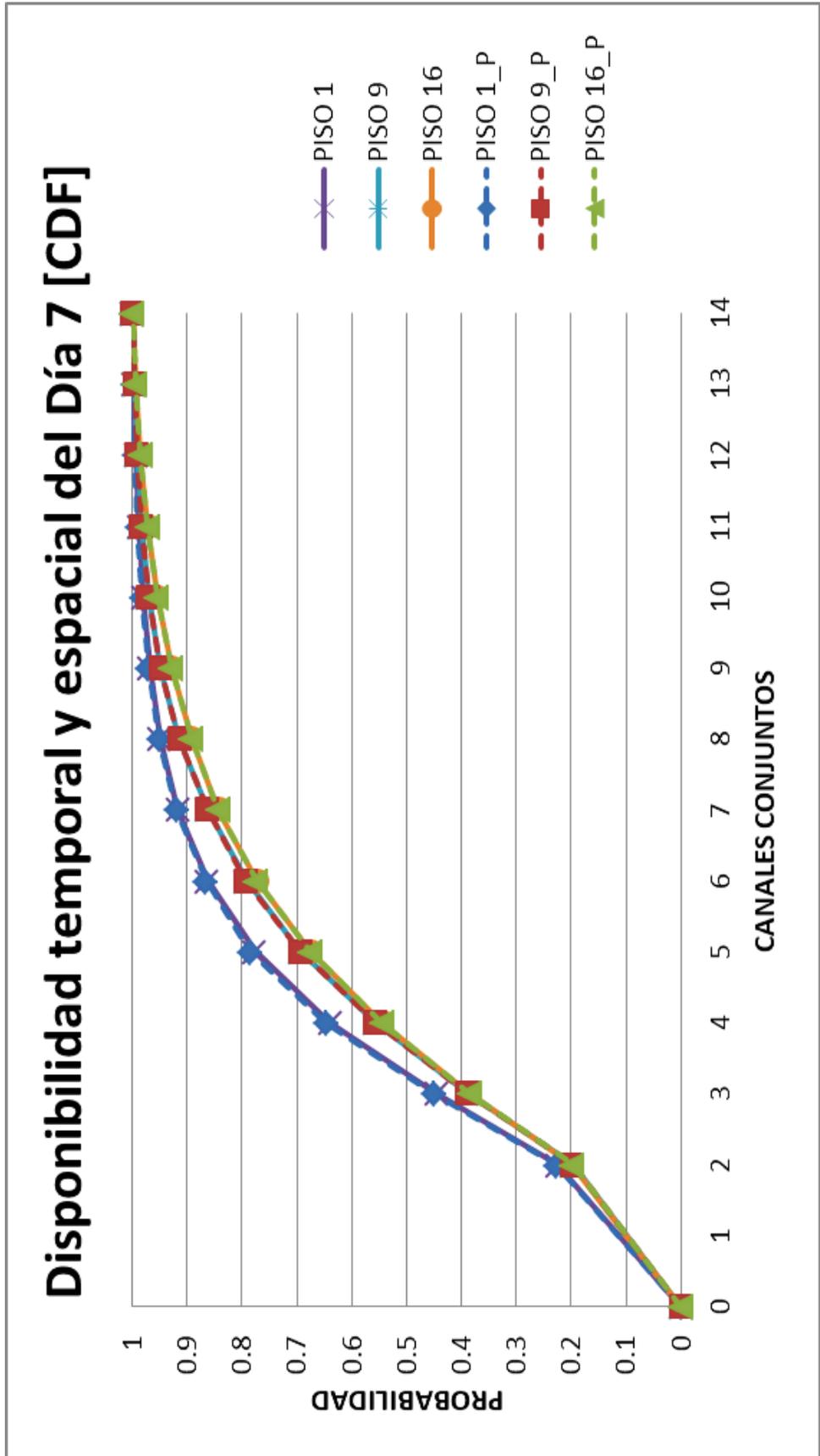


Figura 4.5 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 7.

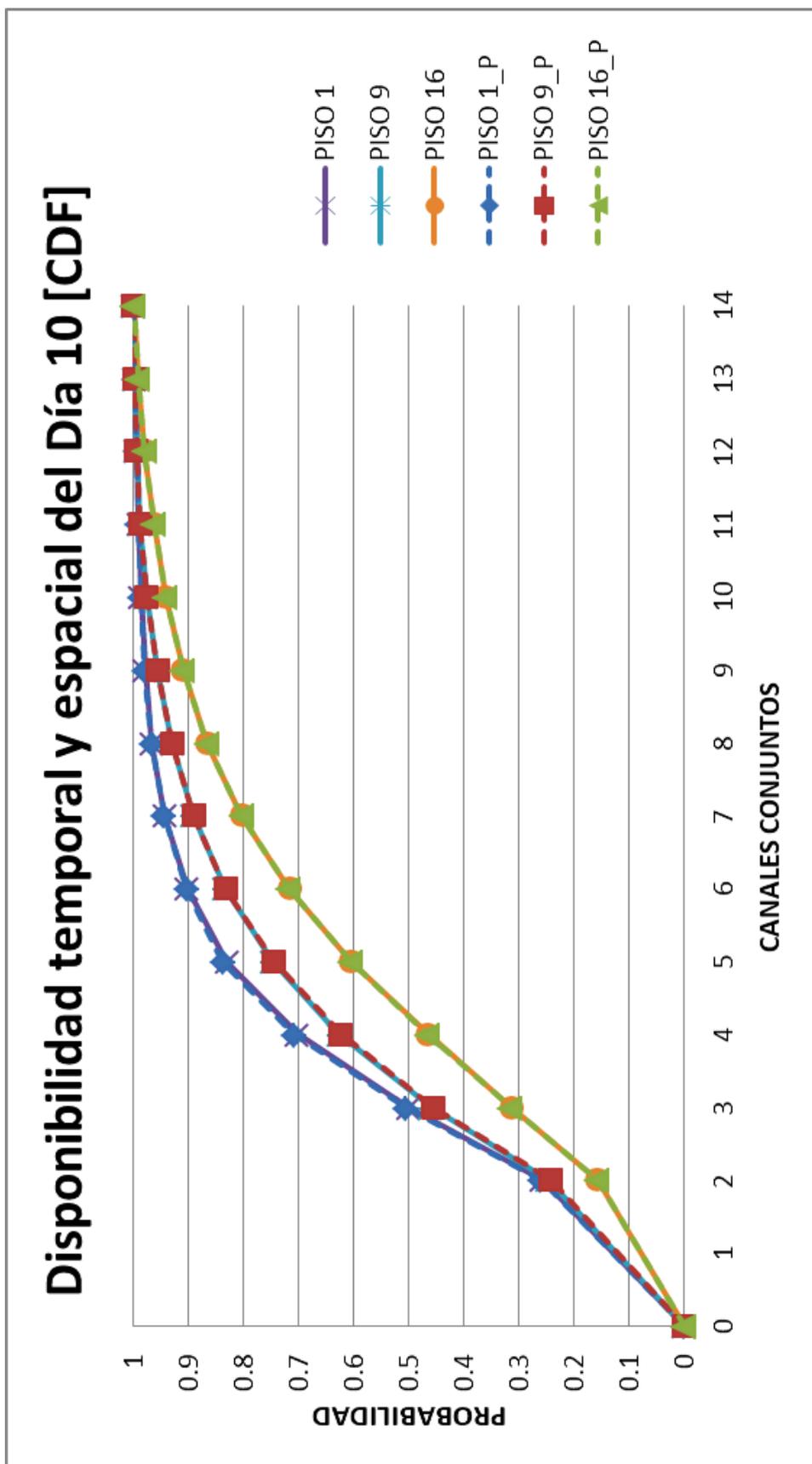


Figura 4.6 Distribución acumulada sobre la disponibilidad temporal y espacial de canales conjuntos del día 10.

En la figura 4.7, figura 4.8 y figura 4.9 se observan las gráficas de la función de distribución acumulada (CDF) del Día 4 para 3, 6 Y 9 Canales Conjuntos de los datos originales y los datos predichos. Se describe la probabilidad de que existan grupos de 3, 6 y 9 canales conjuntos disponibles durante todo el día, realizando también una comparación entre los tres pisos.

