TÍTULO: OPTIMIZACIÓN DE LA FLOTA DE VEHÍCULOS DE SERVICIO ENTRE TERMINALES DE CONSOLIDACIÓN DE MERCANCÍAS CON ESTACIONALIDAD DE LA DEMANDA

AUTORA: MARÍA JOSÉ SUÁREZ MEJUTO

TUTORES: MIQUEL ÀNGEL ESTRADA ROMEU

RESUMEN

Las empresas de logística nacen de la necesidad por parte del sector secundario, de reducir el coste unitario de un producto con la intención de ofrecer un precio muy competitivo. Su objetivo es el de conseguir un coste unitario de transporte mínimo satisfaciendo todas las necesidades temporales y de calidad que exija la demanda. Es por lo tanto, un problema de optimización de una red de muchos orígenes a muchos destinos, en la cual es necesario el dimensionamiento de la flota al igual que la organización de las rutas.

El presente estudio pretende en primer lugar, ofrecer una mejora del servicio de paquetería a partir del análisis de la estacionalidad de la demanda y en segundo lugar ofrecer un modelo de inventario y optimizar las rutas de la flota que compone el sistema. Con ello se conseguirán unos modelos que permitan la predecibilidad del comportamiento de las terminales y la reducción de costes, por movimientos en vacío, a la empresa.

Para ello se ha analizado el comportamiento de un sistema de terminales de forma probabilística, analizando cuáles son los factores sobre los cuáles se puede incidir para reducir la aleatoriedad de la red. En concreto se ha considerado el tiempo y el espacio como principales efectos atenuantes de dicha aleatoriedad y se han diseñado modelos en función de ello. Además se han considerado los distintos métodos existentes para la optimización de la red y se ha aplicado uno de ellos, en concreto el algoritmo de búsqueda tabú a un caso real de terminales, primero para la optimización de un día de servicio y para dos días distintos. A continuación se ha evaluado la eficacia de su implementación al problema tanto a casos reales como a teóricos.

Se ha trabajado en la mayor parte de la tesina con datos reales, y finalmente se ha servido de una batería de problemas teóricos para analizar los beneficios de la aplicación del algoritmo.

Al final del estudio se ha demostrado que una buena optimización de la red de transporte puede llevar a reducciones de costes muy significativas para la empresa, quedando patente así el interés que ello puede producir a las empresas de logística.

TITLE: OPTIMIZATION OF THE FLEET OF SERVICE VEHICLES BETWEEN MERCHANDISE CONSOLIDATION TERMINALS FOR SEASONAL DEMAND

AUTHOR: MARÍA JOSÉ SUÁREZ MEJUTO

TUTOR: MIQUEL ÀNGEL ESTRADA ROMEU

SUMMARY

Logistics companies are born out of the secondary sector's need to reduce the unit cost of a product, with the intention of offering a more competitive price. Their objective is to obtain a minimal transport unit cost whilst meeting all the temporal needs and the quality required by the demand. It is therefore a problem of optimizing a network consisting of multiple origins and destinations, in which both accurate sizing of the fleet and organization of the routes are needed.

This study has two aims, firstly to provide improvements to the courier service by analyzing the seasonality of demand and secondly to present an inventory model and to optimize the fleet routes which make up the system. The goal of this is to obtain models which will enable the company to determine the predictability of the terminals' performance and to reduce costs by reducing the empty running of vehicles.

In order to do this the behaviour of the terminal system has been analyzed probabilistically so as to determine the factors which can be altered to reduce the aleatory nature of the network. Time and space have been considered as the main extenuating factors behind the aforementioned aleatory nature and the models have been designed based on this. Furthermore various existing methods have also been taken into consideration for optimizing the network and one of these methods, the tabu search algorithm has been applied to a real-life case of the terminals, first for the optimization of one day's service and then for two separate days. In continuation the efficiency of its implementation has been evaluated, both in terms of real-life cases and theoretical situations.

This work carried out for this dissertation has been based primarily on real data but a set of theoretical problems has also been used in order to analyze the benefits of applying the algorithm.

Finally this study has demonstrated that a good optimization of the transport network can lead to significant cost reductions for the company and therefore it is clear that this is of interest to logistics companies.

INDICE

Página

Capítulo 1.- INTRODUCCIÓN **1**

INTRODUCCIÓN Y OBJETIVOS DE LA TESINA 5
ESTRUCTURA DE LA TESINA 6

Capítulo 2.- ESTADO DEL ARTE Y FORMULACIÓN DEL PROBLEMA **8**

2.1 ESTUDIO DE LA ESTACIONALIDAD DE LOS SISTEMAS DE TERMINALES 9
2.2 PREVISIÓN ANALÍTICA DE LOS MOVIMIENTOS EN VACÍO ENTRE TERMINALES 22
2.3 OPTIMIZACIÓN DE LAS REDES DE DISTRIBUCIÓN 27

Capítulo 3.- METODOLOGÍA **32**

3.1 ZONA DE ESTUDIO 32
3.2 HERRAMIENTAS DE CÁLCULO 33
3.2.1 Agrupación de las terminales en función de su comportamiento respecto a unas variables determinadas 33
3.2.2 Análisis estacional del sistema de terminales 39

3.2.3 Optimización de la red de distribución 41
3.3 BATERÍA DE PROBLEMAS 58

Capítulo 4.- ANÁLISIS DE RESULTADOS 62

4.1 AGRUPACIÓN DE LAS TERMINALES EN FUNCIÓN DE SU COMPORTAMIENTO RESPECTO A UNAS VARIABLES DETERMINADAS 62
4.2 ANÁLISIS DE LA ESTACIONALIDAD DEL SISTEMA DE TERMINALES. 80
4.3 OPTIMIZACIÓN DE LA RED DE DISTRIBUCIÓN 93

Capítulo 5.- CONCLUSIONES Y LÍNEAS DE INVESTIGACIÓN FUTURA 97

REFERENCIAS BIBLIOGRÁFICAS 100

ANEXOS 102

ANEXO 1.- DENDOGRAMAS SEGÚN LOS DISTINTOS MÉTODOS A-1
ANEXO 2.- TABLAS Y ESTADÍSTICOS A-8
ANEXO 3.- ANÁLISIS DE BONDAD DE AJUSTE A-12
ANEXO 4.- DIAGRAMAS BOX-PLOT A-14

CAPÍTULO 1

INTRODUCCIÓN Y OBJETIVOS DE LA TESINA

La logística estudia, analiza y optimiza los flujos de materiales y de información de toda la cadena de producción, desde el proveedor hasta el cliente. Esta definición, representa todo el engranaje que existe desde los inventarios, pasando por la producción, el empaquetamiento, el transporte hasta la distribución física del producto.

Por eso hoy la logística se configura como un área estratégica de las empresas y no únicamente como un eslabón operativo. Una buena logística reduce los costes de las empresas en un porcentaje muy importante. Según ITENE (Instituto Tecnológico del Embalaje, Transporte y Logística, 2004) entre el 7% y el 30% del coste de un producto, es coste logístico, e incluso puede llegar a un 60% como es el caso de los productos lácteos y derivados. Por esa razón se convierte en objetivo principal para muchas empresas reducir esos costes.

Para poder evaluar el coste unitario del transporte de mercancía es necesario hacer un estudio detallado de los flujos de la empresa para después optimizar costes. Ello comporta un análisis de cada una de las terminales, de los flujos entre ellas y de la variabilidad de estos.

Dentro de los distintos sectores del transporte de mercancías se estudiará el de paquetería por carretera. Según el Libro Blanco de Transporte de la comisión Europea (2006), en Europa el transporte terrestre de mercancías por carretera representa el 44% de los modos de transporte frente al 8% que representa el ferrocarril o el 4% que representan las vías navegables. El transporte por carretera presenta la ventaja de ofrecer un servicio puerta a puerta de manera que resulte mucho más sencillo y cómodo para los contratantes, por otra parte la flexibilidad del recorrido presenta ventajas frente a posibles reestructuraciones de las empresas, cambios de rutas, u otras necesidades. En lo referente a su coste total, éste resulta notablemente inferior, y si el volumen de la mercancía es suficientemente grande resulta posible para un cliente utilizar el camión de forma exclusiva. El barco por su parte, no presenta estas facilidades, pero por otra parte, al ser el volumen total transportado mayor reduce el coste unitario del producto, al igual que el tren.

En este caso, se estudiará esencialmente el transporte por carretera, y sólo en algunos casos en los que se indicará previamente se hará mención, o incluso se llegarán a estudiar, algunos casos de trasporte marítimo.

Por lo que respecta a las empresas de transporte, se puede hacer una diferenciación en dos tipos de empresas, las empresas que disponen de una flota propia y empresas que tienen parte de la flota subcontratada. Antiguamente las empresas transportistas tenían tendencia a disponer de su propia flota, sin embargo, en los últimos tiempos se ha tendido a la subcontratación. Esta evolución resulta más marcada en empresas cuya tarea principal no es el transporte sino la producción. Los textos económicos sobre este tema se refieren a dicha evolución como desintegración vertical del transporte. Según Compés (2005) los motivos económicos por los que se ha producido son, por una parte, las restricciones legales al transporte por cuenta propia, unidas a la liberalización progresiva del transporte por cuenta ajena, por otra, la necesidad de reducir el coste unitario del producto. Para conseguirlo se puede, o bien reducir el coste de fabricación del producto o bien el coste logístico. En esta tesina se estudiará la segunda tipología.
Las razones por las cuales el coste unitario del transporte es superior son:

a) la lejanía de los consignatarios respecto a los expedidores y el consecuente coste de encontrar cargas de retorno propias o de empresas del grupo

b) la gran cantidad de consignatarios, lejanos los unos de los otros y que reciben cargas inferiores a la capacidad total del transporte

c) las restricciones temporales que exigen los consignatarios.

Estos costes asumidos por una empresa pequeña con un volumen pequeño de transacciones es mucho mayor que para una empresa que mueve grandes volúmenes. Eso es debido a que en grandes empresas se balancean los desequilibrios debidos a la aleatoriedad del sistema. De ahí nace la necesidad de la subcontratación por parte de las pequeñas empresas. Según ITENE el nivel de subcontratación de las distintas actividades logísticas por parte de las empresas es el siguiente:

- Transporte 57%
- Almacenaje 40%
- Preparación de pedidos 20%
- Reposición de lineales 15%
- Acciones de venta (captura y gestión de pedidos) y facturación 7%

Cerca de un 15% contrata una gestión logística integral, siendo ésta la tendencia que más se está extendiendo, especialmente entre grandes empresas de fabricación y distribución.

Cabe destacar también que lo más habitual es que las empresas subcontraten a más de un operador logístico y las razones que les motiva a utilizar esta forma de subcontratación son, entre otras, la especialización de operador en rutas y destinos (relevancia de casi el 60%) y el ahorro de costes (importancia superior al 40%).

Las empresas logísticas nacen de la necesidad de reducir el coste final de un producto interviniendo sobre el coste de distribución. La causa principal del elevado coste del problema es la aleatoriedad del sistema, que resulta ser una de las características principales de las redes de distribución de paquetería, eso hace que la predicibilidad, la programación y diseño de los inventarios resulte un problema. Dicha aleatoriedad se transforma en gastos para la empresa en forma de camiones parados o recorridos a carga inferior a la máxima posible. Las consecuencias de esa aleatoriedad pueden estudiarse

mediante el análisis estacional del sistema de terminales que conforma la red. Fundamentalmente la aleatoriedad inherente a un sistema está mitigada por flujos de carga entre grupos de terminales y por flujos de carga sobre un período de tiempo extenso. Por lo tanto, es menester conocer cómo esas variables alteran el problema.

Por otra parte, para poder reducir los costes, es necesaria la optimización del sisema. Ésta ha sido uno de los aspectos del transporte más estudiados hasta el momento. Existen diversos métodos para la resolución del sistema. Si bien la obtención de la solución óptima absoluta requiere un tiempo computacional demasiado elevado, existen algoritmos que proporcionan una solución muy eficiente al problema. En Estrada (2007) el problema de optimización combinatoria se resuelve eficazmente mediante un algoritmo de Búsqueda Tabú, consiguiendo una reducción del coste del sistema en relación a otros procedimientos heurísticos del 7% en problemas de carga completa (*Full Truck Load, FTL*) y superior al 12% en problemas a carga fraccionada (*Less than Truck Load, LTL*). La diferenciación entre ambas tipologías consiste en que en los *FTL* el servicio es de puerta a puerta, la planificación temporal del envío se adapta a un único cliente con las restricciones temporales del proveedor. Se basa en una gestión dinámica de los recursos debido a la variabilidad temporal de la demanda y requiere el estudio de los movimientos en vacío. En *LTL* el tamaño del envío de un solo cliente es muy inferior a la capacidad del vehículo de transporte. Se utiliza un servicio de transporte consolidado con lo cual las empresas requieren instalaciones propias de consolidación (*HUBS*). En este caso los horarios de envío son fijos y se deben ceñir a un conjunto de terminales o proveedores. Estos dos problemas son tratados mediante algoritmos distintos ya que las técnicas o posibilidades de optimización de *FTL* son muy diversas a las técnicas de *LTL*. En las primeras resulta un algoritmo en el cual tan solo se debe encontrar el camino más corto o alargamientos de rutas y en el segundo se pueden seguir una serie de estrayegias como son la de *Hub and Spoke* o bien la de paradas múltiples o *peddling* para la consolidación de la carga.

Por lo tanto la empresa de paquetería necesita establecer una red con servicios a carga parcial *LTL* y una red con servicios a carga total *FTL*. La red total consistirá en la unión de dos redes operativas, una red capilar y una red troncal. La red capilar dispondrá de un mayor número de vehículos de capacidad menor, deberá tener en cuenta los efectos de la congestión en la planificación de rutas y los efectos aleatorios serán más

significativos, (aumentará en caso de que la empresa ofrezca sus servicios a clientes particulares en vez de a empresas). La red troncal por su parte dispondrá de un número de vehículos limitado, la demanda será variable a escala mensual, tendrá operaciones de cross-docking frecuentes y realizará operaciones de paradas múltiples muy ocasionalmente.

Se dispone de diversas estrategias para organizar la distribución de manera que sea más eficiente y flexible. Las tres estrategias principales son la estrategia de envíos directos (*many-to-many*) que es aplicable exclusivamente a pares origen-destino con un flujo asociado significativamente alto; la estrategia de los envíos *hub&spoke* la cual consiste en el envío a una terminal *hub* donde se consolida la carga y desde allí se distribuye. Por último, la estrategia de envíos con paradas múltiples (*peddling*) en cuyo caso la consolidación se realiza en el mismo vehículo.

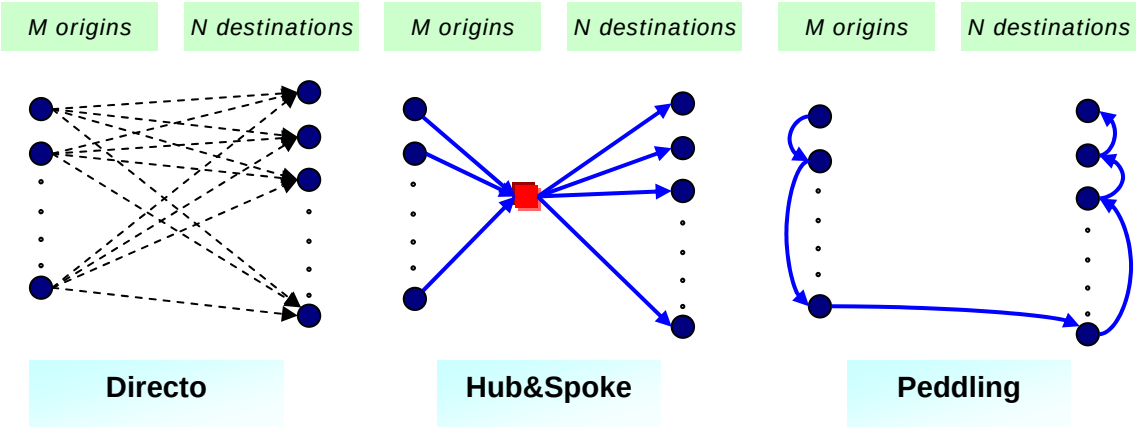


Fig. 1.1. - *Esquema de las estrategias principales ara una red de distribución de mercancías.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

1.1 OBJETIVOS

Se desea mejorar un sistema de terminales de mercancías, para ello se buscarán unos modelos probabilísticos que permitan la predicción de la variabilidad del sistema como principal valor de la aleatoriedad del problema y en segundo lugar plantear una posible solución para la optimización de la flota y de su uso. Será menester, por lo tanto para cumplir dichos objetivos, unos análisis previos. En primer lugar es necesario conocer

como datos de partida para la tesina, las características de las terminales que componen el sistema de terminales. Se dispone de los datos suministrados por una empresa nacional líder en el sector logístico, que ha proporcionado los flujos de carga entre las distintas terminales. A partir de un análisis cluster, análisis de conglomerados, se analizará la similitud de comportamiento entre las distintas terminales del sistema de forma estadística. Para ello se utilizará el programa SPSS que como resultado del análisis proporcionará una agrupación de las terminales según su comportamiento. Para corroborar que dicha clasificación es fiable se hará un análisis discriminante. Este segundo análisis dirá si existe alguna terminal mal clasificada y cuál es.

En segundo lugar se analizará la variabilidad del comportamiento de las terminales, tanto para las terminales individualmente como para grupos de terminales, se buscará las distribuciones de probabilidad acumulada de dicha variabilidad, se analizarán los desequilibrios de flujo y por último el exceso de movimientos en vacío. Se buscará el modelo de comportamiento, tanto de la variabilidad de las terminales como de sus flujos.

Finalmente en la última parte de la tesina se desarrollará un modelo para la optimización de los movimientos dentro de la red, que proporcionará los recorridos óptimos en función de la demanda diaria.

1.2 ESTRUCTURA DE LA TESINA

La tesina esta dividida en 5 apartados: una primera introducción, la segunda referente al estado del arte y la formulación, la tercera sobre la metodología llevada a cabo, la cuarta referente al análisis de resultados obtenidos y finalmente la quinta, que serán las conclusiones.

En la introducción se hace una breve explicación de las razones por las cuales se ha llevado a cabo el estudio, se describen los rasgos principales de las redes de transporte entre muchos orígenes y muchos destinos, se citan los objetivos a los cuales pretende llegar la tesina y finalmente se describe la estructura de la tesina de forma breve para que el lector pueda entender la forma.

El segundo, tercer y cuarto apartado tienen una estructura muy similar puesto que están subdivididos respectivamente en tres partes referentes a las tres áreas de estudio. Que son la agrupación de terminales, el análisis de estacionalidad de una red de terminales en función de diversas consideraciones temporales y espaciales y, por último, la optimización de la red para la organización de rutas en un solo día y en varios días con demanda variable.

El estado del arte se presenta en el segundo apartado y explicará para cada una de las tres áreas citadas, la situación en la que se encuentran los estudios que se han hecho hasta el momento y explicará los conceptos teóricos relevantes para poder escribir y entender la presente tesina.

Los métodos y herramientas utilizados para hacer el estudio se presentarán en el tercer apartado, así como los programas informáticos y las posibilidades que éstos han ofrecido. Por otra parte presenta los datos de los que se ha partido, los trata y justifica las hipótesis y teorías utilizadas en cada caso.

El cuarto apartado presenta los resultados obtenidos, los discute y los compara con los de otros estudios similares.

Por último, el quinto apartado incluirá las conclusiones más significativas derivadas de la tesina así como las líneas de investigación futura.

CAPITULO 2

ESTADO DEL ARTE Y FORMULACIÓN DEL PROBLEMA

La optimización de las flotas de camiones ha sido un tema principal de investigación en lo que a líneas de investigación en organización de recursos respecta. En concreto, lo ha sido el problema de los movimientos en vacío (*backhauling*). Las primeras formulaciones sobre el tema están editadas por Koopmans en "Transportation Problem" (Koopmans, 1949) y a partir de él ha habido varios autores que han escrito sobre el problema dinámico con algunas excepciones como Crainic y Delorme (1993). Los modelos estáticos proporcionan una solución óptima para un problema ideal estacionario. Y los modelos dinámicos dividen el problema en un sistema de intervalos temporales en los que cada paso esta asociado a un intervalo de tiempo distinto. En cualquier caso, el elemento estocástico más importante en el estudio de paquetería es la demanda de equipamiento, (se puede estar hablando tanto de camiones como vagones o barcos). La naturaleza estocástica de la variable está incorporada en el modelo como el valor esperado y su varianza entre unos orígenes y destinos dados. Esta naturaleza estocástica se refleja físicamente en el número de vehículos disponibles en un determinado punto de la red y en un determinado instante. Ésta es la clave en el problema de recolocación de vehículos, ya que los sobrantes se utilizarán para cubrir las necesidades de otros lugares en los que haya déficit.

En Hall (1998) se estudia la estacionalidad y el balance de equipamiento en exceso y en defecto de un sistema de terminales de forma empírica. Para ello se estudia como los efectos de la aleatoriedad son mitigados mediante la agrupación de las terminales y a lo largo del tiempo. Para ello se utiliza un sistema de agrupación de terminales. En el caso de Hall la agrupación de terminales se hace de forma geográfica y en esta tesina se hará una agrupación por comportamiento. Se comparará después con el comportamiento de un sistema de terminales individuales. El propósito es llegar a un modelo que permita predecir la variabilidad de una terminal en función de su media.

2.1 ESTUDIO DE LA ESTACIONALIDAD DE LOS SISTEMAS DE TERMINALES.

Existen distintos métodos dentro del análisis por conglomerados que permitan ejecutar el mencionado análisis y un punto fundamental en su aplicación será escoger el más adecuado, decisión que se tendrá que basar en la experiencia previa sobre trabajos con datos análogos y en toda la información teórica que se disponga de ellos. En función de cual se escoja, los resultados pueden variar sensiblemente resultando, no agrupaciones contradictorias, pero si clasificaciones que pueden llegar a conclusiones diferentes. La aceptación de unos resultados en vez de otros estará condicionada por múltiples factores entre los cuales se destaca el objetivo de la clasificación que se está realizando.

Podemos encontrarnos dos tipos fundamentales de métodos de clasificación: jerárquicos y no jerárquicos. En los primeros, la clasificación resultante tiene un número creciente de clases anidadas mientras que en el segundo las clases no son anidadas.