En la figura 4.7, se puede observar que el Piso 1 y el Piso 9 tienen casi el mismo comportamiento, ya que se puede apreciar que de hasta 4 grupos en adelante los dos pisos tienden a ser la misma, pero para grupos de 1, 2 y 3 se encuentra una pequeña variación entre los dos pisos. Con esas observaciones se puede decir que el Piso 9 tiene mejor probabilidad de que se puedan encontrar hasta 12 grupos de 3 canales conjuntos disponibles. Se puede notar que la predicción fue muy buena, es por esto que se observan las líneas sobrepuestas con las de los datos originales.

En la figura 4.8, el Piso 1 y el Piso 9 asemejan su comportamiento para 5 grupos en adelante, ya que desde 1 hasta 4 grupos se encuentra una variación entre los dos pisos. Se puede observar que el Piso 1 tiene mejor probabilidad de que se puedan encontrar hasta 9 grupos de 6 canales conjuntos disponible durante todo el día. En cuanto a predicciones, se puede notar que la predicción del Piso 1 no fue buena, a diferencia de la predicción de los Pisos 9 y 16 en los que si fue muy buena observando que se encuentran sobrepuestas con las de los datos originales.

En la figura 4.9, se observa que los tres pisos tienen un comportamiento diferente y los pisos 1 y 9 se llegan a intersecar en dos puntos. El Piso 9 tiene mejor probabilidad de que se puedan encontrar hasta 12 grupos de 9 canales conjuntos disponibles durante todo el día. Las predicciones del piso 1 y 16 fueron muy buenas, mientras que las del piso 9 no fue tan buena.

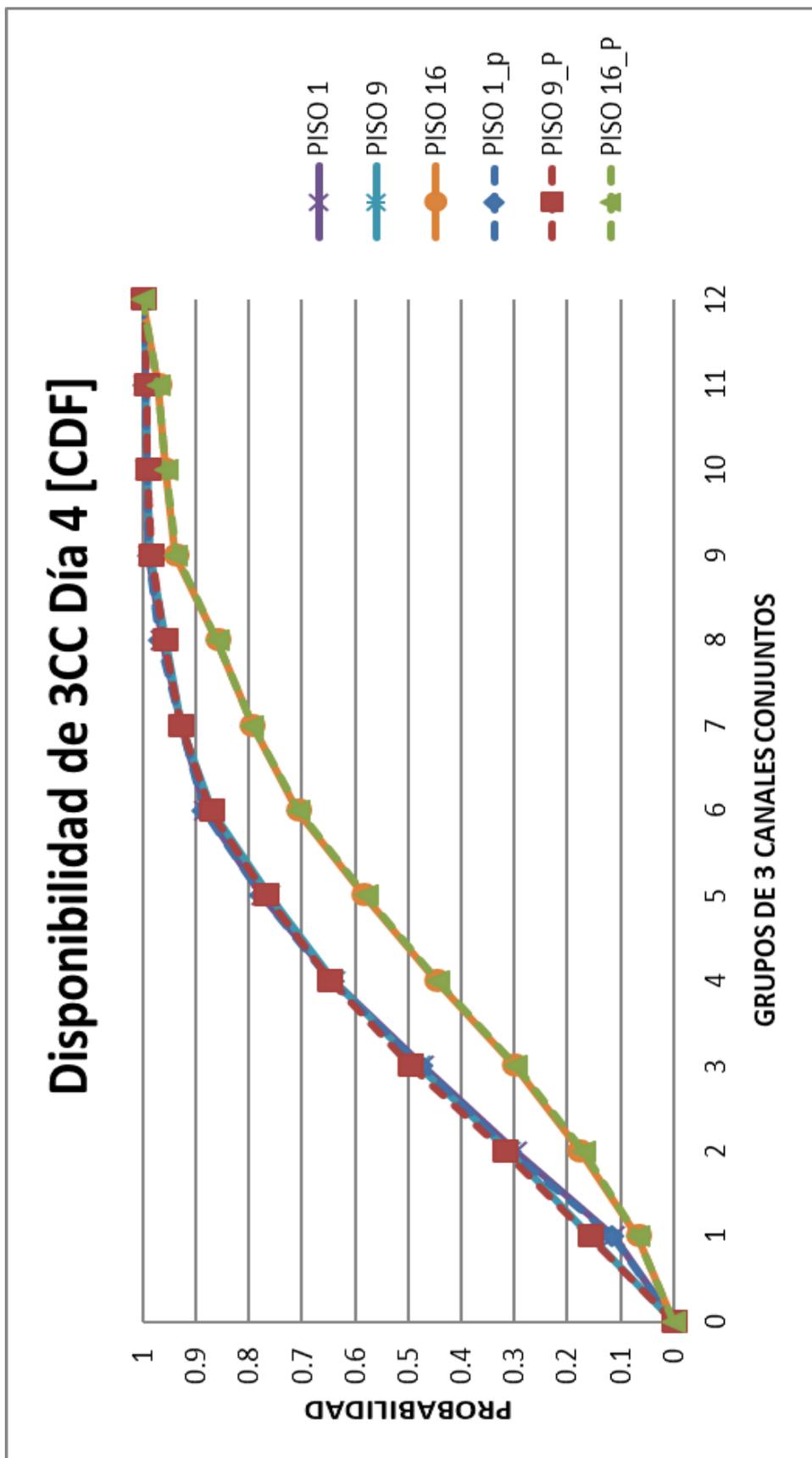


Figura 4.7 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 4.