Las técnicas no jerárquicas comprenden muchas variantes (grupos disjuntos, particiones estocásticas, métodos de optimización,...) pero todas se caracterizan por tener como objetivo realizar una sola partición de los individuos en k grupos, lo que implica que se tiene que fijar este número previamente. El procedimiento más común de estas técnicas es el K-medias, es decir esa en la que a cada caso se le asigna a un cluster en base a que su distancia respecto al centro del mismo sea mínima. Este procedimiento es conveniente cuando los datos a clasificar son muchos o cuando se quiere refinar una clasificación obtenida utilizando una técnica jerárquica.

Otra técnica comparable con las no jerárquicas, pero que sin embargo por no ser definitoria sino explicativa no puede ser incluida en la clasificación anterior es el análisis discriminante. Aunque su utilidad es similar al de las k-medias no está definida como técnica aglomerativa. Ésta ayuda a identificar las características que diferencian (discriminante) a dos o más grupos y a crear una función capaz de distinguir con la mayor precisión posible a los miembros de uno u otro grupo. Se trata de una técnica de análisis multivariante que es capaz de aprovechar las relaciones existentes entre una gran cantidad de variables independientes para maximizar la capacidad de discriminación. El discriminante parte de un número preestablecido de grupos y estudia a partir de métodos probabilísticos la probabilidad de pertenencia de un caso a un grupo en concreto.

Y por último el análisis factorial que es una técnica que consiste en resumir la información contenida en una matriz de datos con V variables. Para ello se identifican un número reducido de factores F, siendo el número de factores menor que el número de variables. Los factores representan a las variables originales, con una pérdida mínima de información. Se puede distinguir entre análisis factorial exploratorio, donde no se conocen los factores a priori, sino que se determinan mediante el mismo análisis, y por otro lado el análisis confirmatorio donde se propone un modelo, según el cual hay unos factores que representan a las variables originales, siendo V superior a F. En este caso, somete el modelo establecido a priori a una comprobación.

Las técnicas jerárquicas reciben su nombre de la estructura inclusiva o divisiva de los grupos que se van formando, éstas permiten utilizar la representación gráfica más utilizada en estos casos, el dendograma, el cual describe con detalle todo el proceso y por la cual cosa resulta ser una herramienta fundamental para estudiar los resultados. Más adelante se explicará su interpretación.

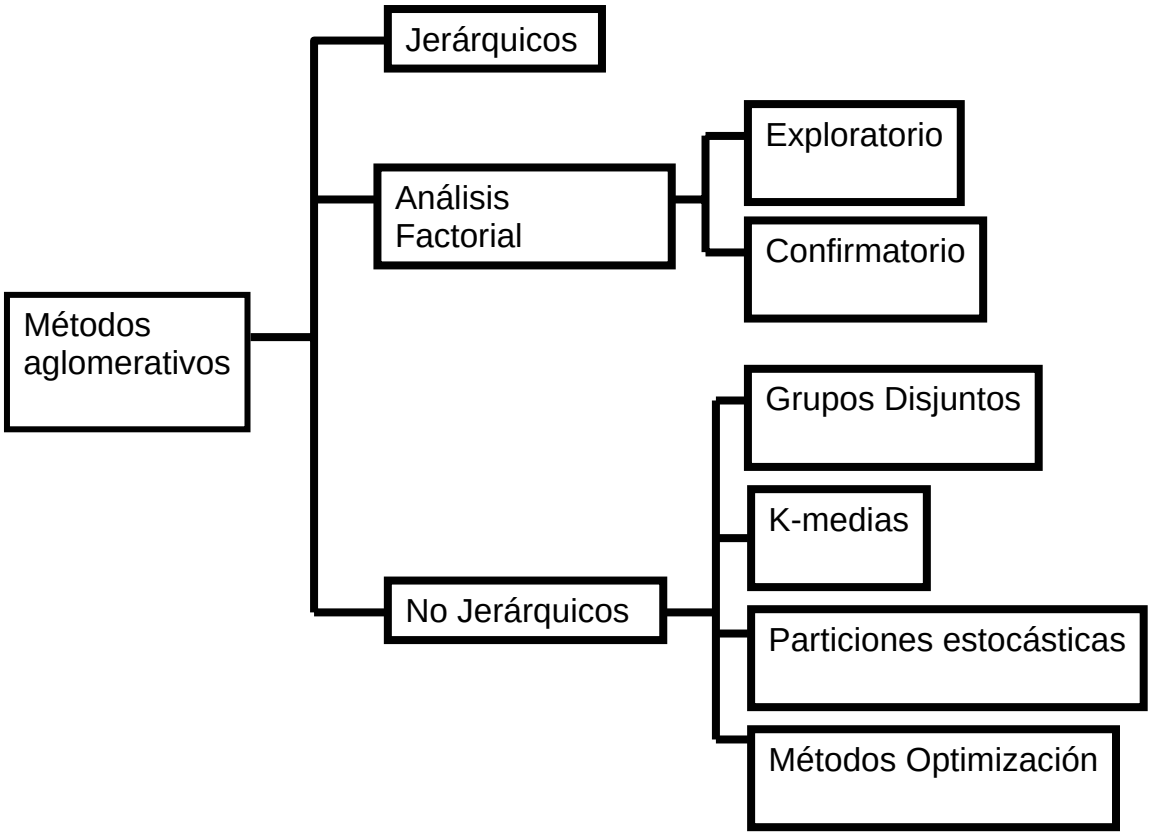


Fig. 2.1. *Métodos de clasificación multivariante.*

El Análisis Clusters (o Análisis de conglomerados) es una técnica de Análisis Exploratorio de Datos para resolver problemas de clasificación. Su objetivo consiste en ordenar el conjunto de casos o de variables en grupos (conglomerados o clusters) con el objetivo de que, por un lado, los objetos pertenecientes a un mismo grupo sean muy semejantes entre sí y, por el otro, los objetos pertenecientes a grupos diferentes tengan un comportamiento de disimilitud con respecto a las variables analizadas. En el caso de que se agrupen variables obtendremos el resultado de un análisis factorial y en el caso de que agrupemos casos obtendremos un análisis de conglomerados normal. El análisis cluster es un método que permite descubrir asociaciones y estructuras en los datos que no son evidentes a priori pero que pueden ser útiles una vez que se han encontrado. Los resultados de un Análisis Cluster pueden contribuir a la definición formal de un esquema de clasificación tal como una taxonomía para un conjunto de objetos, a sugerir modelos estadísticos para describir poblaciones, a asignar nuevos individuos a las clases para diagnóstico e identificación, etc.

Ésta es una técnica exploratoria puesto que no utiliza ningún tipo de modelo estadístico para llevar a cabo el proceso de clasificación. Los métodos pueden dividirse en aglomerativos y divisivos en función de si se parte de que cada caso es un conglomerado y se van produciendo sucesivas uniones o si se parte de un único cluster y se van dividiendo en función de las disimilitudes.

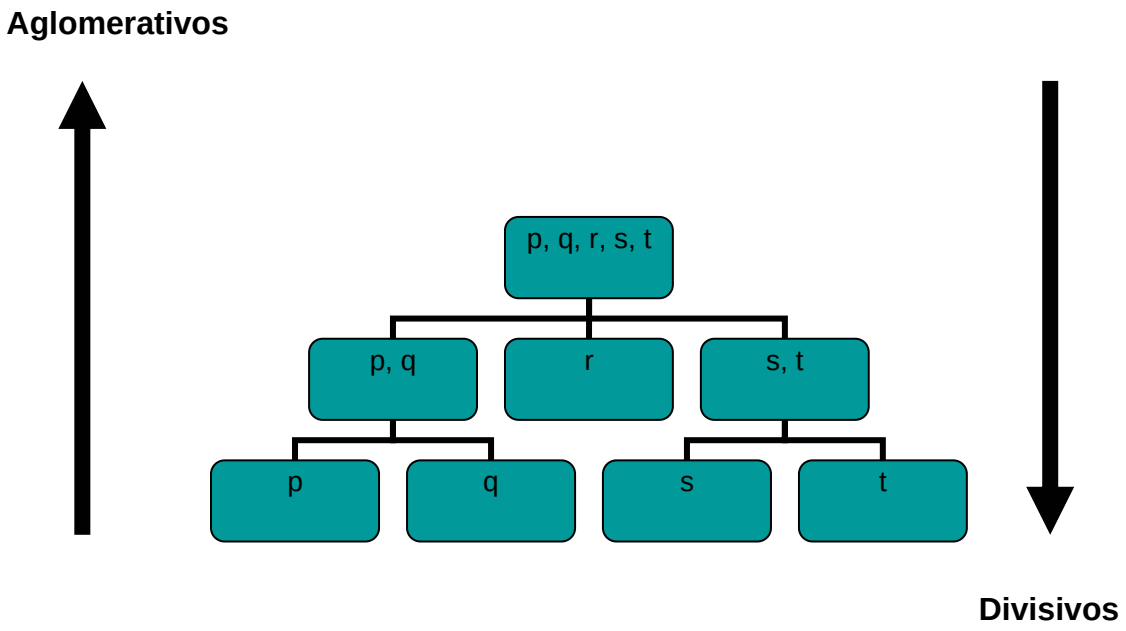


Fig. 2.2.- *Diferencia de formación entre los métodos aglomerativos y divisivos.*

En la figura 2.1 se puede observar los dos tipos de agrupaciones jerárquicas donde *p, q, r, s* y *t* podrían ser o bien casos o bien variables o bien grupos de casos o variables según sea el problema en cuestión.

A continuación se presentan en detalle las técnicas jerárquicas aglomerativas según las indicaciones de Ramirez y Pawlowsky (1998) y del trabajo mas enfocado hacia el SPSS de Visauta (1998). El objetivo en común de estas técnicas es el de obtener una agrupación de casos que sea coherente y capaz de mostrar las diferencias respecto a una variable de los distintos elementos que lo componen.

Se procederá a plantear las características de las distintas medidas de distancia y los distintos métodos y se describirán sus principales diferencias pero antes de ello se debe constatar que dichas medidas dependen del tipo de datos con los que se esté trabajando,

así en el caso de que trabajemos con datos binarios, frecuencias o intervalos se utilizaran medidas distintas. Se debe destacar que en el caso de estudio, los datos son de intervalo y se presentaran, tan solo, las medidas específicas para ellos.

MEDIDAS DE DISTANCIAS

- *Distancia euclídea:* es la raíz cuadrada de la suma de las diferencias al cuadrado entre los dos elementos en las variables consideradas. Es la más habitual de las distancias y cumple las propiedades de distancia métrica. Además es invariante para traslaciones.

$$D(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \tag{2.1}$$

El principal problema que presenta es que tiene una gran dependencia respecto de las escalas de las variables por lo que es conveniente transformarlas previamente para hacerlas independientes de las unidades de medida. Una solución podrían ser estandarizarlas y otra realizar una canonización, con lo que se igualarían los recorridos de cada una de las variables al mismo intervalo (0,1).

- *Distancia euclídea al cuadrado:* directamente es la suma de las diferencias al cuadrado entre los dos elementos en las variables consideradas. Cumple las mismas características que la distancia anterior y, por lo tanto, también es necesario transformar los datos antes de calcular las interdistancias. Es la distancia recomendada para ciertos métodos de clusterización, como el del Centroide o el Ward.

$$D^2(x,y) = \sum_i (x_i - y_i)^2 \tag{2.2}$$

- *Distancia métrica de Chebychev:* es la diferencia máxima en valor absoluto entre los valores de los elementos.

$$C(x,y) = MAX_i \left| x_i - y_i \right| \tag{2.3}$$

- *Distancia de Minkowsky:* Se recogen toda una serie de familias de distancias definidas por la raíz p de la suma de las diferencias en valores absolutos elevadas a p de los valores de los elementos. Son distancias métricas invariantes para traslaciones pero no para cambios de escala por la cual cosa las ideas de transformación de los datos comentados por las distancias euclídea no son extensibles a esta familia.

$$M(x,y) = \left(\sum_i |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{2.4}$$

Para $p = 2$ se obtiene la distancia euclídea y para $p = 1$ la distancia de Manhattan o City-Block.

- *Distancia de Mahalanobis*: esta distancia presenta un enfoque menos geométrico del concepto de distancia, con el que se pretende incluir las covarianzas entre las variables a la hora de evaluar su semejanza. Se define como cuadrado de la siguiente función.

$$Mah(x,y) = (x-y)^t \boldsymbol{S}^{-1} (x-y) \tag{2.5}$$

Donde $\boldsymbol{S}^{-1}$ es la inversa de la matriz de covarianzas empíricas. Esta distancia es invariante a cambios de escala.

TIPOS DE MÉTODOS

Para definir una técnica de análisis cluster es necesario elegir una distancia entre elementos de la muestra y, como se había comentado, también un criterio que permita establecer los grados de semejanza entre los grupos. Este criterio es el llamado método, del cual existen distintos tipos, que son los siguientes:

- *Método del Single Linkage:* también llamado método de vecinos más próximos o de la distancia mínima. Se establece la distancia entre dos grupos con la mínima distancia entre los individuos de cada uno de ellos. Es decir, para evaluar la distancia

entre un grupo y otro se cuantificará la distancia de cada elemento del primero con cada elemento del segundo y la menor de todas ellas será la distancia entre grupos.

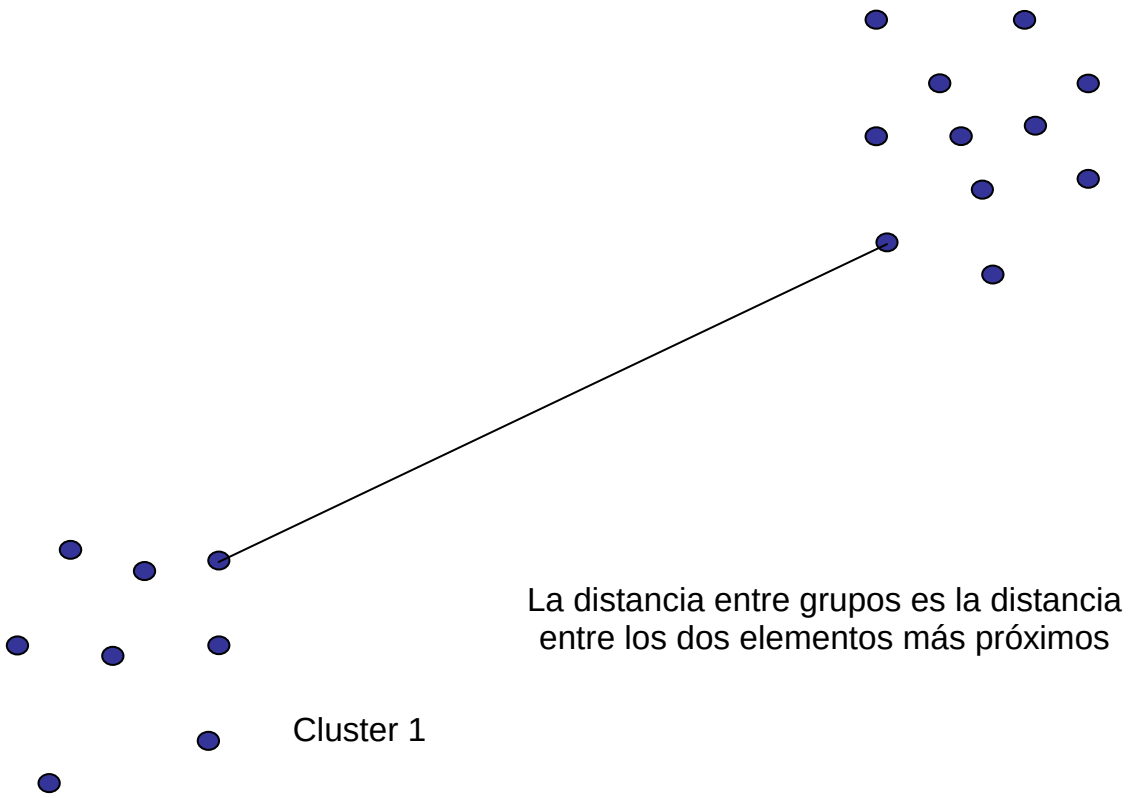


Fig. 2.3. *Metodo del Single Linkage.*

La característica más destacable de este método es que tiene tendencia a formar grupos alargados. Eso implica que, si los grupos son cercanos los unos de los otros en el espacio de las variables, es habitual la formación de encadenamientos artificiales entre ellas mediante valores puente intermedios. No es un método útil para resumir datos sin embargo es útil para detectar *outliers* (estarán entre los últimos en unirse a la jerarquía). Pueden usarse medidas de la similitud o de la disimilitud y tiende a construir clusters demasiado grandes y sin sentido. Es invariante bajo transformaciones monótonas de la matriz de distancias.

- *Método del Complete Linkage:* también llamado método de vecinos más lejanos o de la distancia máxima. Este caso es muy similar al anterior ya que se establece la distancia entre grupos a partir de la distancia entre dos elementos concretos, uno para cada conjunto. La diferencia está en que ahora son los elementos más lejanos los que cumplen esta función representativa.

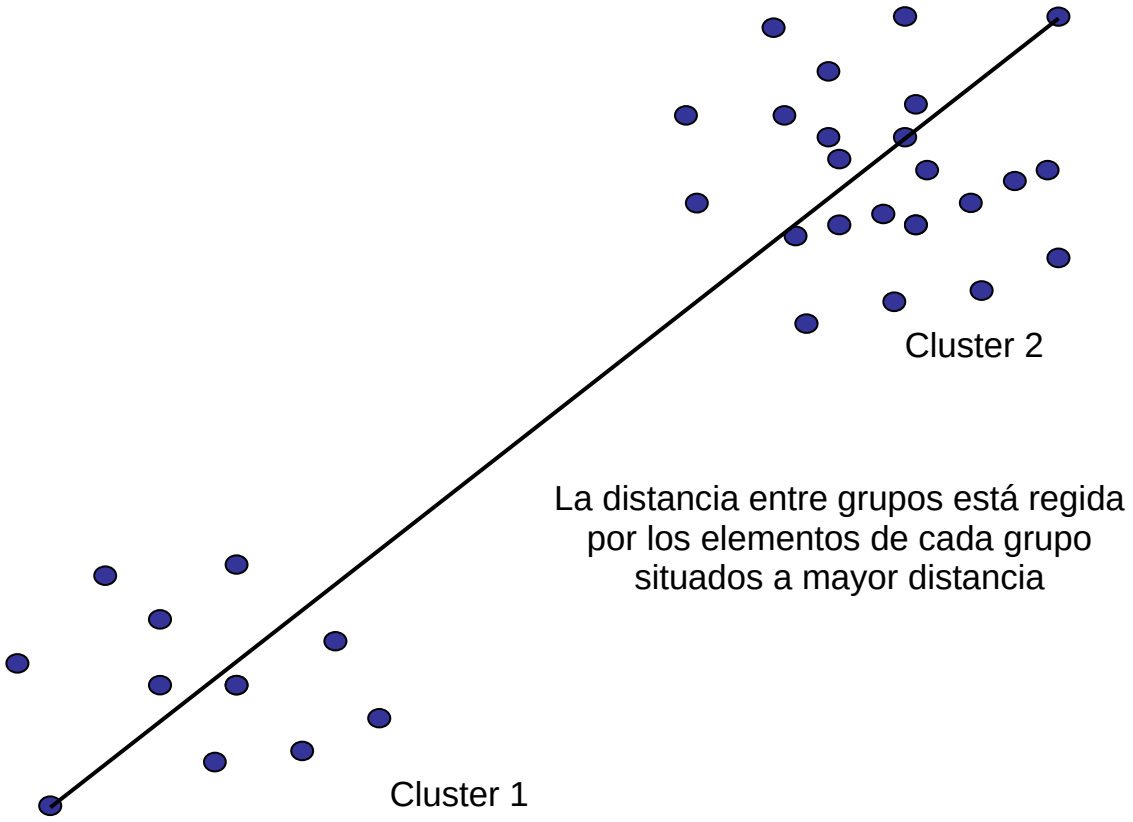


Fig. 2.4. *Método del Complete Linkage.*

Este método tiende a formar clusters pequeños y muy compactos. Presenta problemas en el caso que existan valores atípicos en la muestra sin embargo sirve para detectarlos. Se pueden utilizar medidas tanto de similitud como de disimilitud. Es invariante respecto a transformaciones monótonas de la matriz de distancias.

- *Método del Average Between-groups Linkage:* también llamado método de distancias promedio entre grupos. Este método y el siguiente son opciones intermedias entre los dos anteriores. En este caso se define la distancia entre dos grupos como la media de las distancias entre todas las combinaciones posibles dos a dos de los elementos de uno y otro grupo. Utiliza, por lo tanto, todos los pares de distancias y no solo los elementos más próximos o más alejados como los métodos anteriores.

- *Método del Average Within-groups Linkage:* también conocido como método de distancias promedio intragrupos. Se combinan los grupos de manera que la media de las distancias entre todos los pares de sujetos dentro del cluster resultante sea la menor

posible. Por tanto, la distancia entre dos grupos se toma como promedio de las distancias entre todos los posibles pares de casos dentro del grupo en cuestión.

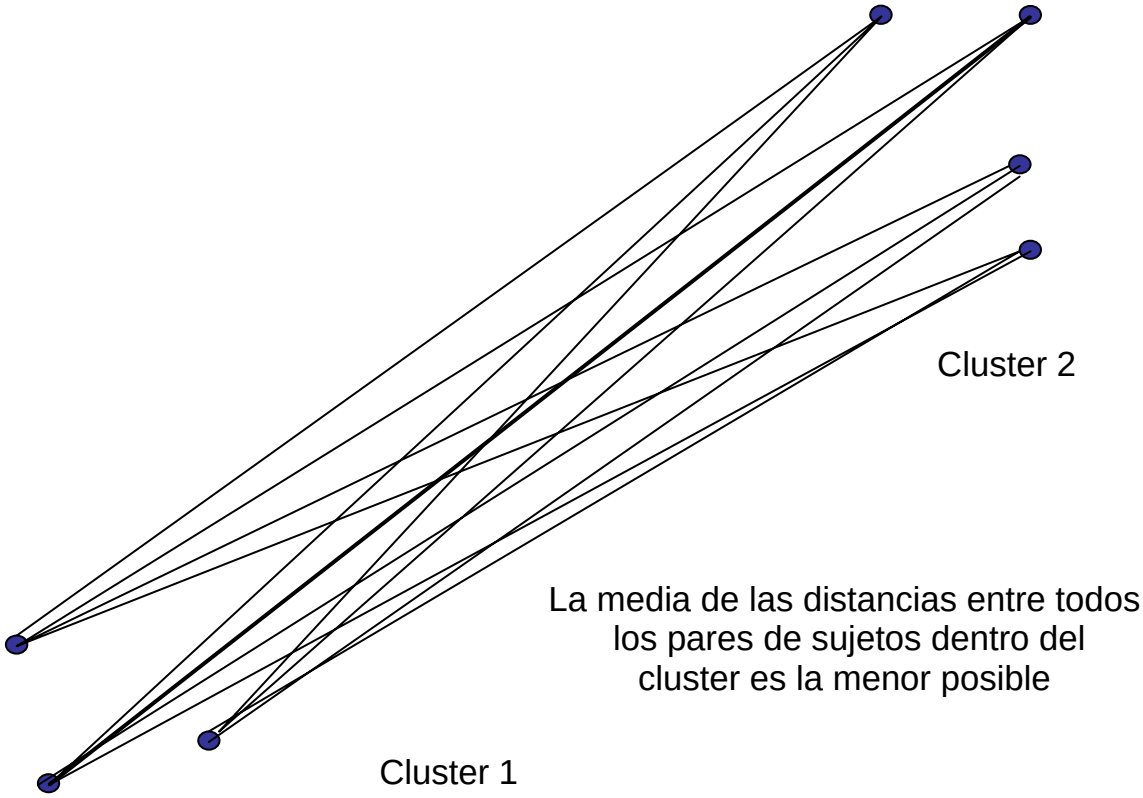


Fig. 2.5. *Método del Average Within-groups Linkage.*

- *Método del Centroide:* la distancia entre dos grupos será la existente entre sus respectivos centroides, es decir, la distancia entre las medias de cada grupo para todas las variables.

Un inconveniente que se puede presentar con el uso de este método es que las distancias entre grupos pueden disminuir de un paso al otro. Es lo que se conoce como inversión y se debe a la reubicación del centroide en cada iteración, con la cual cosa aumenta la dificultad a la hora de interpretar los resultados. Para conocer más detalles de este fenómeno, se puede consultar el trabajo de Everitt (1977). Es necesario para este método que las variables estén en escala de intervalo. Las distancias entre grupos se calculan como las distancias entre los vectores medios. Si los tamaños de los grupos a mezclar son muy diferentes, entonces el centroide del nuevo grupo será muy próximo al de mayor tamaño incluso podría llegar a estar dentro de este grupo.

- *Método de la Mediana:* en el método anterior del centroide el centro del grupo resultante es una media ponderada de los centroides de los clusters individuales, y los pesos son proporcionales a la medida de éstos. En el método de la mediana el centroide del grupo resultante se calcula por simple promedio de los dos, independientemente del número de elementos de cada uno de ellos.

Proporciona clusters ni demasiado grandes ni demasiado pequeños. Con este método se pueden utilizar medidas de similitud o de disimilitud. No es invariante por transformaciones monótonas de las distancias y proporciona una buena representación gráfica de los resultados

- *Método de Ward:* hasta ahora, los métodos presentados han tenido un planteamiento puramente geométrico, pero este introduce una nueva línea de actuación ya que trata de minimizar la pérdida de homogeneidad que supone fusionar dos grupos en cada iteración. De hecho, todo análisis cluster mira de formar grupos con cohesión interna y que sean externamente aislados pero la particularidad de este método es que utiliza directamente un planteamiento en este sentido.

El método busca minimizar $\sum_r SSW_r$ que corresponde a las sumas de cuadrados intragrupos para cada grupo r, que viene dada por la expresión (2.6):

$$SSW_r = \sum_{m=1}^{n_r} \sum_{j=1}^{p} (X_{rjm} - \overline{X}_{rj})^2 \tag{2.6}$$

donde X_{rjm} denota el valor de la variable Xj en el m-ésimo elemento del grupo *r*. En cada paso el algoritmo une los grupos *r* y *s* que minimizan:

$$SSW_t - SSW_r - SSW_s = \frac{n_r n_s}{n_r + n_s} d_{rs}^2 \tag{2.7}$$

con $t = r \cup s$ y d_{rs} la distancia entre los centroides de r y s.

Dicho de otra manera, el método de Ward lo que busca es calcular para todos los grupos la media de todas las variables. Posteriormente, para cada individuo, se calcula la distancia euclídea al cuadrado respecto esta media y se suma para todos ellos. En cada paso los clusters que se van formando son aquellos que resultan en un menor incremento de la suma global de distancias al cuadrado dentro del cluster. El método suele ser muy eficiente, tiende a crear clusters de tamaños pequeños y muy compactos. Se puede utilizar la matriz de distancias como una tabla de contingencia. Es invariante bajo transformaciones monótonas de la matriz de distancias y puede ser sensible a outliers. No suele generar valores puente.

En la práctica, resulta útil para ver la diferencia analítica entre los distintos métodos para medir distancias, el siguiente cuadro de resumen.

Si dos objetos o grupos *P* y *Q* se han agrupado, la distancia del grupo con otro objeto R puede calcularse como una función de las distancias entre los tres objetos o grupos de la forma siguiente:

$$d(R,P+Q)=\delta_1 d(R,P)+\delta_2 d(R,Q)+\delta_3 d(P,Q)+\delta_4\left|d(R,P)-d(R,Q\right| \qquad (2.8)$$

δ_j con $j=1,...,4$ son constantes de ponderación. Los distintos métodos dependen de los valores que demos a las ponderaciones. En la tabla siguiente se muestran los pesos para algunos de los métodos más comunes.

Tabla 2.1.- *Coeficientes de la ecuación según los distintos métodos*

Método	δ_1	δ_2	δ_3	δ_4
Salto mínimo	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Salto máximo	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
Media	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	0	0
Centroide	$\frac{n_P}{n_P+n_Q}$	$\frac{n_Q}{n_P+n_Q}$	$-\frac{n_P n_Q}{(n_P+n_Q)^2}$	0
Mediana	$\frac{1}{2}$	$\frac{1}{2}$	$-\frac{1}{4}$	0
Ward	$\frac{n_R+n_P}{n_R+n_P+n_Q}$	$\frac{n_R+n_Q}{n_R+n_P+n_Q}$	$-\frac{n_R}{n_R+n_P+n_Q}$	0
Método Flexible	$\frac{1-\beta}{2}$	$\frac{1-\beta}{2}$	β	0

Los parámetros n_R, n_P, n_Q denotan el número de objetos en cada uno de los grupos y β es un valor arbitrario entre 0 y 1.

Según la literatura hasta el momento Pawlowsky (1998), Figueras (2001) , el método de Ward parece ser el que mejores resultados tenga que dar para la búsqueda de la minimización de la variabilidad generada dentro de cada grupo, pero es necesario comprobarlo con los datos de este estudio en concreto para poderlo asegurar.

Los resultados del análisis vienen representados gráficamente en el dendograma que resume el proceso de agrupación de un análisis cluster. Los objetos similares se conectan mediante enlaces cuya posición en el diagrama está determinada por el nivel de similitud entre los casos.

Como se ha dicho anteriormente, una vez realizado el análisis jerárquico y en caso querer una solución contrastada, se puede optar por hacer un segundo análisis, en este caso partiendo de un número de grupos predeterminado. Los dos métodos más utilizados para este fin son el análisis discriminante y el análisis de las k-medias.

El análisis discriminante es una técnica de clasificación donde el objetivo es obtener una función capaz de clasificar a un nuevo individuo a partir del conocimiento de los valores de ciertas variables discriminadoras. A diferencia del A. Cluster, se deben conocer los grupos previamente y a qué grupo pertenecen ciertos individuos, de los que también se conoce sus valores en las variables discriminantes. La clasificación es el objetivo último de este análisis, de igual forma que sucedía con el análisis cluster, pero con la diferencia de que ahora la inclusión de nuevos elementos a un grupo se realiza mediante métodos probabilísticos.

El análisis discriminante tiene una doble finalidad, por una parte explicar la pertenencia de cada caso a uno u otro grupo en función de las variables de su perfil para comprobar su pertenencia o no al grupo preestablecido y cuantificar el peso de cada una de ellas en la discriminación. Por otra parte predecir a qué grupo más probable habrá de pertenecer un nuevo individuo del que únicamente se conoce su perfil de variables.

Hay dos posibles enfoques del análisis, uno basado en la obtención de funciones discriminantes de cálculo similar a las ecuaciones de regresión lineal múltiple que consiste en conseguir, a partir de las variables explicativas, unas funciones lineales de éstas con capacidad para clasificar otros individuos. A cada nuevo caso se le aplican dichas ecuaciones y la función de mayor valor define el grupo al que pertenece. Y otro basado en técnicas de correlación canónica y de componentes principales (Análisis Factorial) denominado Análisis Discriminante Canónico.

El análisis discriminante devolverá también los nodos que según este nuevo método están mal clasificados. Todo ello se hará por medio de las funciones discriminantes. Dichas funciones discriminantes consisten en una combinación lineal de las variables explicativas que nos permitirán clasificar a los sujetos en los diferentes grupos establecidos a priori. Hay tantas funciones discriminantes como grupos menos 1. Y para que esa función sea óptima deberá proporcionar una regla de clasificación que minimice la probabilidad de cometer errores.

La ecuación lineal discriminante, similar a la de regresión múltiple, es la siguiente:

$$D_j = B_{0j} + B_{0j} x_{1j} + B_{2j} x_{2j} + ... + B_{pj} x_{pj} \qquad j = 1,...,k\text{-}1 \qquad (2.9)$$

Donde x_j son las variables, B_{0j} son las constantes estimadas a partir de los datos originales de manera que los valores de la función difieran lo máximo posible entre los distintos grupos para facilitar la clasificación.

A partir de las puntuaciones discriminantes, es posible obtener una regla de clasificación de los individuos en uno de los grupos. Esta regla está basada en el teorema de Bayes, la probabilidad de que un sujeto con una puntuación discriminante determinada pertenezca a uno u otro se estima a través de:

$$P(G_i \mid D_j) = \frac{P(D_j \mid G_i) P(G_i)}{\sum_{i=l}^{g} P(D_j \mid G_i) P(G_i)} \qquad (2.10)$$

donde *P(Gi)* es la probabilidad de que un individuo pertenezca a un grupo determinado de la muestra si no se dispone de ningún tipo de información previa sobre el mismo. *P(Dj|Gi)* es la probabilidad condicional, da una idea de cuan probable es una puntuación discriminante cualquiera para los miembros de uno u otro grupo. Mediante la regla de Bayes se obtendrá la probabilidad a posteriori de cada sujeto para asignarlo al grupo al cual el valor obtenido sea mayor.

2.2 PREVISIÓN ANALÍTICA DE MOVIMIENTOS EN VACÍO ENTRE TERMINALES

Uno de los aspectos operacionales más importantes para estas redes de transporte es organizar los flujos de manera que los transportes en vacío se minimicen lo máximo posible lo que se traduce en dimensionar la flota de camiones. El otro aspecto directamente relacionado con el anterior y de sumo interés también es la optimización de las rutas, que la demanda esté completamente cubierta por los camiones disponibles

y la minimización del número de camiones necesarios para el sistema. Generalmente nos referiremos a todos ellos como costes de desequilibrio. Se han editado una gran variedad de métodos para optimizar el problema de costes de desequilibrio en las redes de transporte y sin embargo fuera de lo teórico muy pocos han sido aplicados para redes reales y de manera empírica.

Los costes de desequilibrio pueden ser atribuidos a variaciones temporales y espaciales del flujo. Para esta parte del análisis se ha seguido la formulación de Hall (1998), en él se analiza el comportamiento de las distintas terminales de forma individual y por grupos para ver como afecta esta variante. A partir de ahí Hall busca una relación entre la variabilidad de la terminal con su media que permita predecir el comportamiento de la primera variable. El estudio de Hall se centra en:

1 Desequilibrios a largo término en flujo medio:

- Entrante y saliente a una terminal
- Entrante y saliente a un grupo de terminales

2 Desviaciones a corto término de la media de flujo entrante y saliente de una terminal, grupo de terminales o línea de transporte.

Según Hall se puede solucionar el problema de dimensionamiento de la flota mediante un modelo analítico tal como se indica a continuación. El modelo asume que el tiempo será discretizado en incrementos de tiempo de longitud Δ. Si la capacidad del camión varía debería utilizarse una unidad común de medida de la capacidad como por ejemplo el peso o el volumen. Y en todo caso, el remolque de los camiones se considera intercambiable para el estudio, o sea, no se distinguirán los camiones según el tipo de cargas, resultando que por ejemplo, un camión refrigerador podría transportar cualquier tipo de paquetería y viceversa. Se podría decir por lo tanto que la formulación dada es para tractoras en vez de camiones. Sea $I_j(t)$ el número de camiones en la terminal *j* al final del período *t*, $a_j(t)$ las llegadas de camiones a la terminal *j* en el período *t*, y $d_j(t)$ las salidas de camiones de la terminal *j* en el período *t*. Se asumirá que se cumple el principio de conservación de camiones con el tiempo:

$$I_j(t) = I_j(t-1) + a_j(t) - d_j(t) \tag{2.11}$$

La salida de camiones está marcada por las necesidades de la flota, ya sea para distribuir cargas al sistema, para intercambios dentro de la terminal o para cubrir necesidades de otras terminales en cuyo caso serían movimientos en vacío.

Las llegadas de tractoras están marcadas por las salidas desde otras terminales en períodos de tiempo previos y de acuerdo a tiempos de viaje entre determinadas terminales. Sea *P* el número de períodos de tiempo que se estudiarán, $\lambda_j(t)$ las salidas de la terminal *j* en el periodo *t*, $\mu_j(t)$ las llegadas a una terminal j en un período de tiempo t, entonces en ausencia de movimientos en vacío entre terminales y *stock-outs*, y asumiendo estacionalidad, la cantidad de camiones en el tiempo *t* serán:

$$E[I_j(t)] = I_j(0) - (\lambda_j - \mu_j)t\Delta \tag{2.12}$$

Siendo

$$\lambda_j = \lim_{p\to\infty} \sum_{t=1}^{P} \frac{\lambda_j(t)}{P} \qquad \text{(2.12) a}$$

$$\mu_j = \lim_{p\to\infty} \sum_{t=1}^{P} \frac{\mu_j(t)}{P} \qquad \text{(2.12) b}$$

y *Δ* la longitud de un período de tiempo.

En un sistema estacionario la ecuación 2.12 sería exacta ya que *t* tiende a infinito. En ausencia de movimientos en vacío el valor del inventario de camiones crece sin límite cuando $\lambda_j < \mu_j$ y disminuye sin límite cuando $\lambda_j > \mu_j$. Para estabilizar el problema, e_j definida como la media de flujo neto de camiones vacíos en una terminal, debe igualar $\lambda_j - \mu_j$.

Por lo tanto la cantidad total de camiones vacíos en un sistema estabilizado está limitado por:

$$T \geq \sum_j \left| e_j \right| / 2 \tag{2.13}$$

Donde T es el flujo total de camiones en vacío.

El sumatorio está dividido por 2 porque cada unidad negativa de flujo vacío cuenta el doble que cada unidad positiva de flujo vacío.

En los sistemas reales el flujo de camiones vacíos puede exceder la ecuación (2.13) debido a la aleatoriedad de las fluctuaciones y debido a las variaciones de los desequilibrios de flujo en las líneas de transporte. Las fluctuaciones aleatorias podrían hacer que $\lambda_j(t)$ fuese en algún momento mayor que $\mu_j(t)$ cuando $\lambda_j < \mu_j$ o viceversa. Este desequilibrio temporal podría convertirse, según su magnitud, en camiones moviéndose en vacío en sentido opuesto al normal. Se refiere a este suceso como: *counter-veiling flow*. Variaciones entre líneas de transporte pueden provocar movimientos en vacío en una terminal de una línea mientras que camiones vacíos salen de la misma terminal hacia otra.

Por otra parte el número de movimientos en vacío viene condicionado por la disponibilidad de camiones y la posibilidad de reordenar sus rutas. Por lo general los flujos entrantes y salientes deben equilibrarse. Los inventarios y la posibilidad de reordenar rutas permiten dicho equilibrio en periodos de tiempo largos, que no instantáneos, reduciéndose así la cantidad de camiones vacíos que se mueven en la red. Un exceso de movimiento se define como cualquier movimiento en vacío por debajo del intervalo *T*.

CONSIDERACIONES TEMPORALES Y ESPACIALES.

Los siguientes modelos se usan para examinar el efecto del intervalo de tiempo en el exceso de camiones, el cual se usa para aproximar los efectos al variar el nivel de número de camiones. Los excesos de la red se calculan como flujo neto en camiones

vacíos que equilibra flujos de entrada y salida en un periodo de longitud Δ para cada terminal:

$$\varepsilon(\Delta) = [\sum_{t,j} \left| e_j(t) \right| / 2P - T] = [\sum_{t,j} \left| e_j(t) \right| - \sum_j \left| \sum_t e_j(t) \right|] / 2P \tag{2.14}$$

Definimos *D* como el conjunto de terminales con déficit, y *S* como el conjunto de terminales con exceso de vehículos. Entonces el exceso de camiones en la red podría ser calculado como

$$\varepsilon(\Delta) = \sum_{j \varepsilon D} \sum_{e_j(t) \lhd 0} \left| e_j(t) \right| / P + \sum_{j \varepsilon S} \sum_{e_j(t) \rhd 0} e_j(t) / P \tag{2.15}$$

Cada camión vacío de una terminal con exceso de camiones debería poder cubrir una demanda de otra terminal con déficit de camiones para mantener el equilibrio, resultando entonces que la ecuación no necesita dividirse por dos.

La variable *ε (Δ)* es naturalmente decreciente con *Δ* ya que el límite superior e inferior tiende a las condiciones medias a largo término a medida que *Δ* aumenta. En un límite tenemos que cuando *Δ* tiende a infinito *ε (Δ)* tiende a 0 ya que los excesos se equilibran con el tiempo:

$$\lim_{\Delta \to \infty} \varepsilon(\Delta) = 0 \tag{2.16}$$

Cuando *Δ* tiende a 0 cada movimiento cargado tiene que ser compensado por un movimiento en vacío en dirección opuesta resultando:

$$\lim_{\Delta \to 0} \varepsilon(\Delta) = \sum_j \min \lambda_j, \mu_j \tag{2.17}$$

Excesos de movimientos en vacío debidos a restricciones de líneas

Muchas empresas de transporte trabajan bajo una política de retorno en vacío, en la cual cada camión pertenece a una terminal en particular y dicho camión debe volver a la

terminal una vez descargado. En dicho sistema, el flujo de camiones entre cada par de terminales debe ser equilibrado. Esta forma de transporte es mucho más restrictiva que en el caso de grupos de terminales debido a que es mucho más fácil conseguir el equilibrio en estas últimas. Sea $\lambda_{jk}(t)$ las salidas de camiones cargados de la terminal *j* a la terminal *k* en el período *t*; $\mu_{jk}(t)$ las llegadas de camiones cargados de la terminal *j* a la terminal *k* en un período *t*; $e_{jk}(t)$ la necesidad neta de camiones vacíos para la terminal *j* desde la terminal *k* en el período *t*, igual a $\lambda_{jk}(t)$ - $\mu_{jk}(t)$ y *r (Δ)* el exceso de camiones vacíos necesarios para toda la red. En ese caso el exceso necesario para la red será:

$$r(\Delta) = [\sum_{t,j,k} \left| e_{j,k}(t) \right| / 2p - T] = [\sum_{t,j,k} \left| e_{jk}(t) \right| - \sum_{j} \left| \sum_{t,k} e_{jk}(t) \right| / 2P \tag{2.18}$$

Que puede ser también expresada de la forma:

$$r(\Delta) = \sum_{j \varepsilon D} \sum_{e_{jk}(t) \lhd 0} \left| e_{jk}(t) \right| / P + \sum_{j \varepsilon S} \sum_{e_{jk}(t) \rhd 0} e_{jk}(t) / P \tag{2.19}$$

Al contrario de *ε (Δ), r(Δ)* no tiende a 0 cuando *Δ* tiende a infinito. Esto es debido a que a pesar de que sean condiciones a largo término los movimientos de los camiones tienen que equilibrarse línea a línea con lo cual no podemos aspirar a una reducción total de los costes. Por otra parte, el limite cuando *Δ* tiende a 0 es el mismo que para *ε(Δ),* nos da un límite superior del movimiento de camiones en el peor de los casos.

2.3 OPTIMIZACIÓN DE LAS REDES DE DISTRIBUCIÓN

En este apartado se describirán las metodologías mas representativa de asignación de mercancía en una red de many-to-many. Las principales aportaciones al tema se han hecho según tres vertientes:

- Modelos EOQ y Aproximaciones continuas. Esta metodología analiza los sistemas logísticos en base a sus variables más significativas, suponiendo que éstas no

sufren variaciones significativas en la región de estudio. El método de las aproximaciones continuas se basa en la analogía entre la información de cada uno de los puntos a unas funciones agregadas que permitirán acotar la dimensión del problema. De este modo se reduce el tamaño del problema significativamente a un número acotado de variables que permiten la resolución analítica del problema (Daganzo, Newell y Robusté).

- Programación matemática. Esta línea de investigación pretende desarrollar modelos matemáticos para la optimización del sistema logístico. Las contribuciones empezaron en los años 50 y continúan hasta ahora. Se pretende formular una función objetivo a minimizar o maximizar a partir de unas variables de decisión a las que se les han aplicado ciertas restricciones. Existe una gran variedad de problemas de programación lineal o no lineal con variables reales o enteras. Esta línea pretende aportar una solución única y óptima. Sin embargo el principal problema que comporta es el elevado coste computacional.

- Algoritmos heurísticos y metaheurísticos. Debido a las limitaciones computacionales de la metodología anterior surge otra línea de investigación la cual estudia el desarrollo de algoritmos heurísticos. Estos proporcionan un equilibrio entre el coste computacional y la bondad del ajuste. La solución óptima absoluta deja de ser el objetivo absoluto del problema y se convierte en una línea de tendencia exclusivamente, considerando pues como solución optima la solución actual que corresponde a la de menor coste hallado hasta el momento. El problema de estos algoritmos es que resultan ineficientes en algunos problemas de *NP-Hard* por lo que surge la necesidad de perfeccionarlos. Se desarrollan así los algoritmos metaheurísticos. Estos son procesos iterativos que guían la búsqueda mediante un algoritmo heurístico subordinado (Dorigo, Glover, Laguna). Trabajan en un instante de ejecución con un estado actual que es variado a partir de una transición de estado, perturbación o movimiento. Esta transición de estado tiene como objetivo explorar nuevas rutas que ahorren en la función de coste. Los algoritmos permiten un incremento positivo en la función de coste a cambio de poder analizar nuevas configuraciones que lleven a nuevos óptimos locales y de este modo llegar de forma más eficiente al óptimo global de la solución. Los cuatro tipos principales son el Recosido Simulado (Kirpatrick'83, Robusté'90, Aykin'95), los

algoritmos Genéticos (Abdinnour-Helm'98), Colonias de Hormigas (Dorigo'96, Barcos'02) y por último el Tabu Search (Glover'86, Abdinnour-Helm'98).