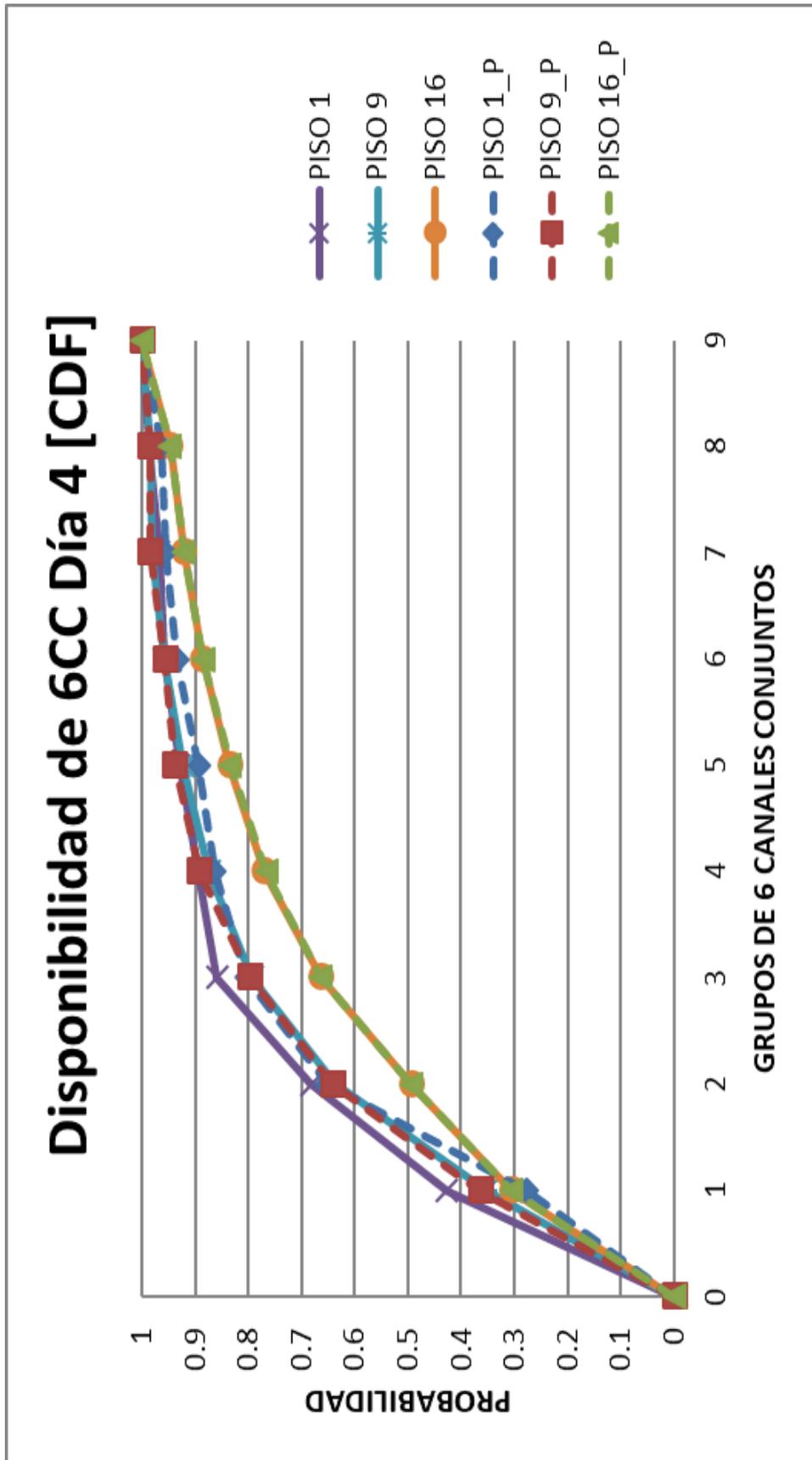


Figura 4.8 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 4.

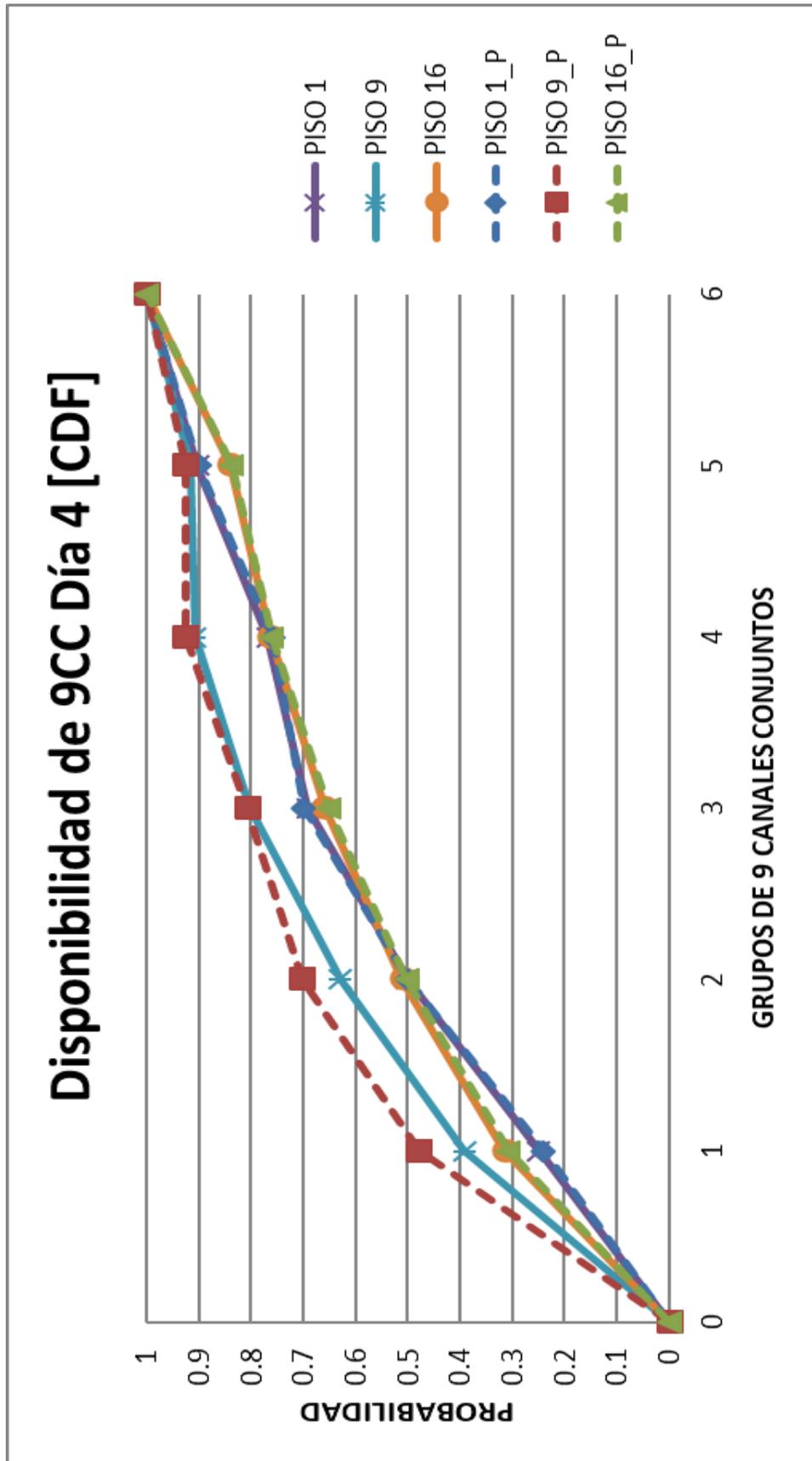


Figura 4.9 Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 4.

En la figura 4.10, figura 4.11 y figura 4.12 se observan las gráficas de la función de distribución acumulada (CDF) del Día 7 para 3, 6 Y 9 Canales Conjuntos de los datos originales y los datos predichos. Se describe la probabilidad de que existan grupos de 3, 6 y 9 canales conjuntos disponibles durante todo el día, realizando también una comparación entre los tres pisos.

En la figura 4.10, se puede observar que el Piso 1 y el Piso 9 tienen casi el mismo comportamiento, ya que se puede apreciar que para grupos de 4 hasta 8 se encuentra una pequeña variación entre los dos pisos; también se observa que el Piso 16 interseca a los pisos 1 y 9 en un punto. Se puede decir que el Piso 9 tiene mejor probabilidad de que se puedan encontrar hasta 12 grupos de 3 canales conjuntos disponibles, a pesar de que en el Piso 16 la probabilidad de encontrar hasta 2 grupos de 3 canales conjuntos es un poco mejor. Se puede notar que la predicción fue muy buena, es por esto que se observan las líneas sobrepuestas con las de los datos originales.

En la figura 4.11, los tres pisos tienen un comportamiento diferente. Se puede observar que el Piso 1 tiene mejor probabilidad de que se puedan encontrar hasta 9 grupos de 6 canales conjuntos disponible durante todo el día. En cuanto a predicciones, se puede notar que la predicción del Piso 1 no fue buena, a diferencia de la predicción de los Pisos 9 y 16 en los que, si fue muy buena, observando que se encuentran sobrepuestas con las de los datos originales.

En la figura 4.12, se observa que los tres pisos tienen un comportamiento diferente y los pisos 1 y 9 se llegan a intersecar en un punto. El Piso 9 tiene mejor probabilidad de que se puedan encontrar hasta 6 grupos de 9 canales conjuntos disponibles durante todo el día. Las predicciones del Piso 9 y 16 fueron muy buenas, mientras que en las del Piso 1 no fue tan buena con pequeños errores con respecto a la original.