La formulación parte de la minimización de una función de costes totales *z* para la suma de cada una de las rutas *r*, dada por la fórmula:

$$z = \sum_{r \in R} (c_d d_r + c_s n_r + F + c_t W_r) \tag{2.20}$$

Donde c_d es el coste por unidad de distancia, d_r la distancia recorrida por cada uno de los vehículos, c_s es el coste por parada, n_r el número de paradas, c_t es el coste unitario de consolidación de la carga y W_r las unidades de carga a consolidar. *F* es el coste fijo de la ruta F debido a la contratación del vehículo.

El factor común entre los distintos modelos desarrollados para la optimización de las redes es que todos cumplen unas determinadas etapas que son: localización de las terminales *hub*, asignación de la mercancía asociada a cada origen-destino a la red y el proceso de determinación de rutas.

Todos los modelos parten de una red de terminales jerarquizada en la que se encuentra una primera red llamada red primaria que sería la red que distribuye entre terminales de rotura o *hubs*, una segunda red que sería la red secundaria que es la que envía de hubs a las distintas terminales y por último una red capilar que no estaría en la base de nuestro estudio y que sería la red formada entre los particulares y cada terminal. Se trata de redes a nivel local.

Los modelos pueden partir de un número y localización prefijada de los *hubs* o bien puede determinar el número de *hubs* y su localización óptima. La capacidad de los *hubs* es un aspecto básico dentro del modelo porque son las terminales que controlarán la eficiencia del sistema.

A continuación se hará una introducción a la formulación matemática del problema de *backhauling* o movimientos en vacío.

FORMULACIÓN MATEMÁTICA DEL PROBLEMA DE BACKHAULING

En Barcos *(2003)* se plantea la formulación matemática del problema de balanceo de vehículos entre dos jornadas laborales consecutivas con desplazamientos sin carga entre terminales fuera de la jornada de servicio. La minimización del coste total de todos los desplazamientos en vacío está dada por la fórmula:

$$Min \sum_{i \in E_D} \sum_{j \in F_{D+1}} (2c_p + c_d D_{ij}) x_{ij} \qquad (2.21)$$

Sujeto a:

$$\sum_{j \in F_{D+1}} x_{ij} = A_D(i) - G_{D+1}(i) \qquad \forall i \in E_D \qquad (2.21)a$$

$$\sum_{i \in E_D} x_{ij} = G_{D+1}(j) - A_D(j) \qquad \forall j \in F_{D+1} \qquad (2.21)b$$

$$x_{ij} \geq 0 \quad \text{y entero} \quad \forall i \in E_D ; \forall j \in F_{D+1} \qquad (2.21)c$$

Siendo E_D el conjunto de todas aquellas terminales con un exceso de vehículos al final de la ruta del día *D* y F_{D+1} el conjunto de todas las terminales con falta de vehículos para iniciar la ruta en el día *D+1*. *Donde* c_d es el coste unitario dedistancia y c_p el coste unitario de parada.

El problema propuesto anteriormente considera una demanda constante entre terminales que obliga a mantener un número fijo de flota de vehículos. Adicionalmente, la variable de decisión "envíos x_{ij}" del problema representa los desplazamientos entre terminales destino final de ruta en día *D* y terminales origen de ruta en día *D+1*.

La variable x_{ij} constituye movimientos realizados sin carga, es decir sin prestar servicio y por lo tanto ineficientes.

Sin embargo, existe un conjunto de movimientos o estrategias a realizar durante el servicio activo que pueden tener asociado un coste menor para la totalidad del servicio. Este conjunto de movimientos recibe el nombre de palancas de decisión.

CAPÍTULO 3

METODOLOGÍA

En este apartado se definirá la zona de estudio tratada, los datos utilizados y las herramientas en las que se ha basado la tesina la realización del estudio. De cada subapartado del estudio se explicarán las consideraciones tomadas y se justificará su utilización.

3.1 ZONA DE ESTUDIO

El estudio se ha aplicado a una empresa líder en el sector de transportes tanto aéreo como por carretera que ha suministrado los datos de los flujos mensuales de carga entre sus 87 terminales por un período de 23 meses. De estas 87 terminales se han seleccionado 56 que son sobre las que se trabajará por la deficiencia de datos de algunas de ellas. De estas 56 terminales 6 están emplazadas en islas y 4 en Portugal. Este dato es significativo ya que tan sólo en la parte del Análisis Cluster se trabajará con el total de ellas. A partir de ese momento, y debido a que el estudio que se está realizando considera tan sólo el transporte por carretera, se trabajará con un total de 50 terminales. La representación geográfica de dichas terminales viene indicada a continuación en la figura 3.1.

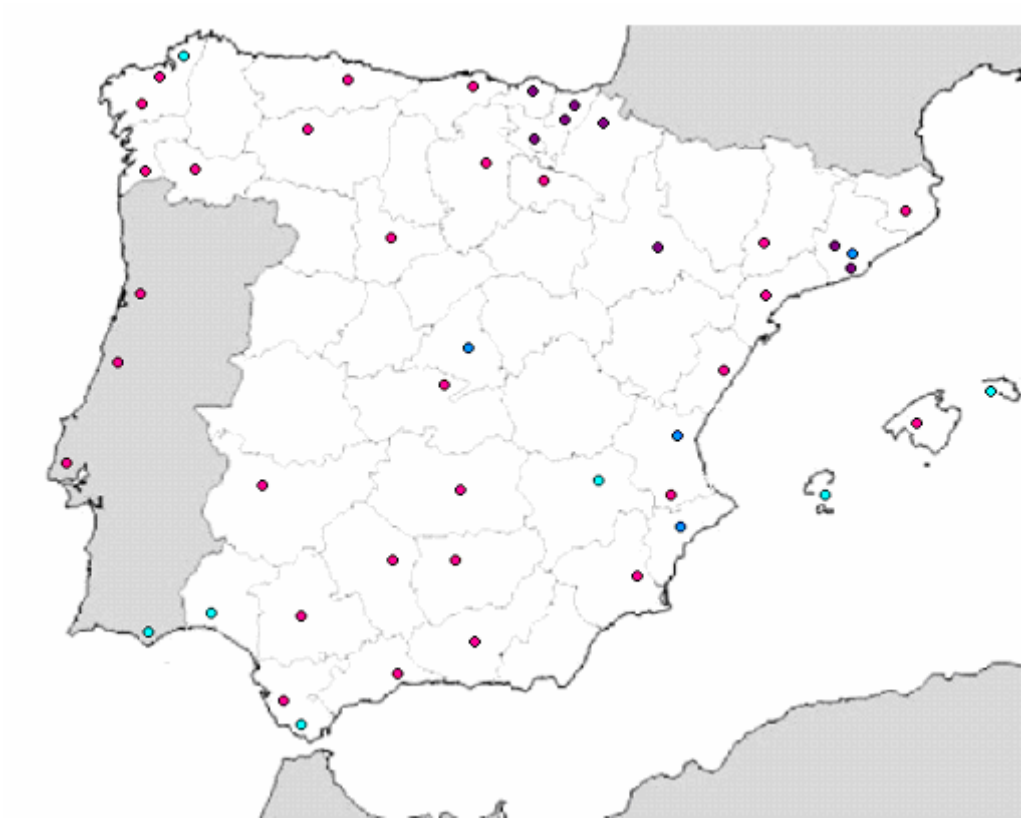


Fig. 3.1- *Situación geográfica de las terminales en estudio.*

3.2 HERRAMIENTAS DE CÁLCULO

Las herramientas de cálculo utilizadas son en todos los casos programas informáticos de los cuales se ha tratado de tener un conocimiento exhaustivo de ellos para poder extraer el máximo de información posible. En el último apartado se ha trabajado a partir de la tesis doctoral de Miquel Estrada. En este caso el estudio de ese programa se ha limitado ha su utilización y aplicación en vez de su diseño o programación por ser un temario demasiado extenso para ser tratado en una tesina.

3.2.1 Agrupación de las terminales en función de su comportamiento respecto a unas variables determinadas

Una vez estudiado y depurado el conjunto de información de la que se disponía inicialmente se puede pasar a clasificar las terminales disponibles en función de las variables correspondientes, que son en este caso, el peso y el número de bultos. El procedimiento ideal para realizar las agrupaciones deseadas es el análisis cluster o de

conglomerados debido a que, respecto la distribución de los valores de las variables, busca formar grupos lo más homogéneos posibles y al mismo tiempo que sean muy distintos los unos a los otros. Es imprescindible para conseguir estas particiones el concepto de distancias entre los distintos individuos y variables. Los casos o terminales en este caso, se irán agrupando uno a uno según su proximidad hasta conseguir un único grupo.

Estudiaremos fundamentalmente métodos jerárquicos aglomerativos que son los que nos interesan en nuestro caso por no partir de un número predefinido de clusters y permitirnos la elección del número de clusters en función de nuestro objetivo o de la coherencia de los resultados.

Se ha visto hasta aquí que los elementos diferenciadores de las diversas técnicas son la distancia entre elementos individuales y la definición de separación entre grupos, llamada método. Por lo tanto, el uso de la técnica adecuada de clusterización pasa por la elección de una distancia y un método adecuado, los cuales, a su vez se tienen que escoger según la naturaleza de los datos, la experiencia en problemas similares y la coherencia de los resultados finales.

La validez de la elección, efectivamente, no se puede asegurar hasta que los resultados del análisis se vean coherentes y por esta razón, se tiene que utilizar un número suficiente de técnicas diferentes para evaluarlas todas, con una parte aleatoria de la muestra, considerando una, o como mucho, dos como buenas, las cuales serán las que den resultados más satisfactorios. Estas técnicas son las que se utilizaran posteriormente con la base de datos completa.

El programa informático utilizado para llevar a cabo dicho análisis es el SPSS, el cual permite realizar una clusterización sobre el conjunto de variables. El caso particular de la utilización del SPSS como herramienta estadística simplifica en gran medida la interpretación del dendograma ya que el programa ofrece la posibilidad de mostrar en el listado los diferentes individuos que pertenecen a cada grupo, opción de gran utilidad cuando la base de datos es muy extensa.

PRIMERA SELECCIÓN

Se reduce la información inicial del conjunto de variables a dos en concreto los kilos de entrada y de salida por ser las variables más significativas en lo que a transporte de mercaderías se refiere, en el caso de disponer de la variable volumen se utilizaría esa, pero en este caso, no se dispone de ella. De todas las distancias disponibles, la más común y de menor coste computacional es la euclídea al cuadrado en segundo lugar la euclídea. La mayor parte de estas distancias derivan de la distancia de Mahalanobis que no se utiliza por el elevado coste computacional. Por lo general, por lo tanto, se utilizará la distancia euclídea o la euclídea al cuadrado en función de cual sea más recomendable para el método utilizado.

Se parte de los datos de los casos sin tratar, en nuestro caso, al quererse hacer un análisis en función de múltiples variables se deberá seguir un proceso de estandarización o bien de corrección de los datos, de manera que sean comparables los unos con los otros y que no produzcan en el resultado final una incoherencia del mismo debido al factor de escala.

Una vez hecho el tratamiento de datos se calcula la distancia para cada par de casos según la distancia escogida para el problema teniendo en cuenta el método a utilizar. Por lo tanto, resulta evidente, que la elección de la distancia a utilizar será significativa y modificará los resultados obtenidos. Con ello obtenemos la llamada matriz de distancias la cual me permite cuantificar su grado de similitud o semejanza.

De esa matriz se localiza el valor máximo o mínimo y se hace la primera agrupación. A continuación se recalcula la distancia de esa nueva agrupación a todos los demás casos, (no es necesario calcular la matriz entera sino sólo la fila y columna afectadas por la agrupación) obteniendo así la nueva matriz de distancias en este caso de dimensión *(n-1) x (n-1).*

En cada iteración se conocen, por lo tanto, los dos grupos que se han fusionado y a que nivel de similitud lo han hecho. Se debe decir que, generalmente, estas similitudes son decrecientes, es decir, que las similitudes de los grupos nuevos respecto del resto acostumbran a ser menores que los existentes previamente. Eso deja de cumplirse para

algunas técnicas y se produce el fenómeno conocido como inversión, que ya se estudiará más adelante.

Estos resultados se representan en un dendograma de manera que se puede observar gráficamente los clusters o conglomerados que se formaron y a que distancia lo hicieron de manera que la persona que los analiza, y bajo sus criterios, determinará si el número de clusters es lo suficientemente pequeño como para que sea interpretable y suficientemente grande como para darle un sentido físico sin resultar una simple fragmentación del número de casos. No obstante, no existe un criterio único y preciso sino que dependiendo de la distancia y método utilizados proporcionará unos resultados u otros. Resultará por lo tanto muy importante analizar cuales son las diferencias entre estos distintos métodos y distancias y cómo repercuten en el resultado.

Se elegirá el dendograma vertical ya que el número de casos es muy extenso. Por lo tanto en el eje de abcisas vendrán indicadas las distancias a las cuales se han ido formando los distintos clusters en una escala de 0 a 25, y en el eje de ordenadas se colocaran los casos a clasificar ordenados según su proximidad. A partir del dendograma se deberá escoger el número de clusters idóneo para la clasificación.

Teóricamente, para cada método existe un número ideal de grupos que reflejan una agrupación de calidad, en función de las distancias entre los clusters formados. Por lo tanto se escogerá de entre ellos el que tenga mayores distancias entre grupos y menores distancias entre elementos de ese mismo grupo. Los dendogramas resultantes de este primer análisis se incluirán en los anejos.

La figura 4.5 indica la formación del número de cluster y la distancia de formación según los distintos métodos. En el eje de abscisas se indica el número de clusters que se han formado y en el de ordenadas la distancia de formación. Esta grafica se suele utilizar para elegir el número de clusters, pero en este caso se utilizará tan solo para comprobar algunos aspectos de los resultados obtenidos según los distintos métodos.

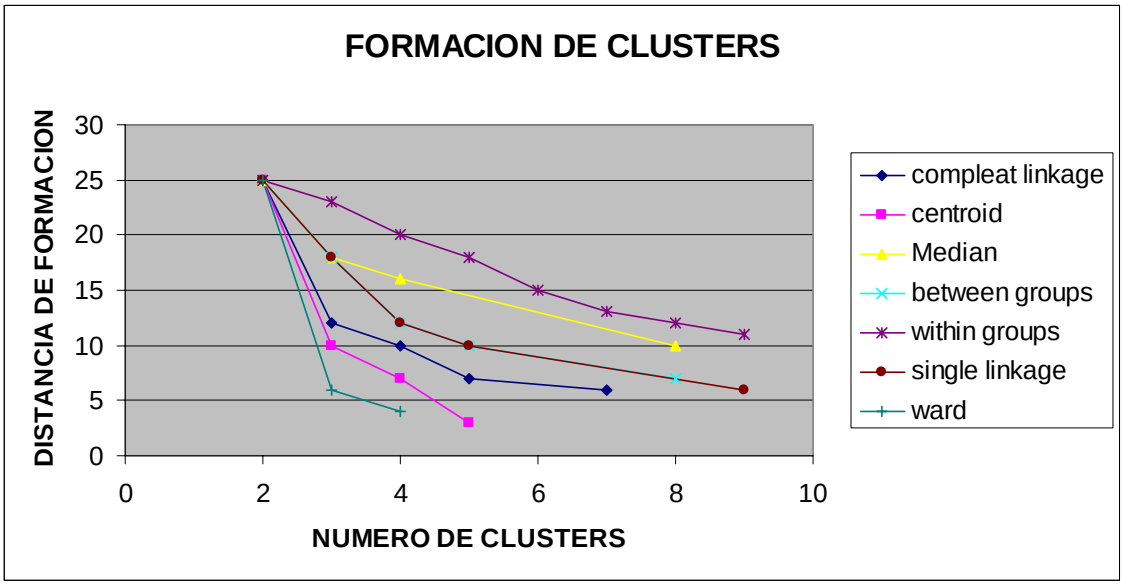


Fig. 3.2 *– Distancia de formación de los clusters según los distintos métodos y distancias.*

SEGUNDA SELECCIÓN

Se repetirá de nuevo el proceso entero, sin embargo esta vez se utilizarán todas las variables y se resolverán por los dos métodos que hayan dado mejores resultados. Se considerará que el resultado de esta segunda selección será el resultado definitivo del análisis cluster.

Una vez interpretado el dendogrma, el siguiente paso hacia la clasificación es la elección del número de clusters. El objetivo del problema, como ya se ha remarcado anteriormente, es conseguir una agrupación tal que la distancia entre los elementos de un mismo grupo sea mínima y la distancia entre los elementos de grupos distintos sea muy grande. No existe un criterio universalmente aceptado. Una forma sencilla de determinarlo es considerando los resultados directos del dendogrma, es decir, mediante una gráfica representar las distancias a las que se forma cada grupo y a partir de ahí localizar en qué iteraciones del método dichas distancias dan grandes saltos.

Véase que la nomenclatura de los grupos por SPSS no es la misma que la dada en las figuras, eso es debido a que por defecto el programa nombra en sucesión de grupos a los elementos encontrados, así el primer elemento de la lista será del "grupo 1" y el siguiente elemento encontrado en la lista que no pertenezca al grupo de uno, pertenecerá

por defecto al "grupo 2" y así sucesivamente. Se tomó la decisión de cambiar la nomenclatura puesto que se consideró más comprensible que el número del grupo fuese en concordancia con el volumen de movimientos de las terminales. La relación de cambio es la siguiente

Tabla 3.1. *Relación de nomenclatura entre el programa SPSS y la tesina.*

MAPAS	SPSS
1	3
2	1
3	2

La columna primera es la nomenclatura que se le ha dado en este estudio a los grupos resultantes del análisis cluster. La segunda es la nomenclatura que por defecto da el programa SPSS en la realización de la clasificación por conglomerados.

Debido a la condición de subjetividad de la clasificación se hará un análisis de la bondad de ajuste, la técnica que se utilizará para analizar dicha bondad es el análisis discriminante.

El análisis discriminante parte de una clasificación previa y analiza por métodos probabilísticos la probabilidad de pertenencia de un caso a un grupo determinado. Con ello se obtendrán los casos que según el segundo análisis están mal clasificados y se entrará en el análisis de cada caso en particular para decidir a que grupo se asignan definitivamente.

Una vez obtenido el resultado de las agrupaciones finales, corroborado y aceptado como bueno, se analizará el comportamiento de dichos grupos. Como análisis previo a dicho comportamiento se analizará la variabilidad de los grupos por medio de diagramas de *box plot*. Este tipo de gráficos es otro modo de resumir la distribución de los valores de las variables de forma grafica, de manera que se puedan identificar los valores singulares. En este caso cada punto no quedará definido como conjunto de 4 variables sino expresadas por separado, de manera que los errores no puedan quedar camuflados por el balance con las otras variables. En vez de visualizar los valores individuales de

cada terminal, se presentarán estadísticos básicos de la distribución: la mediana, el centil 25 y el centil 75, al igual que los valores extremos de la distribución.

3.2.2 Análisis estacional del sistema de terminales.

El procedimiento a seguir para analizar los costes de desequilibrio es un procedimiento empírico a partir de datos reales. El objetivo de este análisis empírico es desarrollar modelos de probabilidad que pudieran ser utilizados en futuros análisis para predecir el comportamiento de la variabilidad de una terminal en función de su media, o bien hallar el modelo de comportamiento que sigue el flujo de mercancías. Sin embargo el conjunto de datos de los que se parte tiene algunas limitaciones. El estudio basado en la literatura de Hall como ya se ha indicado en el apartado de estado del arte, es un análisis diario y semanal de los flujos de carga. Para que el estudio fuese completo deberíamos tener los datos diarios y sin embargo se dispone de los mensuales, con lo cual se harán las operaciones oportunas para pasarlo a diario, conociendo las limitaciones de tener un límite inferior de datos. En este apartado de la tesina, a diferencia del análisis cluster, no se trabajará con todas las terminales facilitadas por la empresa sino que se extraerán aquellas que están situadas en las islas debido a que quedan fuera del área de estudio.

El análisis empírico queda dividido en 4 secciones, la primera un análisis diario de las variaciones de flujo, el segundo la distribución de probabilidad acumulada y la tercera los desequilibrios de flujo.

ANÁLISIS DIARIO DE LAS VARIACIONES DE FLUJO

Se examinarán las variaciones diarias de dos medidas del flujo de carga: el peso y el número de paquetes. El coeficiente de variación se ha calculado para terminales individuales y para grupos de terminales. Los grupos de terminales utilizados son los obtenidos del análisis cluster en el apartado anterior de la tesina. Como ya se explicó anteriormente estos grupos se crearon a partir de la similitud en el comportamiento respecto a cuatro variables. En el análisis se excluyen los domingos por ser los días con un transporte muy inferior y pueden por ello modificar el resto de datos. Por lo tanto se trabajará con meses de 26 días. El análisis cubre tanto flujos de salida como de entrada.

El objetivo es hallar la magnitud de los costes de desequilibrio de una empresa logística. Estos costes de desequilibrio se traducen, en una empresa de transportes, a costes debidos a movimientos en vacío. Y numéricamente, una manera de considerarlos es hallando la variabilidad del sistema. En este caso, como medida de la variabilidad se ha tomado el coeficiente de variación por ser adimensional y permitir así la comparación entre variables.