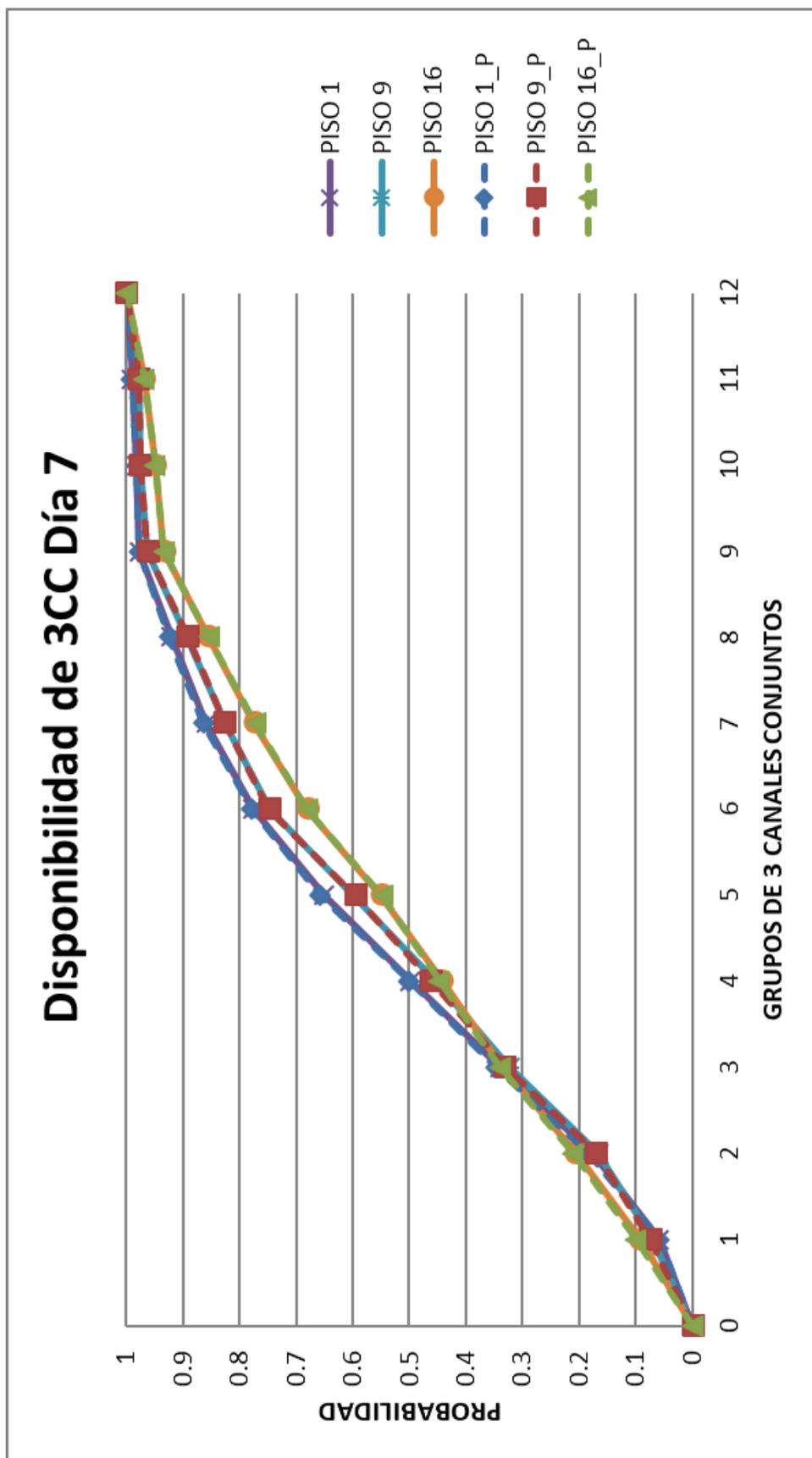


Figura 4.10 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 7.

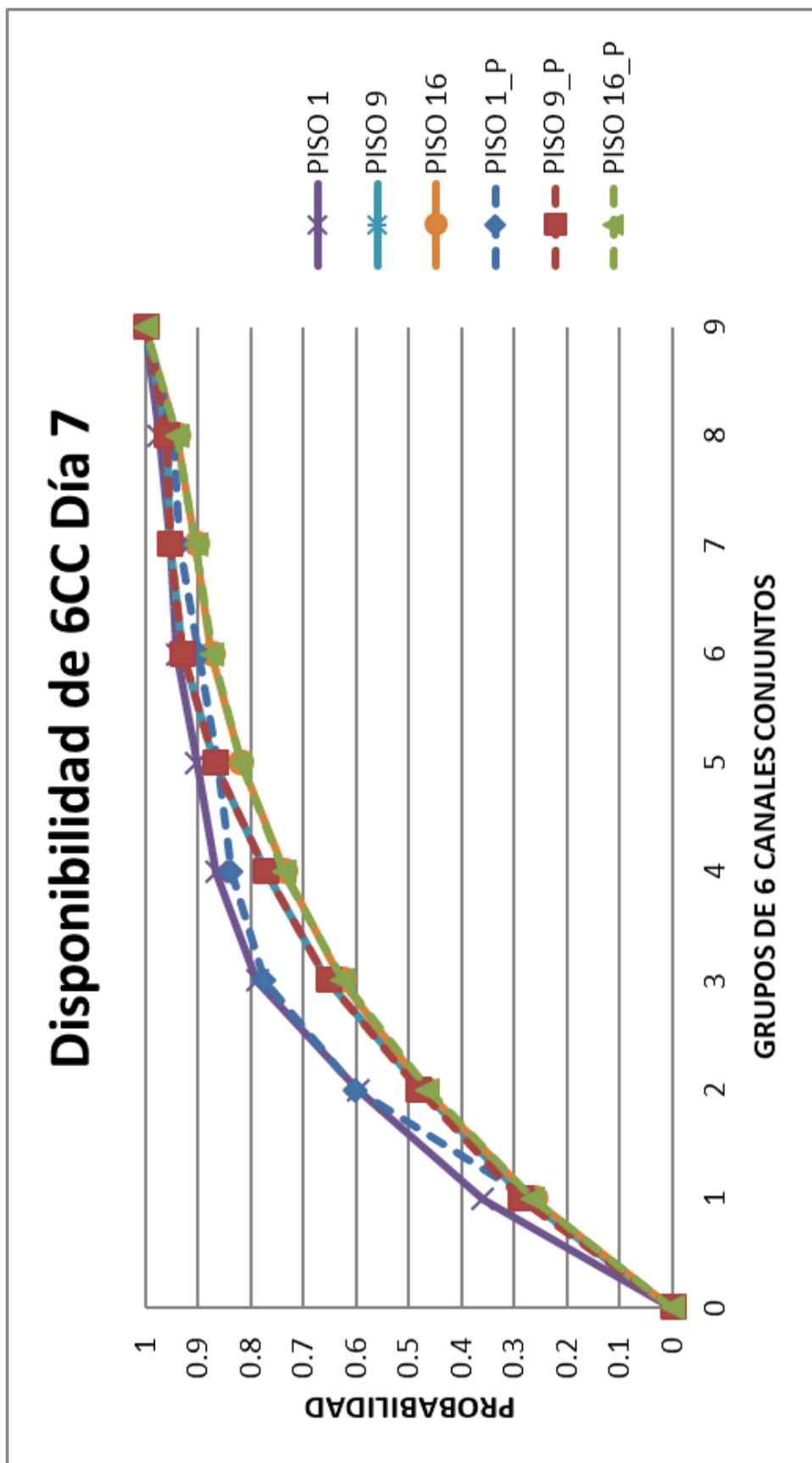


Figura 4.11 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 7.