Se comprobará primero la relación existente entre la media de las terminales sobre la muestra de 23 meses con el coeficiente de variación. Según Hall (1998) la variación del flujo de salida es una exponencial de la media lo que sería equivalente a decir:

$$\text{Ln}\,[CVj] = a + b\,\ln\,[\,\lambda j] \qquad (3.1)$$

Donde a y b son parámetros a estimar. En este caso en concreto se estudiará si el sistema de terminales del que se parte sigue la misma relación y las razones de dicha relación. Se estudia por otro lado el comportamiento de una variable a la que se referirá como flujo neto y que se definirá como la suma de las generaciones con las atracciones de cada una de las terminales. En ese caso la relación anterior pasaría a ser la siguiente:

$$\text{Ln}\,[NCVj] = a + b\,\ln\,[\,\lambda j+\mu j] \qquad (3.2)$$

Para el cálculo de a y b o bien se puede utilizar la herramienta de análisis de datos de excel o bien se pueden calcular por las formulas:

$$b = \frac{\sum(\ln CV \times \ln \lambda j) - \overline{\ln CV} \times \sum \ln \lambda j}{\sum(\ln \lambda j)^2 - \overline{\ln \lambda j} \times \sum \ln \lambda j} \qquad (3.3)$$

$$a = \overline{\ln CV} - b\overline{\ln \lambda j} \qquad (3.4)$$

Para saber si el ajuste es adecuado o por el contrario no se puede aceptar como bueno se utilizará el coeficiente R^2 .Cuanto más se acerque a 1 mejor será el ajuste. R^2 se define como:

$$R^2 = 1 - \frac{\sum e^2}{\sum \left(\ln CV - \overline{\ln CV}\right)^2} \tag{3.5}$$

Por último se buscará el modelo que sigue el comportamiento del flujo, según Estrada (2007) el comportamiento del flujo neto de las cargas sigue una distribución Weibull. Se ordenarán los datos y se hará el ajuste a dicho modelo. El proceso se repetirá tanto para el cociente entre los flujos de entrada y salida como para el flujo de salidas exclusivamente. En ese caso se hipotizará como en Hall (1998) que éstos siguen una distribución Normal y Gamma respectivamente y se analizará dicho comportamiento. Para saber si el ajuste es bueno o no, se utilizará el método de Kolmogorv-Smirnov aplicado mediante la herramienta de SPSS. La Prueba de Kolmogorov-Smirnov para una muestra compara la función de distribución acumulada observada de una variable con una distribución teórica determinada, que puede ser la normal, la uniforme, la de Poisson o la exponencial. La Z de Kolmogorov-Smirnov se calcula a partir de la diferencia mayor (en valor absoluto) entre las funciones de distribución acumuladas teóricas y observadas.

Se constata que el ajuste de modelo se puede realizar tanto mediante excel como SPSS, en este caso se utilizarán ambas herramientas.

3.2.3 Optimización de la red de distribución

Se ha utilizado el algoritmo de Tabú Search para la resolución del problema de optimización de la flota y de organización de las rutas de los vehículos. Esta misma metodología fue estudiada en Estrada (2007) para la resolución del problema de diseño de la red de transporte de una empresa de paquetería.

El algoritmo de la Búsqueda Tabú o Tabu Search, TS ideado por Glover (1986) es un procedimiento iterativo de búsqueda y exploración en el dominio de soluciones que trata de evitar los óptimos locales y el comportamiento cíclico del algoritmo mediante la prohibición de unos movimientos de la solución actual y una estructuración de la memoria de soluciones pasadas.

El método de TS es inicialmente determinista ya que no existe un criterio probabilístico para la optimización de la solución. Éste se basa en sustituir la solución generada sin ninguna condición de coste, eso es, tanto si es mejor que la solución en curso como si no. El principal problema es que podrían crearse procesos cíclicos. Para evitar el problema, el algoritmo dispone de una lista de las operaciones o soluciones prohibidas o tabú. Se define un vecindario constituido por soluciones que no contengan atributos tabú. La lista tabú no es estática ya que el problema impide la utilización de ciertos atributos durante un número determinado de iteraciones.

OPTIMIZACIÓN DE LA RED PARA UN DÍA DE SERVICIO

En este caso el problema consiste en la reorganización de las rutas con la ayuda de unas determinadas estrategias que permitirán reducir el coste total del servicio.

La asignación de flujos de mercancías al conjunto de rutas de la red troncal se puede expresar como un problema de programación matemática, con la identificación de una función objetivo de costes a minimizar y varias restricciones. En algunos casos gracias a la consolidación de la carga en algunas terminales, existe la posibilidad de realizar algunos envíos directos. Sin embargo entre otras terminales con menor volumen de envío resulta necesaria la realización de una ruta con paradas múltiples.

La capacidad de los vehículos juega un papel fundamental en la asignación de los recursos a las tareas o envíos de paquetes. En este punto, se debe diferenciar el problema de envíos completos de los envíos con volumen de carga inferior a la capacidad. Los envíos completos se realizan cuando todo el volumen de mercancía entre un origen y un destino puede ser transportado en un sólo vehículo sin violar las restricciones del sistema (capacidad o restricciones temporales).

Sin embargo cuando la demanda de mercancía entre un origen y destino supera la capacidad del vehículo, la mercancía se divide o fracciona en distintos envíos asociados a distintas rutas para cumplir la restricción de capacidad. En estos casos, al igual que los envíos con una carga muy inferior a la capacidad, habrá como mínimo un vehículo que presentará una ocupación baja, al no aprovechar toda su capacidad potencial para el

transporte. Estos casos conocidos como *LTL* , constituyen un campo estratégico para aplicar estrategias de consolidación.

De este modo, los envíos entre delegaciones con volúmenes de carga inferiores a la capacidad del vehículo constituyen el principal aspecto de optimización de los recursos del sistema.

Debido al elevado coste computacional el problema se descompondrá en dos fases: en la primera fase se realizará la asignación de envíos a carga completa (P1) con los distintos vehículos disponibles, mientras que en la segunda fase se planificarán todos los envíos con carga inferior a la capacidad del vehículo (P2).

En resumen, tenemos un problema, cuyos datos de entrada son:

- Grafo completo *G(N,E)* donde N es el conjunto de delegaciones y *E* los arcos entre las delegaciones
- Conjunto de centros que actuarán como *Hub*
- Volumen de mercancías asociados a las tareas de envíos
- Distancia entre cada par de delegaciones *(i,j)*
- Total de costes de cada una de las operaciones.

El total de costes se divide en: coste de servicio de un vehículo o coste temporal que representa a la empresa el alquiler diario de un vehículo si realiza la subcontratación del servicio. En el caso de flota propia sería el coste diario de amortización del recurso, salario del conductor, seguro,…por otra parte está el coste unitario de recorrido de un vehículo. Es el coste asociado a la superación de una unidad de distancia con el vehículo, e integra los costes de carburantes y mantenimiento. Coste unitario de transferencia. Es el coste que representa la transferencia de una unidad de volumen de mercancía en la terminal *hub* entre dos vehículos. Esta operación suele requerir operaciones de manipulación de la mercancía, maquinaria y tecnología diferencial al tratamiento de la mercancía sin transferencia. Coste de parada que es el coste que representa la parada de un vehículo en una delegación para efectuar la carga y/o descarga de mercancía. Y por último el coste unitario de congestión en delegación que es el coste de penalización que supondría el hecho de tener un vehículo en la delegación

durante una unidad de tiempo a la espera de que un muelle de carga quedase libre para efectuar la carga y/o descarga de mercancía.

El problema consiste en la minimización del coste total del transporte del sistema para cada par de terminales *(i,j)*.

$$z = \sum_{r \in R} (c_d d_r + c_s n_r + F + c_t W_r) \tag{3.6}$$

Donde el primer término de la ecuación representa el coste dependiente de la distancia, donde el parámetro c_d es el coste de la unidad de distancia (Euros/km) y d_r es la distancia total recorrida en la ruta *r*. El segundo término reproduce el coste de las operaciones que se deben hacer en cada parada en la ruta *R*, y el parámetro c_s (euros/parada) es la unidad de coste de parada y n_r el número total de paradas en *r*. El tercer término representa el coste fijo de cada ruta *r* donde F (Euros/vehículos) representa el coste de dejar parado un vehiculo durante un día de servicio y por último, el último término representan los costes debido a la consolidación en las terminales *hub*. Por lo tanto c_t está medida en Euros/m^3 y representa el coste por unidad de volumen transferido y W_r es la cantidad de carga transferida a la terminal *hub*.

El problema que se trata tiene una gran dificultad computacional, por ello se divide el problema en dos subproblemas P1 que es el problema de identificación de las rutas directas entre delegaciones con vehículo a carga completa y otro problema P2 de envió entre delegaciones a carga fraccionada.

El problema P1 por ser un problema a carga completa pasa a ser un problema de *sheduling* puro ya que el único factor a tener en cuenta es la asignación de tareas a vehículos para poder satisfacer las restricciones temporales del problema. La estrategia de paradas múltiples y de consolidación en *hub* no tienen sentido por ser envíos a carga completa.

En el caso del problema P2 se pueden llevar a cabo diversas transiciones de estado o movimientos que permitan modificar el estado actual para estudiar nuevas posibilidades.

El esquema completo del problema sería el siguiente:

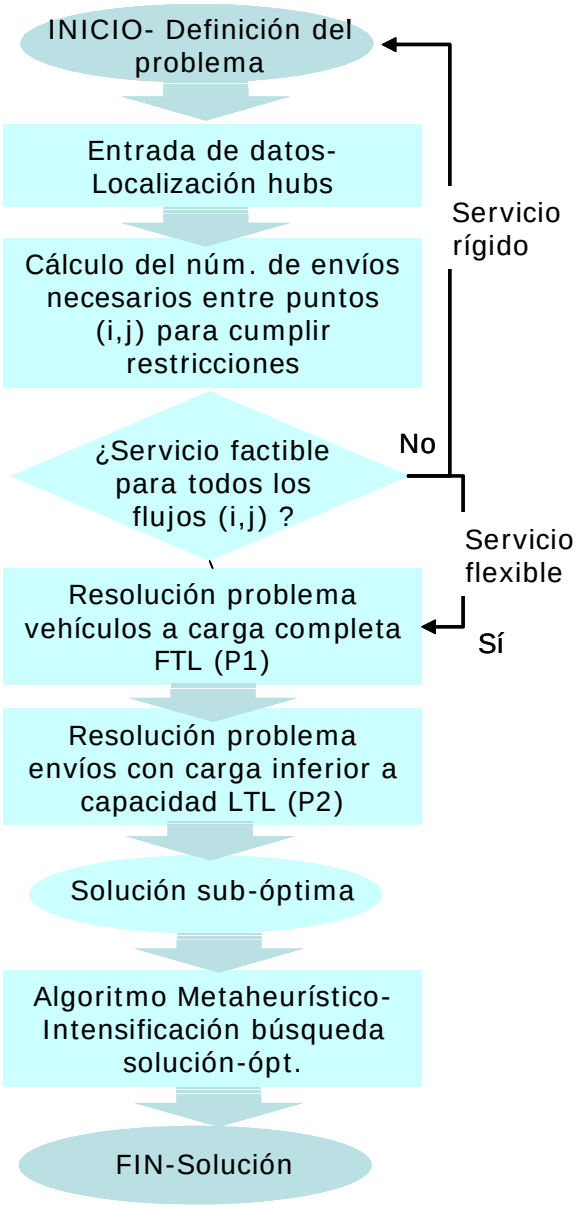


Fig. 3.3. – *Esquema del problema.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

Para conseguir la reducción de coste el programa aplica unas estrategias o movimientos que permiten romper con la línea de búsqueda actual y seguir por nuevas rutas que permitan una reducción de coste mayor.

Dichos movimientos son los siguientes:

- M1. *Opt-Swapping movement* o movimiento de intercambio de cadenas de envíos entre dos rutas. Este movimiento afecta exclusivamente a dos rutas ya existentes r, $s \in R$ pertenecientes al conjunto de soluciones que presentan un nodo común $i \in Nr$ y Ns. El concepto se basa en intercambiar la subcadena de nodos de una ruta a la otra a partir del nodo común de las dos rutas, tal como muestra la figura 2.5. Nr y Ns corresponden al conjunto de nodos de la ruta r y s respectivamente.

  El movimiento únicamente cambia la longitud de la ruta y tiempos de visita de las delegaciones posteriores al nodo i de cada ruta y por lo tanto, el cambio en los tiempos muertos de espera en las delegaciones. Sin embargo, los demás términos presentes en la función de coste no sufren cambios por lo que no es un movimiento que reduzca significativamente el coste total de las rutas. Esta perturbación se ha implantado para poder cambiar la estructura de la ruta y así pasar a morfologías todavía no exploradas y más eficientes. Por lo tanto, es un movimiento que pretende reducir o cambiar las estructuras de aquellas rutas que superan el tiempo máximo de visita en las delegaciones situadas a final de ruta.

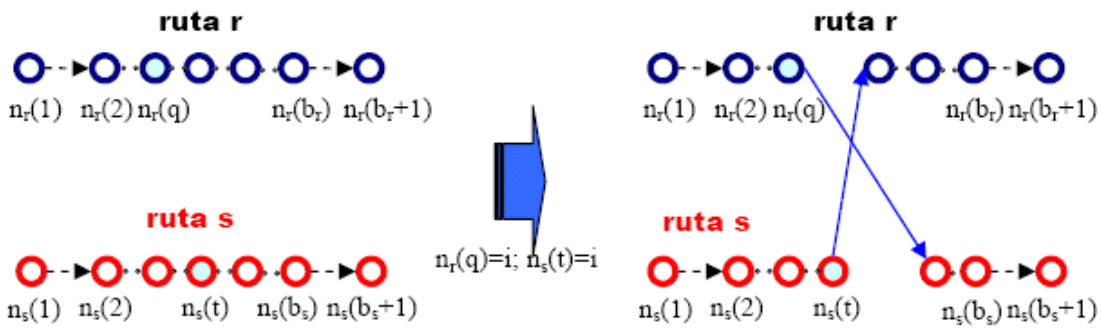


Fig. 3.4. *Esquema del funcionamiento del movimiento M1*. Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

- M2. *Opt-Swapping movement* o movimiento de intercambios de cadenas de envíos entre tres rutas. Esta perturbación de la solución actual se basa en la modificación de la cadena de nodos de rutas existentes $r, s, u \in R$.

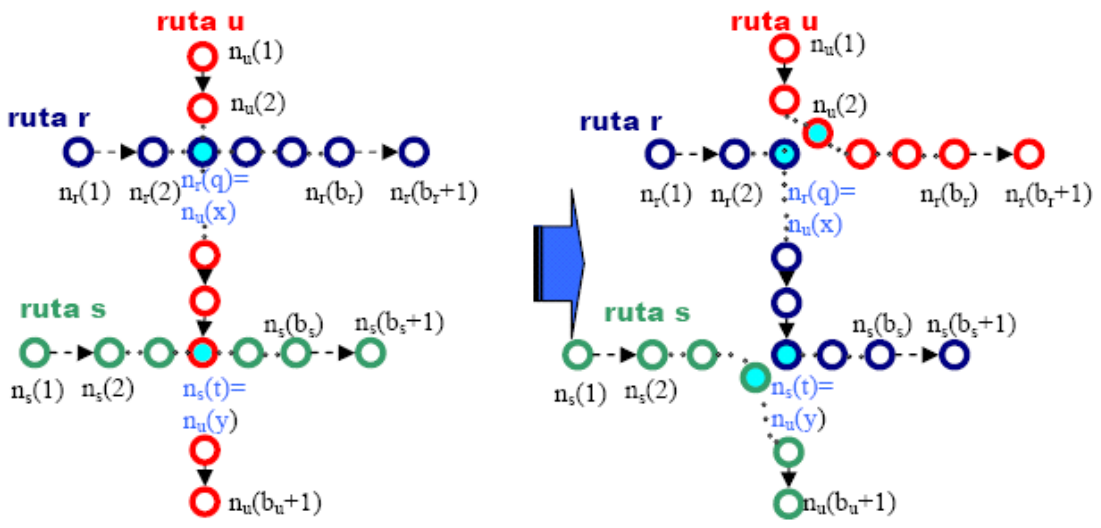


Fig. 3.5. *Esquema del funcionamiento del movimiento M2.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

Se escoge aleatoriamente uno de los nodos y se rompen por él las rutas de manera que el número de rutas se mantiene pero la composición de cada una de ellas no.

De nuevo este cambio lo que aporta es una nueva redistribución de las rutas y no una aparente mejora de la función de costes. Servirá por lo tanto para explorar nuevas rutas que mas adelante puedan producir una reducción de costes.

Esta perturbación permite a su vez cambiar la extensión de todas aquellas rutas que vulneran las restricciones temporales en delegaciones, por lo que es una herramienta útil para pasar de una solución infactible a una factible.

- M3. *Hub&Spoke improving movement* o movimiento de envió con estrategia *hub.* Las rutas existentes se hacen llegar a una terminal *hub* donde se consolidan las cargas y se carga de nuevo en los camiones correspondientes a dicha consolidación para su posterior reenvío. La contribución principal de este movimiento es la eliminación del recorrido entre los nodos *i* y *j,* con la contrapartida de la creación de una nueva línea que parte de *j*. Es especialmente interesante la aplicación en el caso de la existencia de dos rutas que acaben y empiezan en *i* y *j* con lo cual se podría llegar a eliminar una ruta.

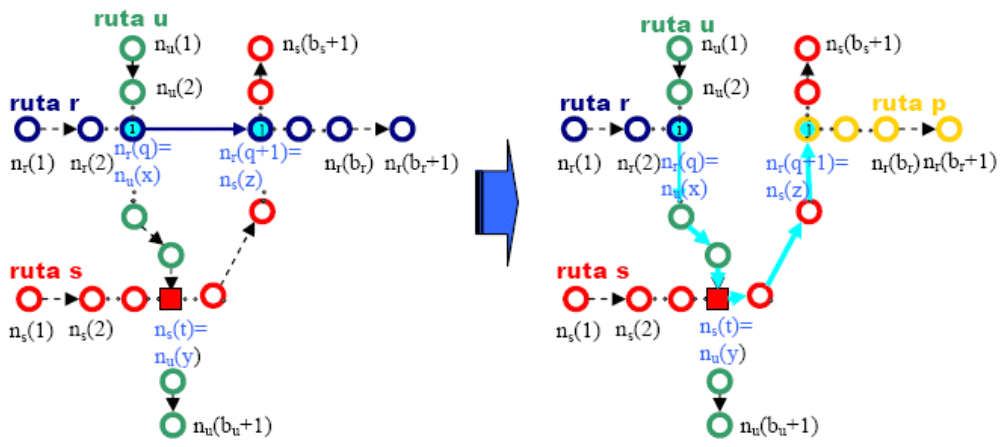


Fig. 3.6. *Esquema del funcionamiento del movimiento M3*. Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

- M4 Peddling improving movement o movimiento de envío con estrategia peddling. Esta perturbación al igual que la anterior, también produce reducciones de coste significativas por medio del envío de todas las cargas que son transportadas a lo largo de un arco *(i,j)* de la ruta *r* perteneciente a *R* a través de otra ruta s perteneciente a *R* con paradas múltiples. Como viene indicado en la figura 2.8

  Esta perturbación permite la eliminación del recorrido del arco *(i,j)* en la ruta r asi como la eliminación de la parada en la delegación *j* de la ruta *r*, si bien también produce unos sobrecostes al igual que se producía con el movimiento anterior. Por un lado, el movimiento conlleva la creación de una nueva ruta *p* constituida por todos los nodos de la cadena formada entre la posición siguiente a la de rotura *q* y el punto final de la ruta *r*. Adicionalmente, la ejecución de esta perturbación también puede desencadenar la superación de la restricción de la capacidad en la ruta *s*.

  En este sentido, los costes son muy similares a los del movimiento *hub* sin la consideración de la manipulación de la mercancía en la terminal de consolidación.

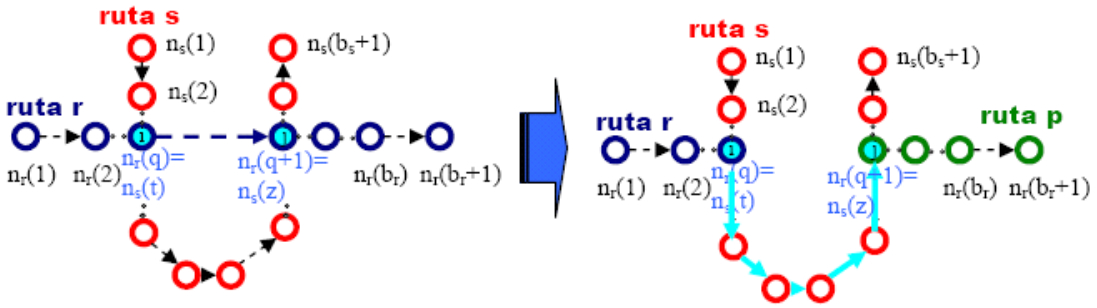


Fig. 3.7. *Esquema del funcionamiento del movimiento M4.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

- M5. Insertion movement o movimientos de fusión de dos rutas existentes. Este movimiento o perturbación de la solución produce ahorros significativos de coste al tratar de encadenar dos rutas que compartan el mismo nodo inicial y final. La configuración final conseguida con este movimiento se detalla en la Figura 2.10

  Este movimiento es muy interesante por reducir considerablemente el coste de la solución aunque no conduzca a nuevos espacios de soluciones temporales de distribución al incrementar en gran medida el numero de nodos a visitar por un mismo vehículo.

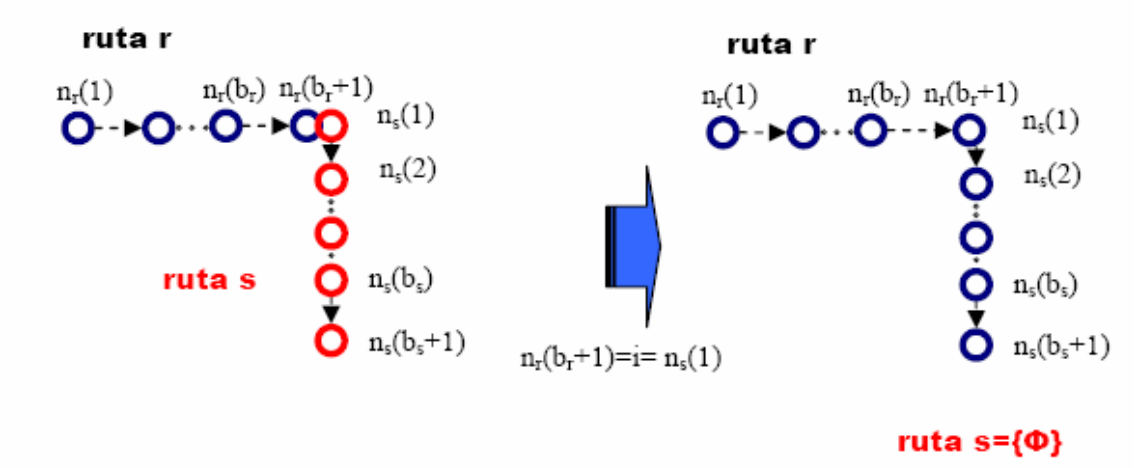


Fig. 3.8. *Esquema del funcionamiento del movimiento M5.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería.

Se constata que para la ejecución del problema, el algoritmo utiliza tres tipos de memoria, la memoria a corto plazo que es la encargada de evitar la búsqueda de soluciones ya consideradas para que no se realicen ciclos. Consta de una lista de atributos llamados tabú, e impide su utilización durante un número determinado de iteraciones N1, se utiliza también una memoria a largo plazo que impide la repetición de movimientos ya realizados y potenciar la exploración de nuevas rutas (diversificación) y por último consta del criterio de aspiración que es la componente que desempeña la función de ensamblar las memorias a corto y largo plazo.

Hasta ahora se ha tratado de las posibilidades de reducción de coste en la organización de la flota para un solo día, se estudia el problema de asignación de flujos a carga incompleta correspondiente a la organización de envíos de un día determinado D. En este contexto, las rutas se establecen sin tener en cuenta la localización de la flota en la finalización del servicio del día anterior.

De ahí nace el nuevo problema, cómo adaptar la organización diaria para una flota de camiones que no está dispuesta de forma óptima. Una vez estudiado el problema de la organización día a día se debe estudiar cómo equilibrar la diferencia de demanda entre días consecutivos.

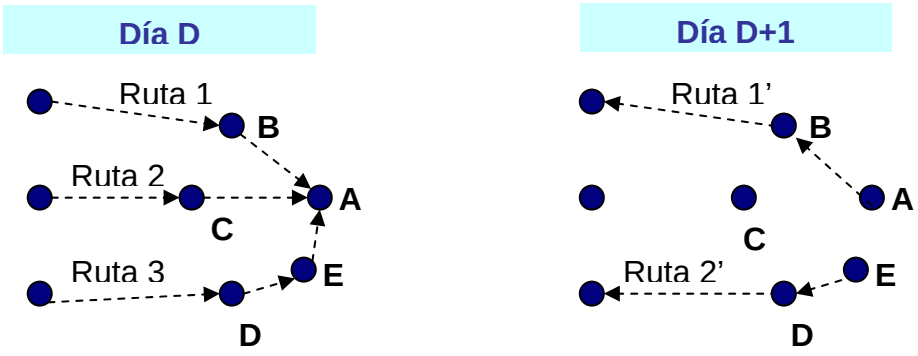
Para ello se puede, o bien solucionar el problema por medio de la contratación o la parada de los camiones en las terminales que se necesite o bien resolver el nuevo problema de optimización del problema de *backhauling* por medio de la adaptación de un algoritmo al nuevo sistema.

OPTIMIZACIÓN DE LA RED PARA VARIOS DÍAS DE SERVICIO.

Se trata de un nuevo problema de optimización que se estudia mediante nuevos algoritmos.

Debido a la variación temporal de la demanda a escala diaria y a la variación espacial de la demanda del territorio los flujos de mercancías son asimétricos en la red y su valor variará en función del día de la semana y el mes en el que nos encontramos. Es por este motivo que el número de vehículos utilizados diariamente puede variar y que el balance

entre vehículos entrantes en una terminal el día *D* y vehículos salientes de la misma terminal el día *D+1* no está compensado. Este hecho obliga a enviar vehículos entre las terminales sin carga en la finalización de la distribución del día *D* para preparar las salidas de las rutas de sus terminales de origen del día *D+1*. El hecho de realizar este rebalance de vehículos entre terminales sin carga constituye una ineficiencia de los recursos ya que no se produce ningún servicio y por lo tanto, es un fenómeno a eliminar. En este punto, resulta clave determinar las terminales óptimas para finalizar o empezar un servicio y por lo tanto, la programación de las rutas del día *D* deben estar coordinadas con las del día *D+1*. En este subcapítulo se afronta la resolución del problema de reasignación de los tramos finales de las rutas de un día *D* y los tramos iniciales del día siguiente *D+1* para minimizar las ineficiencias del sistema.



Segmento final ruta 1: B-A
Segmento final ruta 2: C-A
Segmento final ruta 3: D-E-A

Segmento inicial ruta 1':E-A-B
Segmento final ruta 2': E-D

Fig. 3.9.- *Ejemplo de la variación diaria de demanda.* Fuente: Estrada, M. 2007 Análisis de estrategias eficientes en la logística de distribución de paquetería

Si se define *A(i)* y *G(i)* respectivamente como el número de rutas que acaban en la terminal i en el día *D* (atraídas) y el número de rutas que empiezan en la misma terminal para el día *D+1* (generadas), su diferencia representa el déficit (o sobrante si es negativo) de vehículos entre la gestión de envíos entre ambos días. Si este análisis se extiende a todas las terminales de la red, el número total de ineficiencias o movimientos adicionales para balancear vehículos vendrá determinado por el valor V_T de la ecuación (3.7):

$$V_T = \sum_{i=1}^{N} \left| G_{D+1}(i) - A_D(i) \right| \tag{3.7}$$

No obstante, existe un número determinado de movimientos, eliminaciones o generaciones de vehículos entre las rutas del día *D* y *D+1* que son inherentes al sistema y que por tanto no pueden ser consideradas como ineficiencias causadas por una mala gestión de rutas. Esta afirmación es válida si se parte de una asignación de servicio a toda la demanda diaria de servicio y no para rutas fijas. Si se considera R_D el número óptimo de rutas de servicio del día *D* y R_{D+1} el valor correspondiente al día *D+1,* el número mínimo total de variaciones entre orígenes y destinos de terminales viene determinado por el término $V_{\min}$ de la ecuación (3.8), es decir la diferencia entre el número de rutas asociado a cada día:

$$V_{\min} = \left| R_D - R_{D-1} \right| \tag{3.8}$$

Por lo tanto, se puede determinar el número total de movimientos de balanceo a considerar como ineficiencias reales por el término V_{in} de la ecuación (3.9):

$$V_{in} = V_T - V_{\min} \tag{3.9}$$

En el ejemplo de la figura 2.11 se puede observar que la diferencia entre el número de rutas es igual a $V_{\min}$ = 1. Sin embargo, si realizamos el cálculo del término V_T se puede comprobar que su valor es de 3. Este hecho se explica porque el punto *E* pasa a ganar una ruta generada y el punto *A* pierde 2 rutas de salida en el día *D+1* en relación a las rutas del día *D*. En este sentido, para poder hacer el rebalanceo de vehículos entre las dos programaciones diarias, se suprimirá el servicio activo de un vehículo que acabe la ruta en *A* y otro de los vehículos que acaban en este punto se desplazará sin servicio hasta el punto E. El tercer vehículo que acaba en *A* no deberá realizar ningún movimiento ya que existe una ruta que inicia servicio en este punto en el día *D+1*.

Todos estos movimientos forman parte del conjunto de palancas de decisión citados anteriormente. Éstas son unas estrategias que permiten reducir el coste de la totalidad del servicio.

En particular los movimientos que se plantean son:

Movimiento 1.Directo interdía. Desplazamiento muerto entre terminal día *D* y terminal día *D+1*.

El movimiento en cuestión es el tradicional en el balanceo de vehículos al finalizar la jornada laboral de reparto de un día *D*. El vehículo que cubre una ruta que finaliza en una terminal en la que el día *D+1* tiene excedente de vehículos, dejará de ser activo o se desplazará hacia otras terminales con déficit de vehículos para el día *D+1*.

En particular, será de interés aplicar el movimiento entre las terminales *(i,j)* si existe una ruta $r \in R_D$ con una terminal i en la posición $n_r(m_r)$ a final de ruta que presente $A_D(i)>G_{D+1}(i)$. Adicionalmente, debe existir una ruta s R_{D+1} con una terminal j en la posición ns(1) a inicio de ruta que cumpla A_D (i)<G_{D+1} (i).

El coste debido a este movimiento es

$$C1 = d_{ij} \cdot c_d + n_r \cdot c_s \qquad (3.10)$$

Movimiento 2. *H-P-F*. Reasignación de un arco en posición final de ruta en el día *D* con estrategia *hub* o *peddling*

Este movimiento puede tener una aplicación potencial en el caso de disponer un punto *i* en la posición final n(m) de una ruta $r \in R_D$ con la condición de *A(i)>G(i)*. Si el punto j situado en la posición penúltima de la ruta *n(m-1)* presenta la condición *A(i)<G(i),* el par de puntos *(j,i)* puede ser asignado a otra ruta existente $s \in R_D$ mediante una estrategia de paradas múltiples o de consolidación en *hub*.

Movimiento 3.*H-P-I*. Reasignación de un arco en posición inicial de ruta en el día *D+1* a otra ruta con estrategia *hub* o *peddling.*

Este movimiento es relativamente parecido al anterior pero en este caso únicamente afecta a la programación de rutas del día *D+1*, por lo que puede ser preferible ya que

afecta a un escenario temporal más lejano al tiempo de planificación (mayor margen de maniobra). En este caso, puede resultar eficiente la reasignación del par de puntos *(h,k)* con h ocupando la primera posición *n(1)* y *k* la segunda posición *n(2)* de una ruta u R_{D+1} si $A_D(k) < G_{D+1}(k)$ y $A_D(l) > G_{D+1}(l)$. La reasignación puede aplicarse a una ruta $v \in R_{D+1}$ con una estrategia *peddling* o a dos rutas $v, w \in R_{D+1}$ con la estrategia *hub & spoke* sin violar las restricciones del sistema.

Los costes de los movimientos 3 y 4 son los mismos dependiendo del tipo de estrategia que se lleve a cabo. En el caso de que se realice una estrategia *hub* el coste será de:

$$C3 = c_t W_t - c_d \cdot d_{ij} \tag{3.11}$$

En el caso de llevarse a cabo una estrategia *peddling* el coste se reduciría a:

$$C4 = -c_d \cdot d_{ij} \tag{3.12}$$

La razón de este valor negativo para ambas estrategias es el hecho de que se reduce en uno el número de arcos necesarios, y el valor añadido al coste en el caso de estrategia *hub* es el coste debido a la consolidación de carga.

Movimiento 4. Directo *I-F*. Desplazamiento de un arco en posición inicial en ruta del día *D+1* a posición final de otra ruta.

El movimiento se basa en detectar una ruta $r \in R_{D+1}$ que presente un nodo i en origen con exceso de rutas generadas en día *D+1* en relación a rutas finalizadas del día *D*. Por su lado, el segundo punto de la ruta debe tener un déficit de rutas iniciadas en este punto el día *D+1* en relación a las atraídas el día *D*. Adicionalmente, debe existir una ruta s R_{D+1} que presente un nodo final coincidente con el punto origen *i* de la ruta *r* para poder reasignar el cubrimiento del arco al final de la ruta *s*. En particular, las condiciones a cumplir serían:

$$\exists i \in N_r \cap N_s \Big| \quad A(i) < G(i),\ \mathrm{i} = \mathrm{n_r}(1)\ ,\ \ \mathrm{i} = n_s(m) \tag{3.13}$$

$$\exists j \in N_r \Big| \quad A(j) > G(j),\ \mathrm{j} = \mathrm{n_r}(2) \tag{3.14}$$

El coste de este movimiento es *0* ya que lo único que proporciona es posponer el problema al día siguiente.

Movimiento 5. Eliminación de un vehículo dentro de la ruta (*Headquarter*) Este movimiento tiene un coste *CostV2* de aproximadamente 150Euros/h debido al regreso sin servicio.

Movimiento 6. Aumento de la flota en un vehículo para poder cubrir la demanda adicional del sistema. Este movimiento se realiza en el caso de no poder realizar ningún movimiento alternativo que pueda cubrir dicho desplazamiento. En ese caso se subcontratara el servicio de uno o más vehículos adicionales. El coste de este nuevo movimiento es 0.

Las variables con las que trabaja el nuevo algoritmo y que resultan las variables a equilibrar son:

Gatr1 (i) número de rutas que recibe la terminal *i* en el día *D*
Gene2 (i) número de rutas que salen de *i* en el día *D+1*

En este caso también se trabaja con palancas de decisión que son los movimientos para los que esta programado el algoritmo y que sirven o bien para reducir el coste computacional o de servicio, o bien para dirigir la búsqueda hacia nuevas líneas de investigación que posteriormente puedan producir un decremento elevado de coste.

El caso óptimo sería que las dos variables presentadas con anterioridad *gatr1 (i)* se igualara a *gene2 (i)* eso significaría que no serían necesarios movimientos de rebalanceo en la red, sin embargo por lo general no será así. Pueden darse diversas situaciones, que en el día *D+1* sobren vehículos, en ese caso se contabilizarán los vehículos en exceso

mediante la variable *nvehexc(i)* que serían el número de vehículos que sobran para la demanda exigida en el día *D+1* en la terminal *i*. Esta variable se contabiliza después de haber rechazado la posibilidad de realizar ningún movimiento alternativo y que la única posibilidad restante sea parar el vehículo. En ese caso se aplicará el movimiento 6, que como se explicó anteriormente, consiste en eliminar uno de los vehículos del sistema. Este movimiento sería aplicado a la terminal *i* en el día *D+1*.

Otro caso posible sería el inverso, que en el día *D+1* hubiese un exceso de carga no compensable con ninguno de los movimientos alternativos a la distribución de camiones, en cuyo caso se procedería a la subcontratación de un nuevo vehículo para la terminal *i,* ese caso se contabiliza mediante la variable *nvehafe (i)*.

Las dos variables significativas en el algoritmo son *glis(i)* que representa el número de links que salen de la terminal *i* en el día *D* hasta el punto de inicio de ruta en día *D+1* y *glia (i)* que son los links que entran en la terminal *i* en el día *D+1* desde un punto final de ruta en día *D*. Por lo tanto estas dos variables son exclusivas también de este segundo algoritmo.

El coste está relacionado con el número de vehículos trabajando en vacío o no balanceados, el inventario viene determinado por la ecuación (3.6):

$$gatr1(i) - nvehexc(i) - glis(i) - gene2(i) - nvehafe(i) - glia(i) = \text{rutas no balanceadas} \quad (3.15)$$

Se parte de la consideración de que el sistema se ha dimensionado para el instante 0 y por lo tanto:

$$nvehexc=nvehafe=glis=glia=0 \quad (3.16)$$

Por lo tanto la ecuación anterior para el instante $t = 0$ resultará:

$$\sum_{i=1}^{N} \left| gatr1(i) - gene2(i) \right| = numero\ de\ rutas\ o\ vehículos\ no\ balanceados$$

Los cambios o movimientos que efectúa el algoritmo para 2 días consecutivos los realiza por lo tanto sobre las variables añadidas a partir del instante 0 o día 1. *N* es el número de terminales, los cambios serán entrelazados entre las distintas terminales para reducir el coste. El algoritmo teórico tiene como objetivo encontrar la combinación para la cual se cumple que el número de rutas no balanceadas es igual a cero.

En el problema para la reorganización de rutas para un solo día se aplicaba el algoritmo Ogro o *Greedy Algorithm* para obtener la solución inicial y a continuación se aplicaba el Tabu Search para la obtención de una solución refinada. Dicho algoritmo proporciona un coste llamado *Coste 1*.

Ahora al tener que balancear el flujo para dos días, aplicaremos ambos algoritmos para obtener la organización y costes de cada uno de esos días obteniendo de forma análoga a la del día 1 un coste llamado *Coste 2*.

En este nuevo paso se pretende analizar el coste de balancear ambos días para que disminuyan los camiones parados. En el caso de que no utilizáramos el 2º algoritmo el coste final no será Coste 1 + Coste 2 sino que llevará un añadido dependiendo de la variación de demanda entre ambos días. Por lo tanto el coste final será:

$$Coste\ final = Coste1 + Coste2 + \Delta Coste \qquad (3.18)$$

Si las atracciones son mayores que las generaciones significa que la demanda disminuiría y que por lo tanto tendremos un coste añadido, si por el contrario las atracciones son menores que las generaciones significará que la demanda aumentaría y por lo tanto sólo será necesario el aumento de flota que repercute con un coste 0.

Ese será el objetivo del nuevo algoritmo, por lo tanto tenemos un nuevo problema de optimización.

3.3 BATERÍA DE PROBLEMAS

Se realiza a continuación una batería de problemas y se aplica también a los datos de la empresa objeto de este estudio. Se hará el análisis en un principio para un único día y se aplicará el algoritmo sobre el total de las 50 terminales disponibles. Los *hubs* serán calculados por el programa a partir de la optimización de la red a pesar de que la empresa ya tiene asignados 7 centros de consolidación de carga. Se estudiará en primer lugar la optimización de la red para 50 terminales y la optimización a continuación por zonas y en una batería de problemas de dimensión inferior a la estudiada. Se han escogido en concreto dos zonas de estudio. La primera citada con el nombre de Levante, está formada por las terminales situadas en las regiones de Aragón, Catalunya y Comunidad Valenciana. La segunda zona llamada Norte esta formada por todas las terminales situadas en la zona norte de la península, siendo así las terminales correspondientes a las regiones de Girona, Barcelona, Lleida, Navarra, La rioja, Guipúzcoa, Vizcaya, Álava, Cantabria, Asturias y La Coruña. También se han considerado tres problemas teóricos con variaciones del número de terminales, considerándose así un sistema de 16 terminales, otro de 45 terminales y otro de 88. En cada caso y sistema de terminales se han considerado dos casuísticas para poder considerar la variación de demanda en dos días consecutivos. Para los casos de sistemas reales, se consideró como demanda del día *D*, las salidas del mes de enero divididas entre 26 días laborables. Para la demanda del día *D+1* se consideraron las salidas del mes de febrero divididas de nuevo entre los 26 días laborables. Para los casos teóricos se tomó la batería de problemas de calibración de Estrada *(2007).* La nomenclatura utilizada en este apartado es, para la demanda del sistema de terminales del día *D* se le ha llamado caso *a* y para la del día *D+1* se le ha llamado *b.* Así la demanda del día *D* para el sistema de 16 terminales se ha citado como *16 a* y para el de 45 terminales del día *D+1* se le ha llamado *45 b.* Se debe constatar que la variación de la demanda en los problemas teóricos es mucho mayor que en los problemas con casos reales. Mientras que en los casos reales el coeficiente de variación es de 0,3 en los casos teóricos la demanda pasa a ser la mitad del día *D* al día *D+1.*

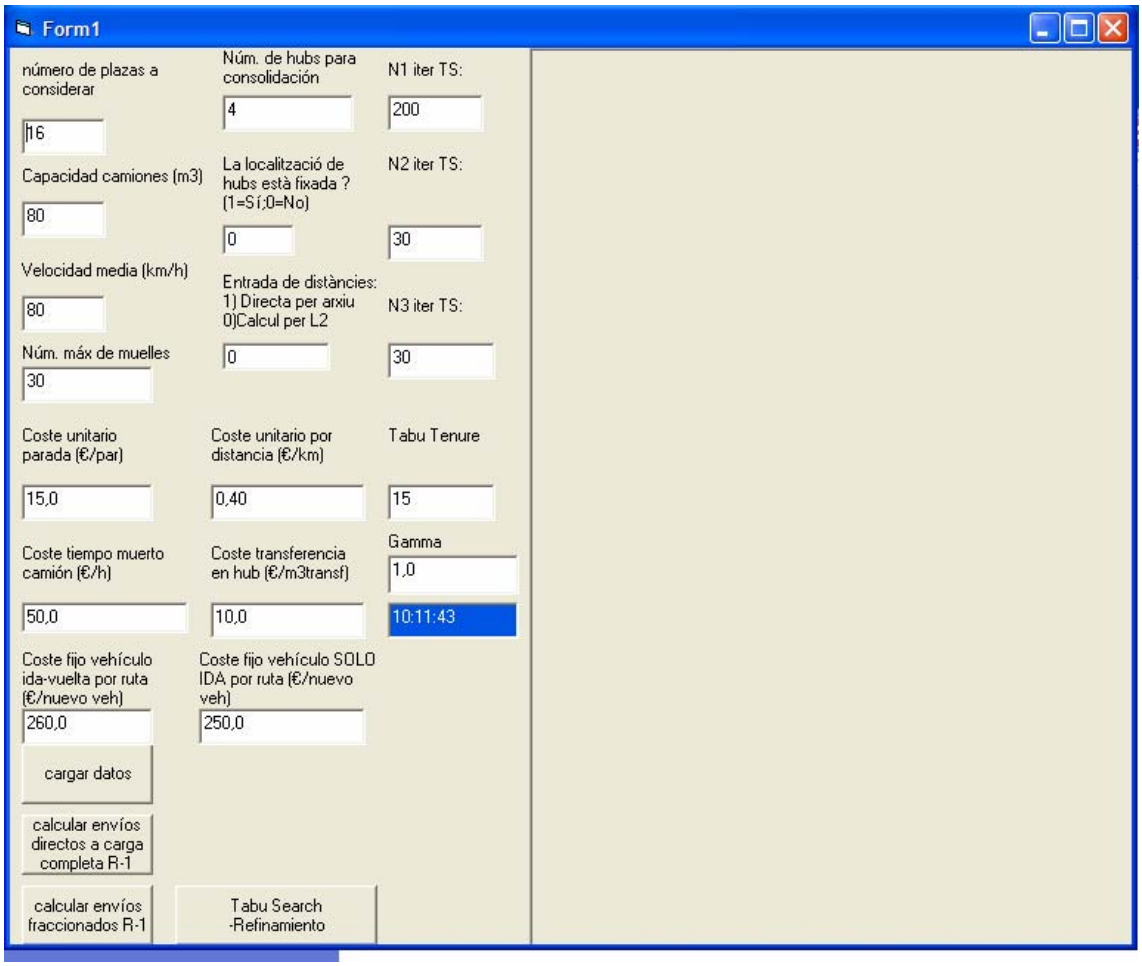


Fig. 3.10. - *Archivo para la introducción de los datos de entrada.*

VALORES CONSIDERADOS EN EL PROBLEMA

Los parámetros que actúan en el problema son, el número de terminales del sistema, el número de hubs, la capacidad de los camiones, la velocidad de desplazamiento, el tiempo de apertura y cierre de las terminales, el número de muelles de cada una de las terminales, las distancias entre cada una de ellas y por último una serie de parámetros a calibrar. Esta calibración está hecha en Estrada (2007): el *N1* que representa el número de iteraciones que se harán para llegar a la solución, *N2* es el número de iteraciones en los que se aceptará una solución infactible y *N3* el número de iteraciones sin mejora de la solución actual. En Estrada (2007) la calibración del algoritmo mediante la batería de problemas obtuvo unos resultados satisfactorios para un $N2 = N3 = 0{,}2N_T$. El parámetro del tabú tenue determina el número de iteraciones en que un atributo de la solución es considerado tabú. Estrada en su tesis encontró un valor aceptable para $\theta=100$. En todo caso la solución del algoritmo es menos sensible a este parámetro que a los de parada e inicialización explicados anteriormente. Otro parámetro a considerar es

el número de terminales de consolidación. La elección del número óptimo de *hubs* para la red. Si no se tuviera en cuenta el coste de instalación se llegaría a la conclusión de que los mejores resultados se obtendrían para $N = N_T$ ya que daría una mayor consolidación de la carga. Una vez introducido el coste de instalación al problema en Estrada (2007) se demuestra que para un $N_T = 25$, el número óptimo de *hubs* sería de *1* o *2*, para problemas de tamaño medio $N_T = 45$, el número óptimo estaría entre *3* y *4*.