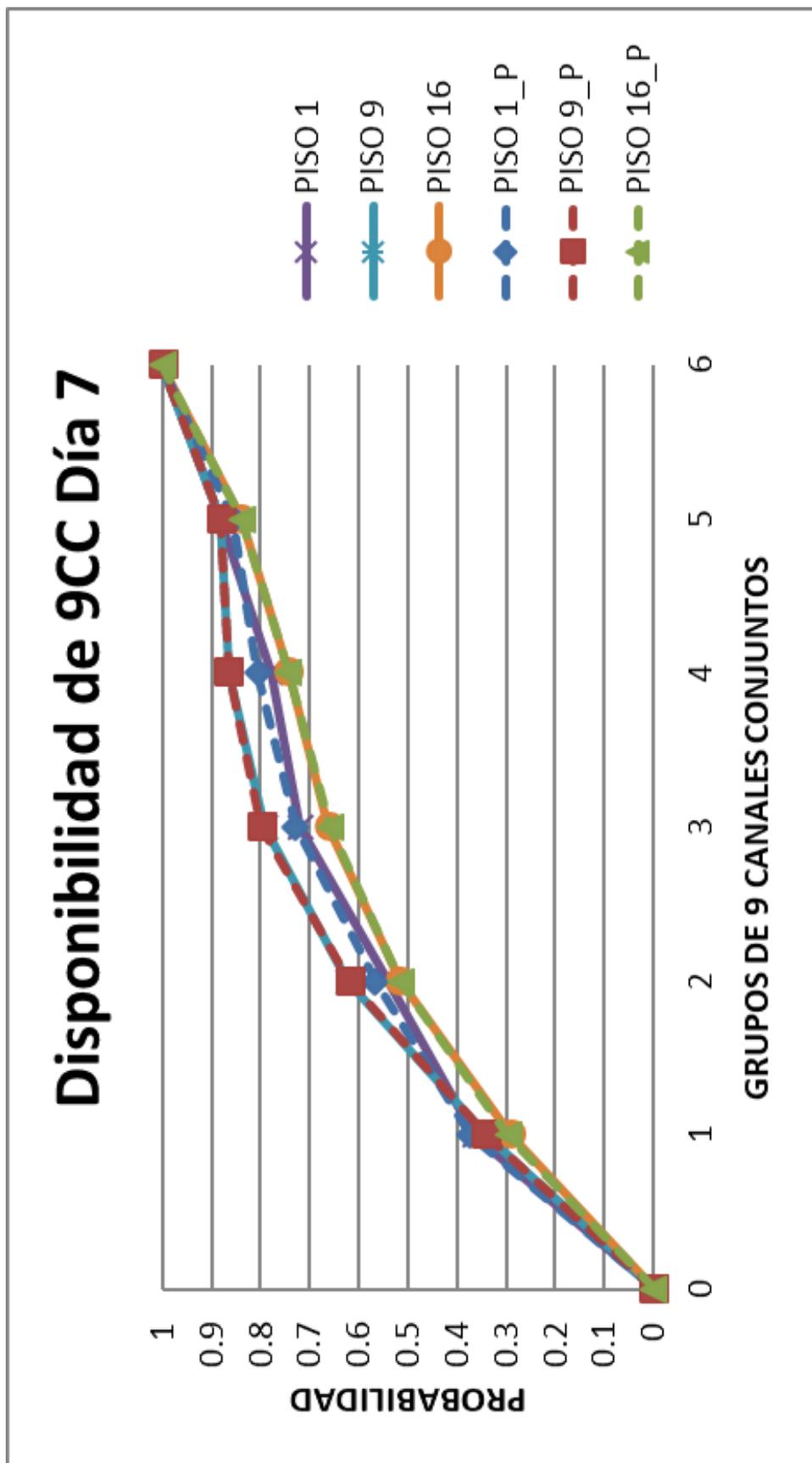


Figura 4.12 . Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 7.

En la figura 4.13, figura 4.14 y figura 4.15 se observan las gráficas de la función de distribución acumulada (CDF) del Día 10 para 3, 6 y 9 Canales Conjuntos de los datos originales y los datos predichos. Se describe la probabilidad de que existan grupos de 3, 6 y 9 canales conjuntos disponibles durante todo el día, realizando también una comparación entre los tres pisos.

En la figura 4.13, se puede observar que el Piso 1 y el Piso 9 se intersecan en un punto, se asemejan un poco entre si ya que se observa una variación pequeña entre sí; a diferencia del Piso 16 que tiene un comportamiento muy diferente con respecto a las anteriores ya descritas. Se puede decir que el Piso 1 tiene mejor probabilidad de que se puedan encontrar hasta 12 grupos de 3 canales conjuntos disponibles, a pesar de que en el Piso 9 la probabilidad de encontrar hasta 2 grupos de 3 canales conjuntos es un poco mejor. Se puede notar que la predicción fue muy buena, es por esto que se observan las líneas sobrepuestas con las de los datos originales.

En la figura 4.14, los tres pisos tienen un comportamiento diferente, a pesar de que el Piso 1 y el Piso 9 asemejan su comportamiento para 7 grupos en adelante. Se puede observar que el Piso 1 tiene mejor probabilidad de que se puedan encontrar hasta 9 grupos de 6 canales conjuntos disponible durante todo el día. En cuanto a predicciones, se puede notar que la predicción del Piso 1 no fue buena, a diferencia de la predicción de los Pisos 9 y 16 en los que, si fue muy buena, observando que se encuentran sobrepuestas con las de los datos originales.

En la figura 4.15, se observa que los tres pisos tienen un comportamiento diferente se intersecan en diferentes puntos. El Piso 1 tiene mejor probabilidad de que se puedan encontrar hasta 2 grupos de 9 canales conjuntos disponibles, mientras que el Piso 9 tiene mejor probabilidad de encontrar hasta 3 grupos de 9 canales conjuntos en adelante. Las predicciones del Piso 9 y 16 fueron muy buenas, mientras que en las del Piso 1 no fue tan buena con pequeños errores con respecto a la original.

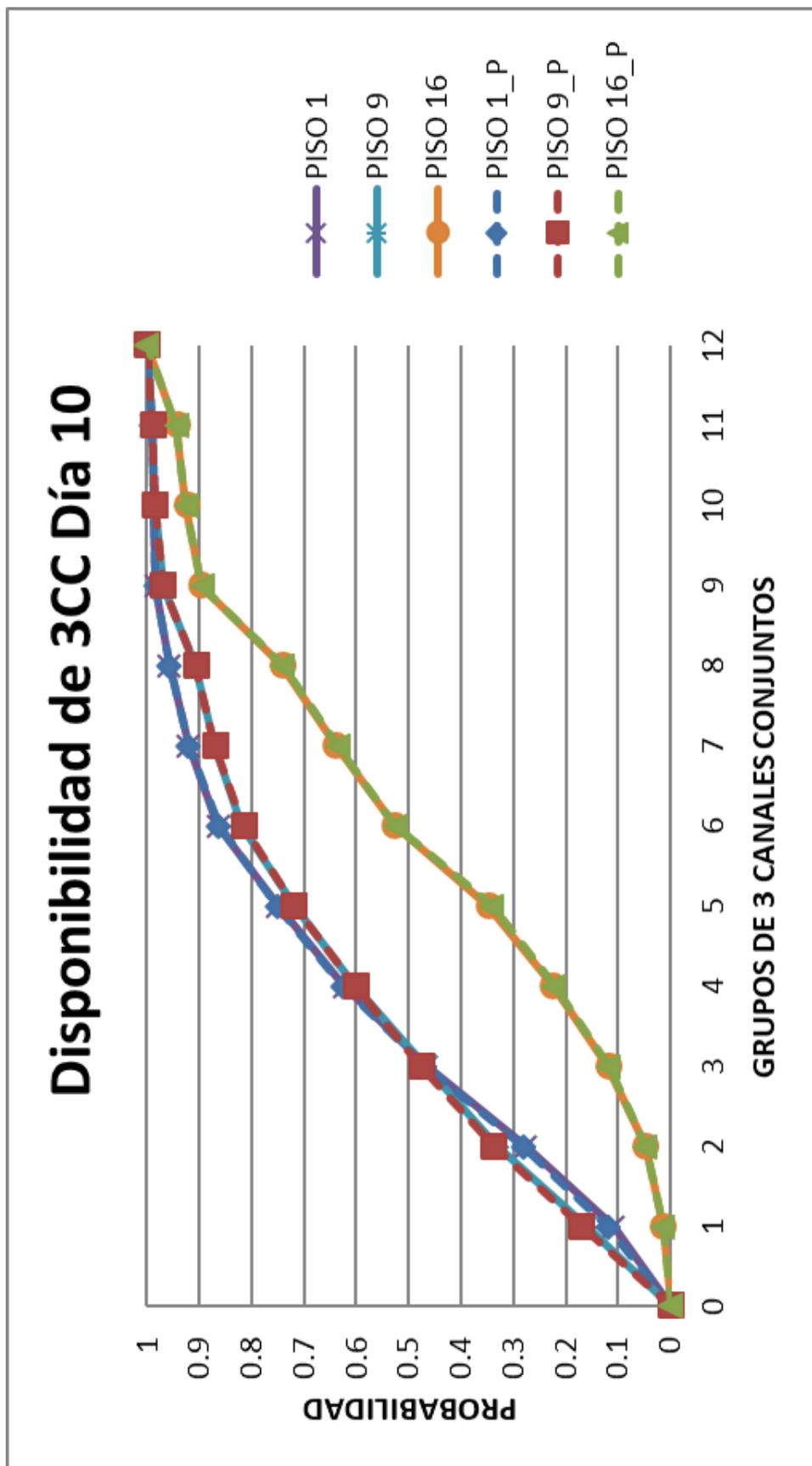


Figura 4.13 Distribución acumulada sobre la disponibilidad de 3 canales conjuntos en el día 10.