Se consideran camiones de una capacidad de $80\,m^3$ que será uno de los factores limitantes del problema. En el caso de los problemas con sistema real de terminales se trabajó con 25 toneladas que sería meramente significativo ya que en ese punto la densidad de la mercancía debería ser muy alta, de lo contrario el factor limitante volvería a ser el volumen. Por otra parte se considera el coste fijo de prestación de vehículo en ruta equivalente a 260 euros por vehículo considerando ida y vuelta y 250 euros por vehículo para ida exclusivamente. El coste kilométrico que en todas las fórmulas se ha citado como c_d se ha considerado 0,4 Euros/km, el coste unitario de parada c_p se tomó el valor de 15 Euros/parada y finalmente el coste unitario de transferencia en *hub* c_t de 10 euros/m^3.

Las ventanas temporales que se tomaron fueron de 15 horas inicialmente, considerando como hora de inicio la hora 0 a pesar de que en la realidad no se inicia a esa hora sino a las 7. La razón de haber tomado esa decisión es el hecho de que de lo contrario el algoritmo entendería el tiempo de circulación de los camiones como horario en dos días consecutivos. En algunos casos en los que los camiones debían hacer rutas muy largas, el algoritmo se truncaba por la imposibilidad de cumplir los límites temporales en esos casos se amplió el abanico horario a 18 horas. La velocidad de circulación se tomó para los problemas teóricos 80km/h y para los problemas reales 100km/h y el número de muelles se puso 30 muelles por cada terminal a pesar de que en Estrada (2007) se concluyó que la probabilidad de que llegaran más de 5 camiones a la vez era muy poco probable, con lo cual a partir de 5 terminales el algoritmo es muy poco sensible a los cambios.

Como se puede observar en la figura 3.3 habrá cuatro pasos en el proceso de cálculo del algoritmo. El primero será para cargar los datos, un segundo en el que se resolverá el primer problema de envíos directos, el tercero correspondiente a la resolución del segundo problema para el cálculo de los envíos fraccionados y por último el cuarto paso que aplica el algoritmo Tabu Search para la obtención de la solución refinada. Dicho refinamiento corresponde a una mejora de un 10% respecto la solución dada por el Greedy Algorithm.

CAPÍTULO 4

ANÁLISIS DE RESULTADOS

En este apartado se explicarán los resultados obtenidos en cada uno de los subapartados del estudio en cuestión y se realizará un análisis de dicho resultado. En cada caso se comprobará la coherencia y la consistencia de los datos obtenidos con datos exteriores de otros estudios similares o bien de la lógica del problema.

4.1 AGRUPACIÓN DE LAS TERMINALES EN FUNCION DE SU COMPORTAMIENTO RESPECTO A UNAS VARIABLES DETERMINADAS

PRIMERA SELECCIÓN DENTRO DEL ANÁLISIS CLUSTER

Se representa a continuación la figura 4.1 en ella se puede comprobar la naturaleza de los datos.

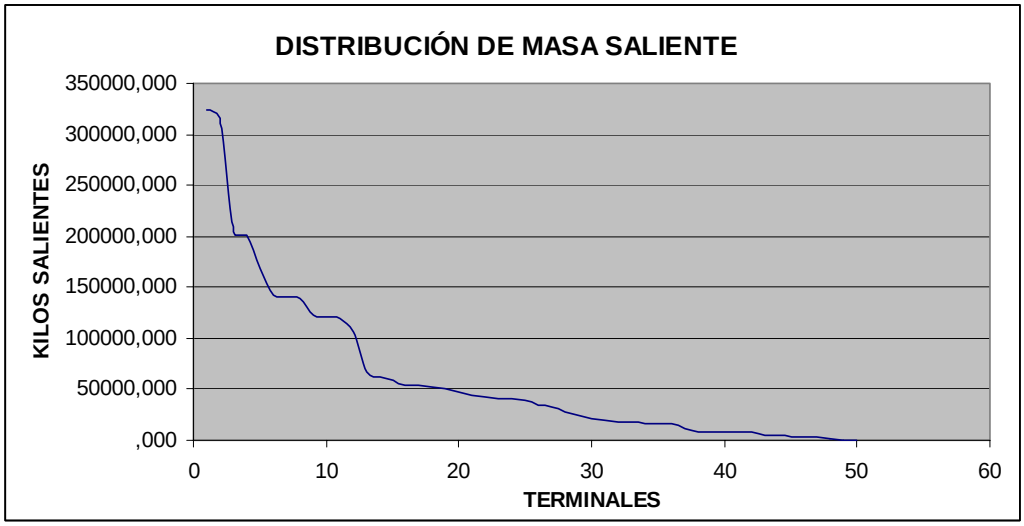


Figura 4.1 – *Distribución del flujo de masa salientes de las distintas terminales.*

En la figura 4.1 se puede ver que disponemos de pocas terminales con un volumen de movimientos muy grande y muchas terminales con un volumen muy pequeño, por lo tanto es de suponer, que hayan grupos de terminales que contengan muchos casos, y otros que contengan muy pocos. Eso es justo lo que se puede comprobar con los distintos dendogramas. De los 7 métodos existentes (y para cada método su correspondiente distancia adecuada) se obtienen 7 dendogramas que se adjuntan en el anejo. En ellos se puede observar que la mayoría de los métodos crean clasificaciones difusas con los grupos poco diferenciados entre sí y unidos mediante valores puente. En el análisis por todos los métodos excepto en el de Ward, para una clasificación con 3 clusters nos dan dos grupos de dos elementos y un tercer grupo con el resto de terminales. Eso va en contra de la primera suposición que se hizo referente al método de Ward como método más conveniente. La razón es que nuestra intención es conseguir grupos de terminales con un comportamiento similar, y una agrupación en la que todas la terminales excepto 4 estén dentro del mismo grupo no resulta significativa. La ventaja del método de Ward, es que debido a su formulación los coeficientes de las distancias tienden a homogeneizar el número de terminales entre grupos.

A continuación se adjunta la figura 4.2 proporcionada también en el apartado anterior en la que se puede ver la formación de los clusters y a la distancia a la que se formaron.

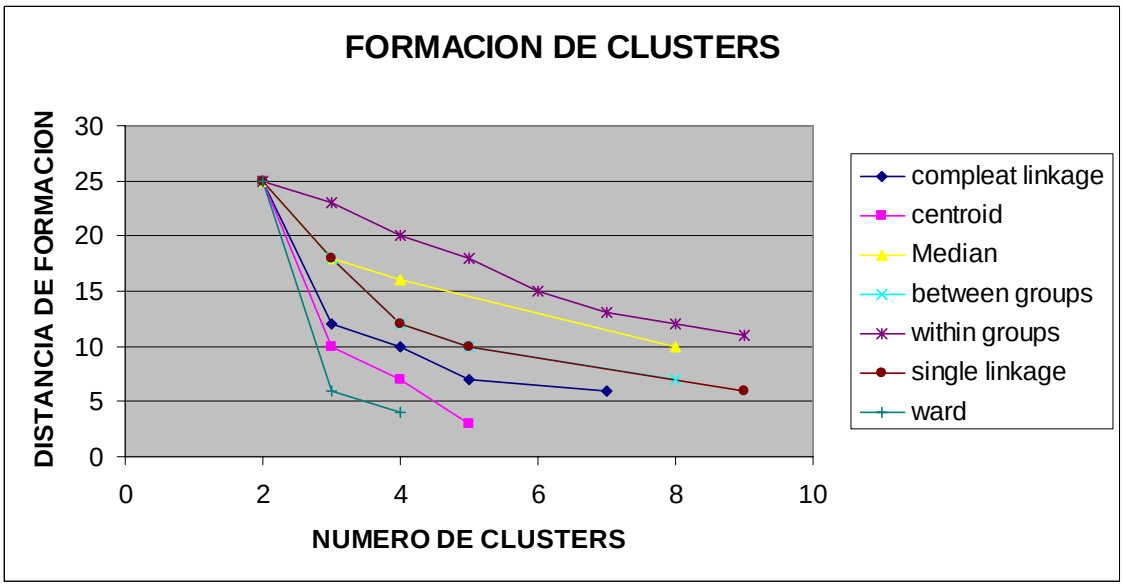


Fig. 4.2 –*Formación de los clusters en función de su distancia de formación.*

Si se analizara independientemente esta gráfica del resultado de los dendogramas, se podría llegar a la conclusión de que el método de promedio intragrupos es el que da mejores resultados ya que proporciona una distancia entre grupos mucho mayor. Ese razonamiento es cierto y de hecho concuerda con la realidad física. La razón es la siguiente. La característica en común de todos estos dendogramas es que uno de los grupos contiene más de un 75% de los casos, eso es debido a la falta de homogeneidad en el comportamiento de las terminales. Si tuviésemos un conjunto de casos con un comportamiento completamente aleatorio, la distribución de clusters sería más homogénea. Por el contrario, se dispone de la distribución mostrada en la figura 4.1

A continuación se explicará el comportamiento de los valores puente mediante uno de los dendogramas obtenidos en el análisis.

```
Dendrogram using Single Linkage

    C A S E      0         5        10        15        20        25
  Label     Num  +---------+---------+---------+---------+---------+

  Case 10    10   -+
  Case 40    40   -+
  Case 13    13   -+
  Case 39    39   -+-+
  Case 9      9   -+ |
  Case 25    25   -+ |
  Case 6      6   -+ |
  Case 12    12   -+ |
  Case 4      4   -+ |
  Case 23    23   -+ |
  Case 30    30   -+ |
  Case 47    47   -+ |
  Case 16    16   -+-+
  Case 33    33   -+ |
  Case 5      5   -+ |
  Case 46    46   -+ |
  Case 17    17   -+ |
  Case 53    53   -+-+
  Case 24    24   -+ |
  Case 15    15   -+-+
  Case 56    56   -+ |
  Case 38    38   ---+
  Case 2      2   -+-+
  Case 27    27   -+ |
  Case 37    37   ---+-+
  Case 22    22   -+ | |
  Case 36    36   -+-+ |
  Case 32    32   -+ | |
  Case 8      8   ---+ |
  Case 11    11   ---+ +-+
  Case 20    20   ---+ | |
  Case 55    55   -----+ |
  Case 42    42   -----+ |
  Case 52    52   -----+ |
  Case 54    54   -------+
  Case 3      3   -------+
  Case 43    43   ---+-+ +---+
  Case 49    49   ---+ | |   |
  Case 29    29   ---+-+ |   |
  Case 51    51   ---+ | |   |
  Case 44    44   -----+-+   +-+
  Case 50    50   -----+     | |
  Case 19    19   -----+     | |
  Case 1      1   -----+-+   | |
  Case 34    34   -----+ |   | |
  Case 28    28   -------+---+ |
  Case 21    21   -------+     +-+
  Case 35    35   --------------+ |
  Case 41    41   -----+---+   | +---------------+
  Case 45    45   -----+   +---+ |               |
  Case 14    14   ---------+     |               +-----------------+
  Case 31    31   ---------------+               |                 |
  Case 26    26   -------------------+-----------+                 |
  Case 48    48   -------------------+                             |
  Case 7      7   ----------------------------------------+-------+
  Case 18    18   -----------------------------------------+
```

Fig. 4.3 – *Dendograma del análisis por el método de Single Linkage.*

En este diagrama se puede observar la existencia de valores puente, son valores que no están clasificados de forma clara en ningún grupo y que por lo tanto, al estar situados en una frontera son difíciles de interpretar y de clasificar, un ejemplo de este efecto es el caso 31. Como se puede observar es una terminal que se ha unido de forma singular a

un grupo de terminales. Si eligiéramos hacer un análisis de 6 grupos, uno de ellos estaría formado por un único elemento, siendo en realidad no muy disímil del grupo 6 al que se ha unido a una distancia de 7 unidades. El método de Single Linkage tiende, como ya se ha dicho antes a crear este tipo de irregularidades. Este dendogrma explica también marcado en rojo el fenómeno citado anteriormente, en el cual los grupos que se han creado tienen una distancia muy grande entre ellos (gran variabilidad entre grupos) pero el número de elementos de cada grupo es también muy diverso. Este efecto se acentúa especialmente para este método ya que tiende a remarcar el valor de los *outliers* o de los valores extremos.

Una vez analizados todos los dendogramas (adjuntos en el anexo) se llega a la conclusión de que el método de Ward y el de Complete Linkage, son los que me unos resultados más coherentes y mejores para poder cumplir el objetivo de este apartado ya que por una parte no tienen valores puente y por la otra proporcionan grupos mayores y más compactos. Se hace una segunda selección a partir de estos dos métodos.

SEGUNDA SELECCIÓN

Se repite el análisis mediante estos dos métodos, en este caso, a partir de todas las variables. Los respectivos dendogramas vienen adjuntos en las Figuras 4.6 y 4.7 respectivamente. Se ha marcado con una línea roja la distancia de formación de 3 clusters en cada uno de ellos y en azul se marcan los distintos grupos formados. Como se puede observar la distancia entre los distintos cluster es mucho mayor para el método del Complete Linkage que para el de Ward. Por su parte el método de Ward crea grupos de mayor tamaño

En el primer dendograma se pueden observar dos clusters con únicamente dos terminales cada uno, al igual que en el análisis por dos variables.

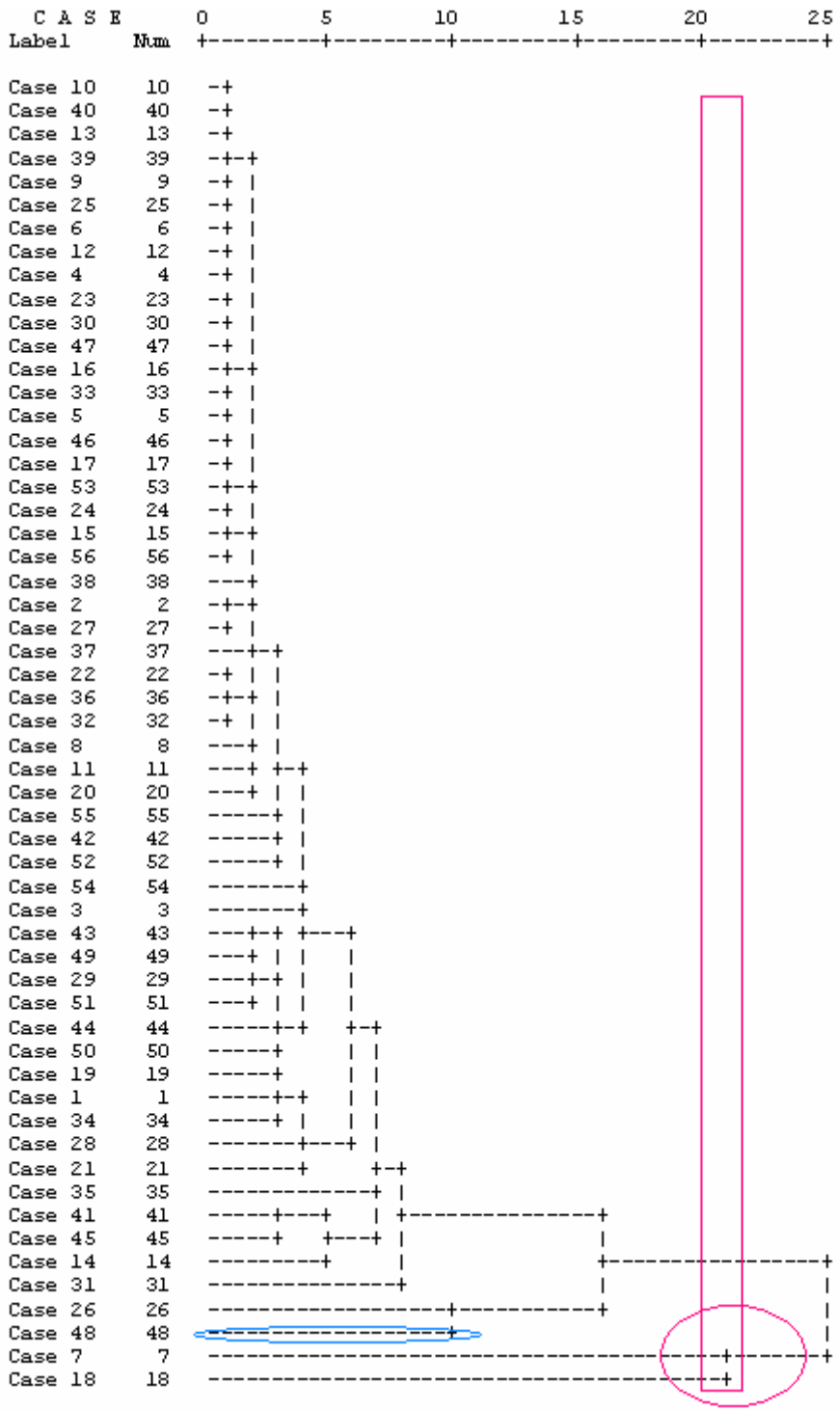


Fig. 4.4. - *Dendograma del análisis cluster por el método de Complete Linkage*

Es por ello que, como se suponía, se concluirá con que el método de Ward es el más adecuado, no por aproximarse más a la realidad, sino por ser el más adecuado para el estudio de la variabilidad en el próximo apartado.

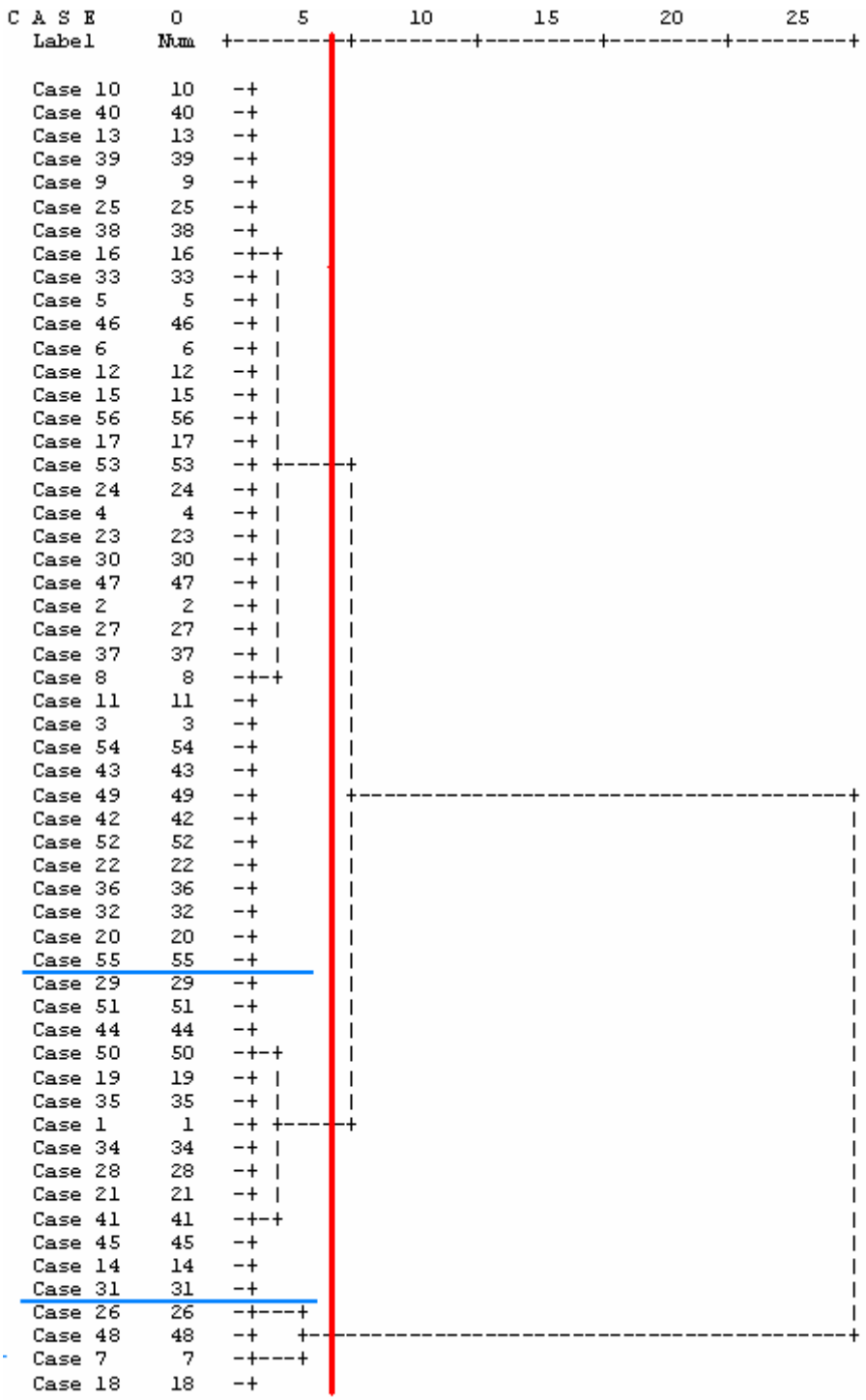


Fig. 4.5. - *Dendograma del análisis cluster por el método de Ward*

ANÁLISIS DEL DENDOGRAMA POR EL MÉTODO DE WARD

En este dendograma se puede observar que existen terminales que se han unido muy fácilmente entre ellas, como es el caso de las terminales 11,3,54,43,49….o bien 40, 13 39, 9 y otras a las que les ha costado más unirse, como es el caso de Coslada con

Alicante, o Valencia con Sta. Perpetua. Ese dato ayuda a intuir que la variabilidad entre el grupo de las primeras terminales citadas será mucho menor que la variabilidad del segundo grupo citado, o sea, Valencia, Alicante, Santa Perpetua y Barcelona. También se puede constatar, que resultaría fácil unir las terminales con un volumen medio de movimientos con las terminales con un volumen pequeño, sin embargo, resulta muy complicado unirlas con el primer grupo de terminales con un volumen muy elevado. Esta unión se produce a una distancia 25 que dentro de la escala de representación es la máxima.