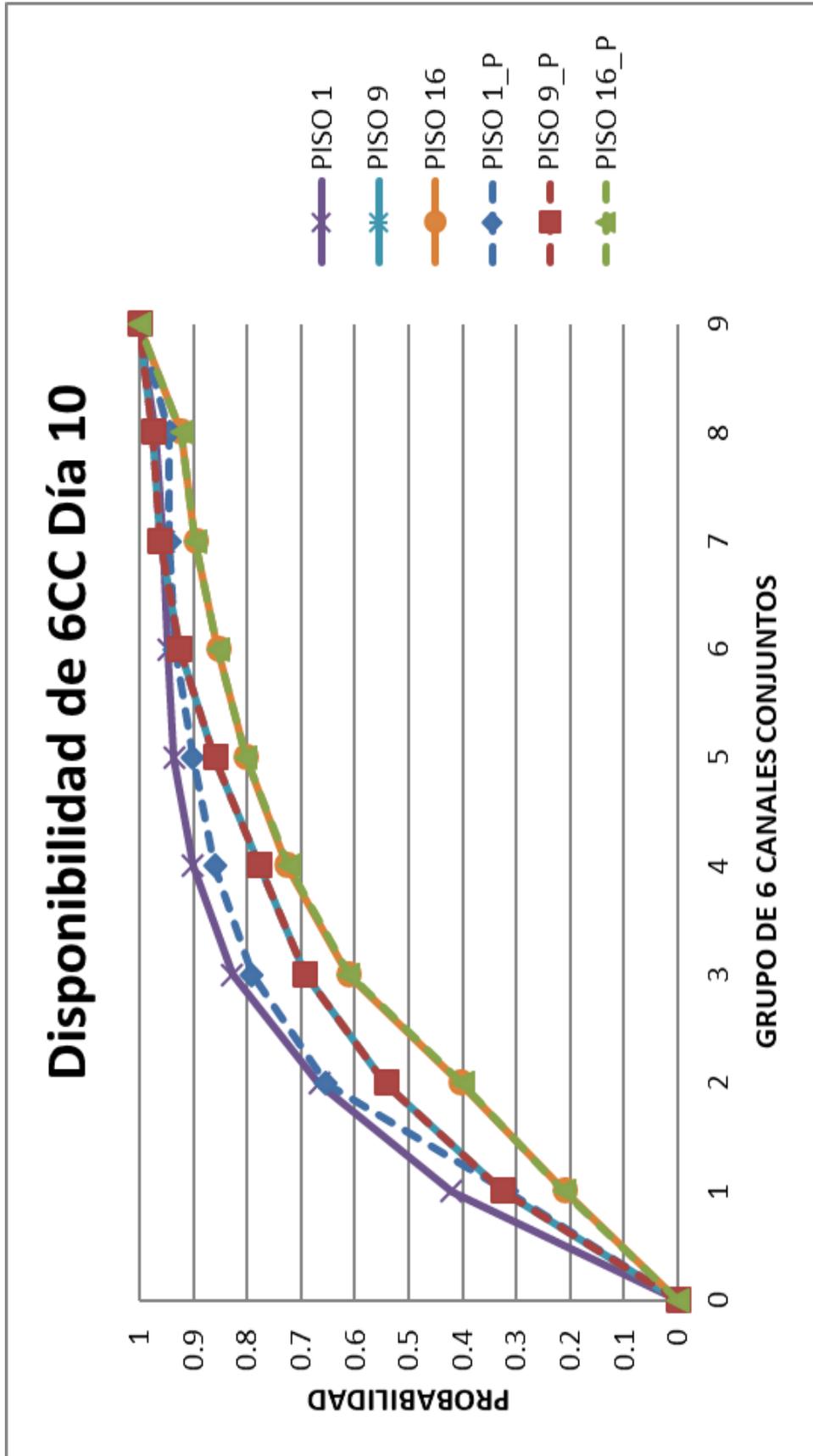


Figura 4.14 Distribución acumulada sobre la disponibilidad de 6 canales conjuntos en el día 10.

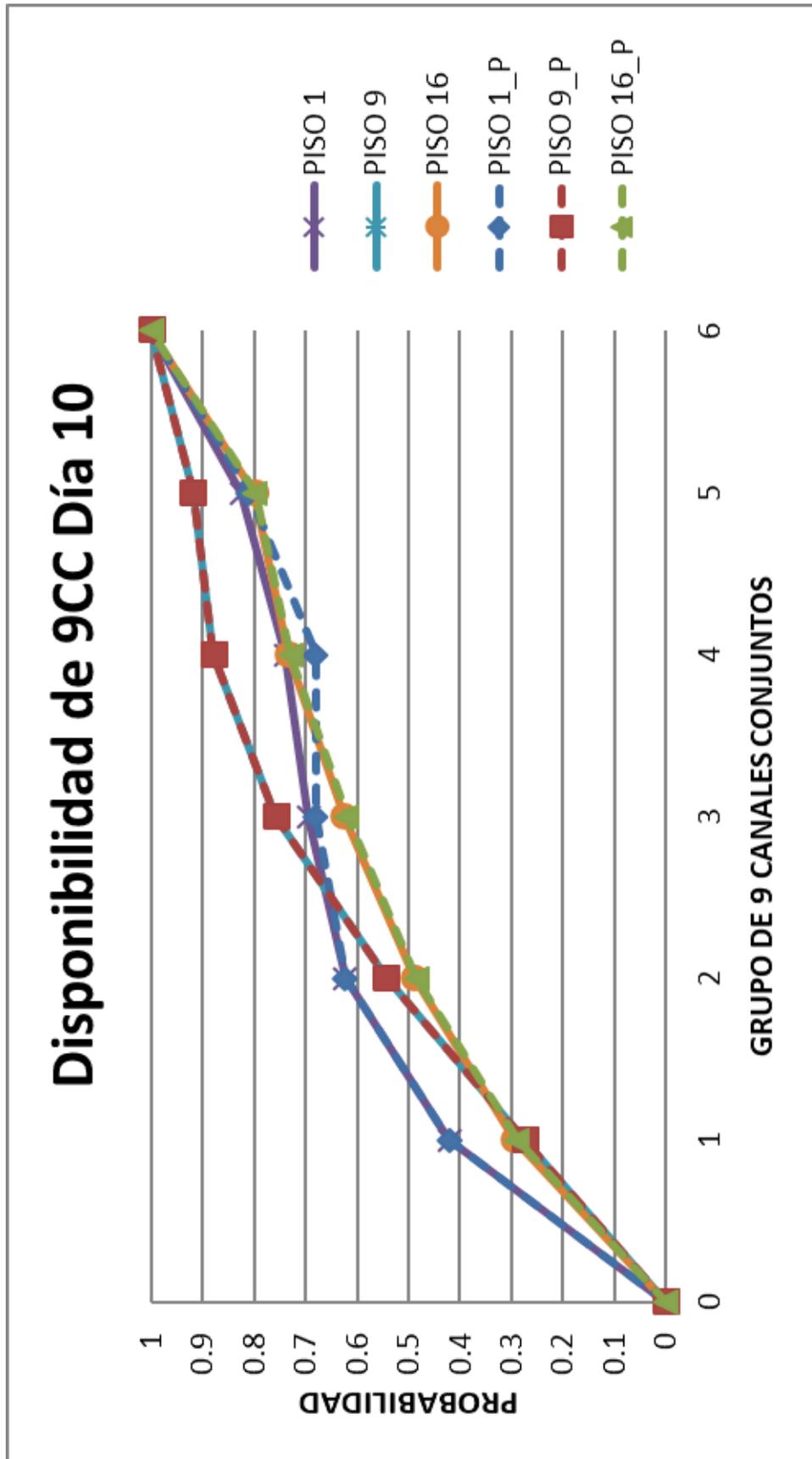


Figura 4.15 Distribución acumulada sobre la disponibilidad de 9 canales conjuntos en el día 10.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Las predicciones que se obtuvieron siempre se mantuvieron con un alto nivel de confiabilidad, es decir al momento de comparar las matrices predichas día a día de los n-canales conjuntos con las matrices originales se obtuvo un error menor al 1% en la mayoría de los días, sin embargo, cuando se predijeron los n-grupos disponibles para n-canales conjuntos el error fue aumentando, llegando a alcanzar hasta un 12% de error en el caso del día 4 para los 9 canales conjuntos.
- El modelo de predicción KNN obtuvo buenos resultados en ambientes indoor y a pesar de que hubo un bajo porcentaje de error, este puede ser corregido con técnicas de ponderación adicionales y el uso de sus parámetros óptimos.
- El comportamiento más irregular entre las predicciones y los datos reales ocurrieron al buscar la disponibilidad según la cantidad de grupos de n-canales conjuntos existentes, también se observó que a partir de 2 grupos disponibles para 9 canales conjuntos habrá una mayor probabilidad de encontrar más grupos disponibles para el piso 9 que para los pisos 1 y 16, a diferencia del comportamiento para menores canales conjuntos donde el piso 1 se mantenía con mayor probabilidad de encontrar mayor cantidad de grupos de canales conjuntos.
- Sobre los canales conjuntos y su disponibilidad en este edificio, debido al comportamiento que se observó en los tres pisos, se puede considerar que para futuras mediciones que vayan a ser realizadas en lugares con condiciones similares sean efectuadas en un solo piso y a partir de los resultados determinar un comportamiento aproximado de los demás pisos, en nuestro estudio se determinó que la mayor disponibilidad de n-canales conjuntos se dio para el primer piso y este comportamiento se mantuvo con el transcurso de los días.

Recomendaciones

- Realizar diferentes predicciones alternando el valor de K y los datos de entrenamiento para obtener el error en la predicción, con esto se podrá escoger el valor de K adecuado para el algoritmo predictivo y así conseguir resultados con un bajo porcentaje de error.
- Graficar las potencias obtenidas durante el día para observar el piso de ruido y según esto escoger el umbral de potencia donde no se vea afectada por el ruido la decisión de disponibilidad del canal.
- Organizar los datos de las mediciones, así como los resultados que se van obteniendo para tener un fácil acceso y un mejor manejo sobre ellos.

REFERENCIAS

- [1] <https://www.laprensa.com.ni/2014/09/06/economia/210889-optimizacion-del-uso-del-espectro-radioelectrico>.
- [2] I. A. Manjarrez Alva, «Radio Cognitiva», Universidad Nacional Autónoma de México, Ciudad de México, 2010.
- [3] A. Llanos, Gestión del espectro radioeléctrico en Ecuador: nueva modalidad para radiodifusión y televisión abierta. Quito: Universidad Andina Simón Bolívar, Sede Ecuador: Corporación Editora Nacional, 2013.
- [4] A. Castro Lechtaler et al., «Utilización de los Espacios Blancos: Una solución al problema de la saturación del espectro de frecuencias», p. 11, 2016.
- [5] A. Galvis Quintero, «ACCESO DINAMICO AL ESPECTRO: ESTADO ACTUAL, TENDENCIAS Y RETOS», p. 21, 2008.
- [6] ITU-Radiocommunication Sector of ITU, «Monte Carlo simulation methodology for the use in sharing and compatibility studies between different radio services or systems», 2017.
- [7] Arcotel. (2017). La regulación en el sector de las Telecomunicaciones genera impacto positivo. Guayaquil, Ecuador. Recuperado de: <http://www.arcotel.gob.ec/la-regulacion-en-el-sector-de-las-telecomunicaciones-genera-impacto-positivo-en-la-ciudadania/>
- [8] Contemporary Engineering Sciences, Vol. 10, 2017, no. 15, 721 - 738 HIKARI Ltd, www.m-hikari.com <https://doi.org/10.12988/ces.2017.7877>
- [9] Félix Barba Barba. (2012). Estudio de algoritmos de localización en interiores, para tecnologías móviles de última (tesis de masterado). Universidad Politécnica de Madrid, Madrid, España.

- [10] Lin, T.N. & Lin, P.C., 2005. Performance comparison of indoor positioning techniques base on location fingerprinting in wireless networks. In Wireless Networks, Communications and Mobile Computing, 2005 International Conference on. IEEE, pp. 1569–1574. Disponible en: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1549647.
- [11] https://www.enacom.gob.ar/-que-es-el-espectro-radioelectrico-_p117.
- [12] Camargo Olivares, J.L. (2009) Modelo de cobertura para redes inalámbricas. Sevilla, España. e-Reading. Recuperado de: <http://bibing.us.es/proyectos/abreproy/11761/fichero/Volumen1%252F7-Cap%C3%ADtulo3+-+Redes+inal%C3%A1mbricas+de+%C3%A1rea+local+%28WLAN%29.pdf>
- [13] Luaces Novoa, J.M(2003). Seguridad en redes inalámbricas de área local. Catalunya, España.: UOC. Recuperado de: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/18804/6/jluacesTFC0113memoria.pdf>
- [14] Us. Robotics. Wireless Lan Networking. Illinois, United States. Recuperado de: <http://support.usr.com/download/whitepapers/wireless-wp.pdf>
- [15] Qing Zhao and Brian M. Sadler. A Survey of Dynamic Spectrum Access
- [16] Xin, ChunSheng, Song, Min. Spectrum Sharing for Wireless Communications
- [17] Daniel Ospina, Andrés F. Ocampo, Roberto Hincapié. Aplicación de esquemas de Acceso Oportunista al Espectro en ambientes inalámbricos heterogéneos, 2011.
- [18] HANS RAÚL MÁRQUEZ RAMOS. Arquitecturas de radio cognitiva: una revisión actual, 2012.
- [19] José Manuel Huidobro. Radio Cognitiva. La radio se vuelve inteligente, 2011.
- [20] Raikel Bordón López, Samuel Montejo Sánchez. La Radio Cognitiva y su Impacto en el Uso Eficiente del Espectro de Radio

- [21] Cesar Hernández¹, Luis Fernando Pedraza Martínez², Fredy Hernán Martínez Sarmiento³. Algoritmos para asignación de espectro en redes de radio cognitiva, 2015.
- [22] D. Hand, H. Mannila and P. Smyth, Principles of Data Mining. MIT Press, 2001.
- [23] D. R. Wilson and T. R. Martinez, Improved heterogeneous distance functions. Journal of Artificial Intelligence Research 6(1), 1997, 1-34.
- [24] López, A. M. (2015) Análisis de Conglomerados. Ciencias del Comportamiento. Recuperado de: <http://personal.us.es/analopez/ac.pdf>
- [26] Business Analytics (2015) Best way to learn kNN Algorithm using R Programming. Recuperado de: <https://www.analyticsvidhya.com/blog/2015/08/learning-concept-knn-algorithms-programming/>
- [27] Ripley, B. D. (1996) Pattern Recognition and Neural Networks. Cambridge

ANEXOS

	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
Día 2	0.759%	0.295%	0.076%	0.021%	0.015%	0.035%	0.030%	0.023%	0.000%	0.017%	0.000%	0.000%	0.000%
Día 3	0.887%	0.214%	0.038%	0.021%	0.123%	0.148%	0.129%	0.255%	0.167%	0.139%	0.185%	0.104%	0.069%
Día 4	0.743%	0.278%	0.120%	0.069%	0.131%	0.113%	0.089%	0.081%	0.069%	0.052%	0.023%	0.000%	0.000%
Día 5	1.207%	0.330%	0.158%	0.063%	0.093%	0.113%	0.079%	0.046%	0.014%	0.017%	0.000%	0.000%	0.000%
Día 6	0.609%	0.168%	0.044%	0.035%	0.054%	0.078%	0.079%	0.081%	0.069%	0.052%	0.046%	0.069%	0.000%
Día 7	0.913%	0.330%	0.164%	0.056%	0.131%	0.217%	0.188%	0.162%	0.250%	0.260%	0.440%	0.417%	0.000%
Día 8	0.785%	0.353%	0.145%	0.069%	0.093%	0.052%	0.069%	0.046%	0.028%	0.017%	0.023%	0.000%	0.000%
Día 9	0.983%	0.272%	0.032%	0.049%	0.069%	0.043%	0.040%	0.023%	0.014%	0.000%	0.000%	0.000%	0.000%
Día 10	0.748%	0.231%	0.088%	0.028%	0.093%	0.078%	0.079%	0.069%	0.083%	0.087%	0.069%	0.000%	0.000%

Anexo 1. Tabla de errores en la predicción de canales conjuntos del día 4.

	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
Día 2	1.026%	0.249%	0.133%	0.028%	0.293%	0.260%	0.188%	0.185%	0.222%	0.191%	0.255%	0.313%	0.278%
Día 3	1.394%	0.677%	0.354%	0.153%	0.093%	0.052%	0.060%	0.035%	0.028%	0.017%	0.023%	0.035%	0.000%
Día 4	1.031%	0.451%	0.215%	0.076%	0.332%	0.321%	0.357%	0.394%	0.444%	0.347%	0.139%	0.104%	0.069%
Día 5	1.042%	0.365%	0.177%	0.028%	0.185%	0.156%	0.169%	0.220%	0.222%	0.174%	0.185%	0.174%	0.347%
Día 6	0.876%	0.411%	0.240%	0.111%	0.085%	0.078%	0.079%	0.069%	0.153%	0.035%	0.023%	0.035%	0.000%
Día 7	0.849%	0.301%	0.170%	0.097%	0.123%	0.035%	0.030%	0.035%	0.042%	0.035%	0.046%	0.000%	0.000%
Día 8	0.929%	0.220%	0.133%	0.014%	0.062%	0.061%	0.060%	0.058%	0.069%	0.087%	0.069%	0.000%	0.000%
Día 9	0.935%	0.284%	0.152%	0.049%	0.069%	0.061%	0.060%	0.069%	0.069%	0.035%	0.023%	0.035%	0.000%
Día 10	1.271%	0.289%	0.069%	0.028%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%

Anexo 2. Tabla de errores en la predicción de canales conjuntos del día 7.

	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
Día 2	0.855%	0.284%	0.177%	0.028%	0.031%	0.043%	0.010%	0.012%	0.028%	0.000%	0.000%	0.000%	0.000%
Día 3	1.047%	0.249%	0.107%	0.049%	0.039%	0.061%	0.040%	0.046%	0.056%	0.035%	0.046%	0.000%	0.000%
Día 4	1.688%	0.301%	0.069%	0.007%	0.046%	0.043%	0.060%	0.058%	0.028%	0.017%	0.000%	0.000%	0.000%
Día 5	1.143%	0.388%	0.189%	0.056%	0.054%	0.052%	0.099%	0.069%	0.111%	0.104%	0.046%	0.000%	0.000%
Día 6	1.090%	0.249%	0.051%	0.000%	0.031%	0.026%	0.030%	0.069%	0.042%	0.052%	0.069%	0.000%	0.000%
Día 7	1.159%	0.289%	0.107%	0.049%	0.069%	0.061%	0.050%	0.035%	0.056%	0.017%	0.000%	0.000%	0.000%
Día 8	1.229%	0.353%	0.076%	0.021%	0.378%	0.295%	0.337%	0.370%	0.444%	0.556%	0.602%	0.625%	0.000%
Día 9	1.058%	0.417%	0.208%	0.174%	0.162%	0.147%	0.258%	0.208%	0.208%	0.208%	0.277%	0.312%	0.000%
Día 10	0.737%	0.208%	0.057%	0.014%	0.093%	0.113%	0.069%	0.069%	0.028%	0.035%	0.023%	0.000%	0.000%

Anexo 3. Tabla de errores en la predicción de canales conjuntos del día 10.

DÍA 4													
Datos Originales													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	45.56%	30.90%	19.52%	11.25%	6.89%	3.91%	2.50%	1.68%	1.29%	0.97%	0.74%	0.59%	0.35%
P9	43.09%	28.77%	19.40%	13.42%	8.77%	5.77%	3.71%	2.35%	1.54%	0.90%	0.70%	0.49%	0.49%
P16	55.49%	41.35%	31.55%	24.12%	18.29%	13.93%	10.85%	8.40%	6.62%	5.30%	4.26%	3.44%	2.78%
Datos Predicción													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	45.97%	31.13%	19.64%	11.32%	6.53%	4.00%	2.57%	1.74%	1.36%	1.03%	0.76%	0.59%	0.35%
P9	43.65%	29.06%	19.59%	13.47%	8.99%	5.97%	3.90%	2.64%	1.82%	1.25%	0.83%	0.60%	0.56%
P16	54.75%	41.39%	31.60%	24.13%	18.25%	13.88%	10.81%	8.42%	6.62%	5.32%	4.26%	3.44%	2.78%

Anexo 4. Tabla de disponibilidad máxima de canales conjuntos en el día 4.

DÍA 7													
Datos Originales													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	52.99%	38.52%	26.25%	16.92%	11.12%	7.48%	5.17%	3.64%	2.65%	2.10%	1.62%	0.97%	0.97%
P9	53.69%	39.27%	29.53%	22.49%	16.80%	12.74%	9.61%	7.05%	5.04%	3.52%	2.57%	2.15%	2.01%
P16	52.87%	40.12%	30.80%	23.66%	18.18%	14.08%	11.13%	8.88%	7.03%	5.59%	4.56%	3.79%	2.92%
Datos Predicción													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	53.46%	38.80%	26.36%	16.95%	11.23%	7.54%	5.24%	3.70%	2.80%	2.21%	1.69%	1.39%	0.97%
P9	53.58%	39.43%	29.67%	22.56%	16.89%	12.76%	9.62%	7.06%	5.09%	3.56%	2.62%	2.15%	2.01%
P16	52.98%	40.29%	30.86%	23.68%	18.17%	14.07%	11.12%	8.90%	7.06%	5.61%	4.56%	3.79%	2.92%

Anexo 5. Tabla de Disponibilidad máxima de canales conjuntos en el día 7.

DÍA 10													
Datos Originales													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	46.49%	31.77%	20.49%	12.15%	7.45%	4.73%	3.15%	2.23%	1.65%	1.32%	1.11%	1.04%	0.83%
P9	41.93%	28.56%	20.40%	14.85%	11.01%	8.06%	5.99%	4.53%	3.36%	2.33%	1.53%	1.04%	0.83%
P16	66.86%	54.23%	44.11%	35.46%	28.19%	22.33%	17.74%	13.97%	11.32%	9.27%	7.66%	6.39%	5.49%
Datos Predicción													
	2 CC	3 CC	4 CC	5 CC	6 CC	7 CC	8 CC	9 CC	10 CC	11 CC	12 CC	13 CC	14 CC
P1	46.79%	31.93%	20.56%	12.18%	7.37%	4.71%	3.11%	2.16%	1.63%	1.34%	1.18%	1.04%	0.83%
P9	42.07%	28.70%	20.47%	14.87%	11.01%	8.06%	5.99%	4.53%	3.36%	2.33%	1.53%	1.04%	0.83%
P16	66.45%	54.19%	44.12%	35.46%	28.15%	22.33%	17.69%	13.99%	11.29%	9.27%	7.69%	6.39%	5.49%

Anexo 6. Tabla de disponibilidad máxima de canales conjuntos en el día 10.