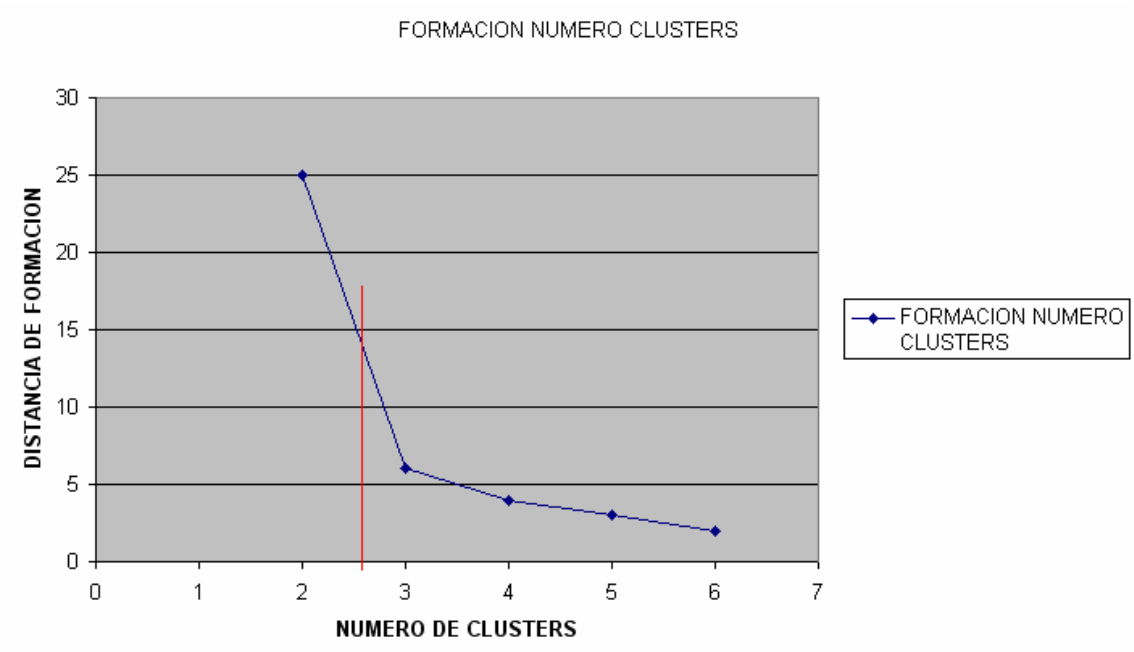


Fig. 4.6 - *Distancia de formación de los clusters por el método de Ward*

En la figura (4.6) se puede observar que existe un salto muy importante entre la formación de 2 clusters y de 3 clusters, eso nos indica que ahí es donde la distancia entre grupos es mayor. Por lo tanto según la gráfica el número óptimo de clusters en este caso sería 3. Por lo tanto se repetirá el análisis con el SPSS guardando ahora la agrupación de las terminales para 3 clusters. La clasificación es la siguiente:

Tabla 4.1 - *Clasificación de las terminales según el análisis Cluster*

Clasificación definitiva Cluster	KII	Ks	NII	Ns	G
10 - STA. PERPETUA - (BCN)	3352631	1685758	194800397	114619441	3
29 - COSLADA (MADRID)	3032789	2368387	202900535	197140906	3
52 - RIBAROJA DEL TURIA (VALENCIA)	1414019	1346248	128329618	107344395	3
87 - TORRELLANO (ALICANTE)	1885749	1068925	124870074	75002233	3
01 - LAZKAO (GUIPÚZCOA)	410748	216593	84052697	31421453	1
20 - DERIO (BILBAO)	790851	757383	86790781	61561370	1
31 - MÁLAGA	161667	678991	12729621	62590589	1
40 - HERNANI (SAN SEBASTIAN)	478845	390895	64836817	36011992	1
60 - PAMPLONA	490234	394503	86358444	36429051	1
61 - COBATILLAS (MURCIA)	236514	727074	23693624	53422981	1
63 - IGUALADA (BCN)	1384997	352019	103596677	28036058	1
70 - VITORIA	523730	271517	74761502	26455144	1
73 - SEVILLA	354050	794056	32439865	83250153	1
80 - EL PRAT (BCN)	1109183	460335	76491239	43939669	1
83 - VIGO	389164	545442	27840463	52505745	1
84 - SAN JUAN DE MOZARRIFAR (ZARAGOZA)	949743	568668	75030725	51517415	1
89 - ALVERCA DO RIBATEJO (LISBOA)	171420	565529	28867628	59002774	1
90 - SILVOTA (ASTURIAS)	300327	655063	25690221	52404831	1
03 - CORDOBA	76767	325832	5245454	29381992	2
04 - ALBAIDA (VALENCIA)	527010	143553	41239764	10294890	2
05 - MANZANARES (CIUDAD REAL)	106118	234861	9437757	18699087	2
08 - ALBACETE	69557	161855	5166184	13837628	2
09 - ALGECIRAS (CÁDIZ)	25683	119091	2368198	11634856	2
12 -PELIGROS (GRANADA)	220196	391157	14972057	34196166	2
13 MENORCA	21595	74927	919545	7281850	2
14 - FARO (PORTUGAL)	9401	65834	611170	5011190	2
15 PALMA	103381	390405	17404657	37222054	2
16 IBIZA	9532	94489	659493	11534367	2
17 STA CRUZ DE TENERIFE	10882	46327	1402208	6523246	2
23 - FERROL (LA CORUÑA)	79350	111013	9969195	10768398	2
26 - SAN JUAN DEL PUERTO (HUELVA)	27476	180151	1780650	14806497	2
27 - LEÓN	88701	261570	4715280	22362651	2
35 - HUECAS (TOLEDO)	308952	283248	26240160	25420578	2
41 - CASTELLÓN DE LA PLANA (CASTELLON)	216337	336102	21386035	25652460	2
42 - SANTIAGO DE COMPOSTELA	77162	223991	10022660	20490103	2
49 - JAEN	52984	268345	2374890	22779527	2
51 - ZAMORA	7866	89100	294645	6571556	2
55 - PUERTO DE STA. MARÍA (CÁDIZ)	60816	334143	5299162	28434031	2
62 - LÉRIDA	101722	263484	10930527	18425851	2
65 - SANTANDER	227639	345735	17155994	29093676	2
67 - SALAMANCA	38352	188630	2893390	16416473	2
74 - TARRAGONA	223900	364529	21085614	25880397	2
75 - MÉRIDA (BADAJOZ)	103597	335354	10553565	30579620	2
76 LAS PALMAS	11	3450	894	171159	2
77 LAS PALMAS	16545	51765	919040	6952686	2
79 - CUENCA	6592	64552	446016	5142310	2
81 - BESCANO (GERONA)	310231	412696	31057587	28023095	2
82 - LA CORUÑA	326190	414338	33055472	44102590	2
85 - LUGO	92931	211104	5511688	17187552	2
86 - SAN CIPRIÁN DE VIÑAS (ORENSE)	160183	239410	11365288	18447357	2
88 - VILA DO CONDE (OPORTO)	404326	481679	35576690	41512155	2
93 - VALLADOLID	170079	333916	32141884	30525719	2
95 - COIMBRA - (PORTUGAL)	77595	256063	3154792	23101329	2
96 - LOGROÑO	485670	260372	37390824	21725560	2
97 - VILLAFRIA (BURGOS)	166036	222848	24516742	20062250	2
98 ALMERIA	71712	86433	7411239	7848232	2

En la tabla 4.1 se puede observar para cada una de las terminales analizadas, la media del número de paquetes que salen, que llegan y el número de paquetes que salen y que entran. Según este análisis la distribución geográfica de las distintas terminales en función del grupo de pertenencia es la mostrada en la figura 4.6:

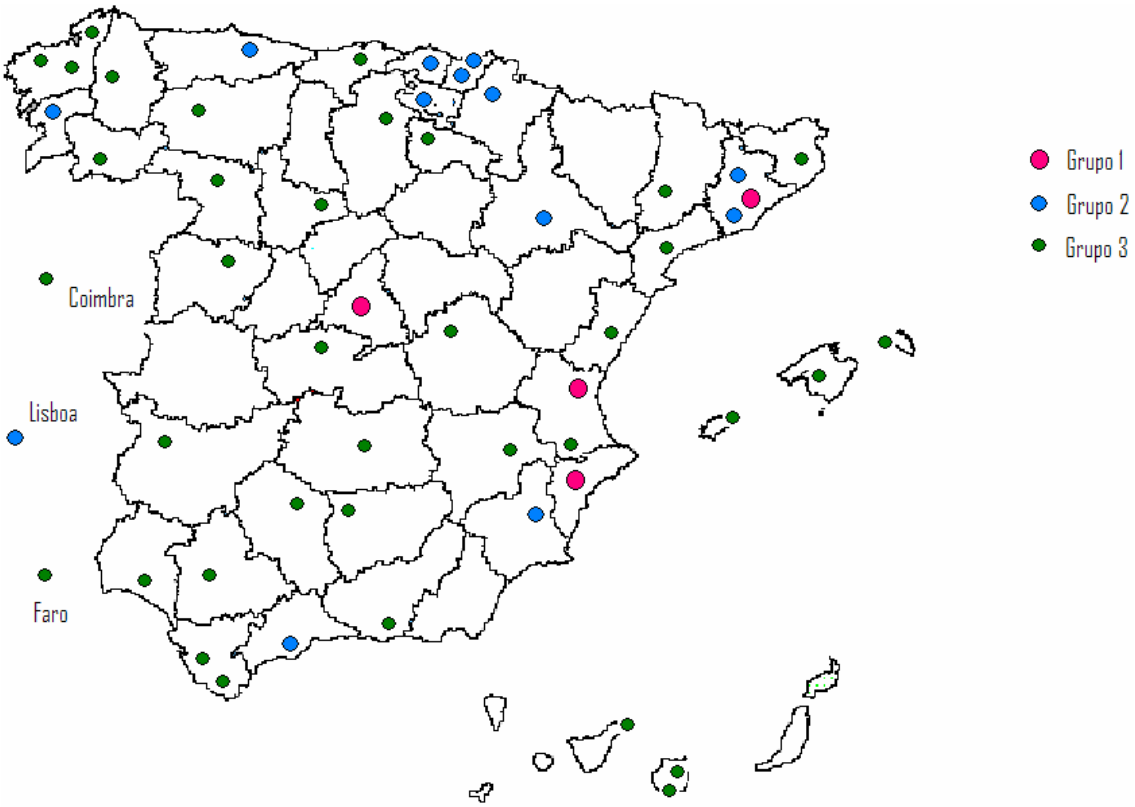


Fig. 4.7 – *Distribución geográfica de las terminales en función del grupo al que pertenecen.*

Esta agrupación por medio del mapa geográfico da una idea del volumen intercambiado entre zonas. Una observación interesante es que excepto el primer grupo, todos los demás disponen de una cierta correlación entre su localización y su volumen. Se puede ver que en toda la zona del País Vasco existe una cierta homogeneidad de volumen de paquetes, al igual que en prácticamente toda la zona de Andalucía, Extremadura, Castilla la Mancha, Castilla León y prácticamente todo Galicia.

Los resultados del análisis discriminante vienen adjuntos en las tablas siguientes.
La primera tabla proporciona los estadísticos de los elementos de cada grupo. Estos estadísticos son los que se utilizarán más tarde para analizar la variabilidad por grupo

Tabla 4.2 – *Estadísticos de cada grupo*

V2		Media	Desv. típ.
1	**Ns**	553676,6	369535,7
	Nll	527004,9	184026,3
	ks	57370022	3,04E+07
	kll	48467802	1,59E+07
2	**Ns**	133765,2	135452,6
	Nll	228219,8	124373
	ks	12280958	1,23E+07
	kll	19686872	1,08E+07
3	**Ns**	2421297	920638,6
	Nll	1617330	560655,1
	ks	1,63E+08	4,19E+07
	kll	1,24E+08	5,20E+07
Total	**Ns**	402138,2	665558,4
	Nll	402138,2	409657,3
	ks	34299238	4,57E+07
	kll	34299238	3,26E+07

La tabla al ser la extraída directamente del programa esta escrita en nomenclatura SPSS. Indica para cada grupo los estadísticos de cada una de las variables, siendo Ns el número de paquetes salientes, el Nll el número de paquetes que llegan, el Ks el número de kilos salientes y Kll el número de kilos que llegan.

Estas dos segundas tablas indican los coeficientes estandarizados de las dos funciones discriminantes ya citadas anteriormente. Estas funciones aplicadas a cada terminal dan la puntuación discriminante de cada una de ellas. La tabla 4.4 proporciona el valor de la función en los centroides de cada uno de los grupos.

Tabla 4.3 – *Coeficientes de las funciones discriminantes*

	Coeficientes de las funciones	
	1	**2**
Ns	-0,133	-0,941
Nll	1,524	-2,174
ks	0,704	1,144
kll	-0,927	2,285

Tabla 4.4 – *Funciones discriminantes evaluadas en los centroides de los grupos*

	Función discriminante	
	1	**2**
1	0,943	1,275
2	-1,222	-0,363
3	8,306	-1,014

La tabla 4.5 Muestra los resultados de la clasificación. Proporciona el número de terminales de cada grupo mal clasificadas y el grupo a donde deberían ir.

Tabla 4.5 – *Tabla de probabilidades de pertenencia a un grupo*

			Grupo de pertenencia pronosticado			
		V2	**1**	**2**	**3**	**Total**
Original	**Recuento**	1	13	1	0	14
		2	3	35	0	38
		3	0	0	4	4
	%	1	92,9	7,1	0	100
		2	7,9	92,1	0	100
		3	0	0	100	100

Como se puede observar, el cuerpo de la tabla consta de dos partes. En la parte superior se indica el número de terminales bien y mal clasificadas para cada grupo, el grupo al cual se habían asignado como datos de partida, y el grupo al que deberían pertenecer según el análisis discriminante. Así en las columnas quedan indicados el número de elementos que tenía cada grupo según el análisis cluster y en las columnas se indica el número de elementos que deberían tener teóricamente. La diagonal indica el número de elementos de cada grupo que ambos métodos coinciden que tiene. Los elementos fuera de la diagonal indican el grupo al que en principio se habían asignado y que el análisis discriminante ha corregido. Así para el grupo 1 el análisis cluster considera que tiene 14 elementos, sin embargo el análisis discriminante indica que uno de ellos debería pertenecer al grupo 2. Por otra parte el análisis cluster dice que el grupo 2 tiene 38 elementos y sin embargo el determinante dice que hay 3 de ellos que pertenecen al grupo 1. Ambos métodos coinciden que el grupo 3 tiene 4 elementos.

La segunda parte de la tabla, la parte inferior indica en forma de porcentaje los elementos bien y mal clasificados.

A continuación se adjunta la tabla 4.6 que indica el grupo según el método cluster y el grupo pronosticado por el SPSS. Marcados en color azul son los casos que según el análisis discriminante con nomenclatura SPSS, deberían pasar del grupo 1 al grupo 2, y en naranja están marcados los casos que deberían pasar del grupo 2 al 1. Ésta es la llamada matriz de confusión.

Como se puede observar en la tabla 4.6 las terminales mal clasificadas son: Málaga, La Coruña, Oporto y Valladolid. Para verificar a que grupo deberían ir, se analizarán los clusters por los demás métodos para estas 4 terminales. Si se analizan los otros clusters resultan que los cuatro puntos deberían estar en el grupo mayor. En realidad, éste es un problema debido a que dichos grupos, en vez de resultar compactos y bien definidos, su línea divisoria forma una franja difusa en la cual es complicado predecir con exactitud a que grupo pertenecen las terminales. En este caso se tomará la decisión de dejar los cuatro puntos en el grupo mayor, o sea, el grupo 3 de manera que el único cambio que sufre la clasificación anterior dada sería en la terminal de Málaga. Esa decisión se ha tomado para contrarrestar el efecto compactante y menos realista del método de Ward. Si se hubiera hecho la clasificación por cualquier otro método estas terminales hubieran estado en el grupo 3. Por lo tanto la clasificación resultante final sería la adjunta en la tabla 4.7.

Tabla 4.6 – *Matriz de confusión*

		Grupo mayor				
			P(D>d | G=g)			
Caso	Grupo real	pronosticado	p	gl	P(G=g | D=d)	dist. al centroide
01 - LAZKAO (GUIPÚZCOA)	1	1	0,01	2	1	10,283
03 - CORDOBA	2	2	1	2	0,975	0,002
04 - ALBAIDA (VALENCIA)	2	2	0,81	2	0,893	0,426
05 - MANZANARES (CIUDAD REAL)	2	2	0,96	2	0,984	0,078
08 - ALBACETE	2	2	0,89	2	0,993	0,229
09 - ALGECIRAS (CÁDIZ)	2	2	0,77	2	0,995	0,522
10 - STA. PERPETUA - (BCN)	3	3	0	2	1	12,199
12 -PELIGROS (GRANADA)	2	2	0,85	2	0,925	0,319
13 MENORCA	2	2	0,66	2	0,997	0,819
14 - FARO (PORTUGAL)	2	2	0,66	2	0,998	0,843
15 PALMA	2	2	0,6	2	0,753	1,036
16 IBIZA	2	2	0,62	2	0,996	0,954
17 STA CRUZ DE TENERIFE	2	2	0,58	2	0,997	1,107
20 - DERIO (BILBAO)	1	1	0,13	2	1	4,065
23 - FERROL (LA CORUÑA)	2	2	0,87	2	0,989	0,286
26 - SAN JUAN DEL PUERTO (HUELVA)	2	2	0,86	2	0,994	0,31
27 - LEÓN	2	2	0,95	2	0,988	0,104
29 - COSLADA (MADRID)	3	3	0	2	1	12,148
31 - MÁLAGA	1	2**	0,39	2	0,515	1,895
35 - HUECAS (TOLEDO)	2	2	0,81	2	0,875	0,412
40 - HERNANI (SAN SEBASTIAN)	1	1	0,87	2	0,986	0,291
41 - CASTELLÓN DE LA PLANA (CASTELLON)	2	2	0,71	2	0,903	0,684
42 - SANTIAGO DE COMPOSTELA	2	2	0,97	2	0,974	0,058
49 - JAEN	2	2	0,93	2	0,99	0,138
51 - ZAMORA	2	2	0,69	2	0,998	0,755
52 - RIBAROJA DEL TURIA (VALENCIA)	3	3	0,01	2	1	8,487
55 - PUERTO DE STA. MARÍA (CÁDIZ)	2	2	0,97	2	0,975	0,066
60 - PAMPLONA	1	1	0,2	2	1	3,224
61 - COBATILLAS (MURCIA)	1	1	0,04	2	0,573	6,285
62 - LÉRIDA	2	2	0,82	2	0,98	0,394
63 - IGUALADA (BCN)	1	1	0,5	2	0,989	1,391
65 - SANTANDER	2	2	0,86	2	0,934	0,312
67 - SALAMANCA	2	2	0,9	2	0,993	0,216
70 - VITORIA	1	1	0,58	2	0,992	1,087
73 - SEVILLA	1	1	0,62	2	0,981	0,97
74 - TARRAGONA	2	2	0,53	2	0,912	1,275
75 - MÉRIDA (BADAJOZ)	2	2	0,95	2	0,944	0,101
76 LAS PALMAS	2	2	0,53	2	0,999	1,286
77 LAS PALMAS	2	2	0,58	2	0,997	1,099
79 - CUENCA	2	2	0,65	2	0,998	0,857
80 - EL PRAT (BCN)	1	1	0,88	2	0,958	0,263
81 - BESCANO (GERONA)	2	2	0,26	2	0,746	2,715
82 - LA CORUÑA	2	1**	0,38	2	0,685	1,918
83 - VIGO	1	1	0,44	2	0,555	1,63
85 - LUGO	2	2	0,91	2	0,992	0,194
86 - SAN CIPRIÁN DE VIÑAS (ORENSE)	2	2	0,91	2	0,984	0,192
87 - TORRELLANO (ALICANTE)	3	3	0,02	2	1	8,226
88 - VILA DO CONDE (OPORTO)	2	1**	0,41	2	0,592	1,763
89 - ALVERCA DO RIBATEJO (LISBOA)	1	1	0,48	2	0,91	1,45
90 - SILVOTA (ASTURIAS)	1	1	0,17	2	0,559	3,609
93 - VALLADOLID	2	1**	0,42	2	0,574	1,74
95 - COIMBRA - (PORTUGAL)	2	2	0,95	2	0,989	0,098
96 - LOGROÑO	2	2	0,69	2	0,82	0,756
97 - VILLAFRIA (BURGOS)	2	2	0,75	2	0,862	0,584
98 ALMERIA	2	2	0,82	2	0,994	0,392

Tabla 4.7 – *Clasificación definitiva*

Clasificación Final	KII	Ks	NII	Ns	G
10 - STA. PERPETUA - (BCN)	3352631	1685758	194800397	114619441	3
29 - COSLADA (MADRID)	3032789	2368387	202900535	197140906	3
52 - RIBAROJA DEL TURIA (VALENCIA)	1414019	1346248	128329618	107344395	3
87 - TORRELLANO (ALICANTE)	1885749	1068925	124870074	75002233	3
01 - LAZKAO (GUIPÚZCOA)	410748	216593	84052697	31421453	1
20 - DERIO (BILBAO)	790851	757383	86790781	61561370	1
40 - HERNANI (SAN SEBASTIAN)	478845	390895	64836817	36011992	1
60 - PAMPLONA	490234	394503	86358444	36429051	1
61 - COBATILLAS (MURCIA)	236514	727074	23693624	53422981	1
63 - IGUALADA (BCN)	1384997	352019	103596677	28036058	1
70 - VITORIA	523730	271517	74761502	26455144	1
73 - SEVILLA	354050	794056	32439865	83250153	1
80 - EL PRAT (BCN)	1109183	460335	76491239	43939669	1
83 - VIGO	389164	545442	27840463	52505745	1
84 - SAN JUAN DE MOZARRIFAR (ZARAGOZA)	949743	568668	75030725	51517415	1
89 - ALVERCA DO RIBATEJO (LISBOA)	171420	565529	28867628	59002774	1
90 - SILVOTA (ASTURIAS)	300327	655063	25690221	52404831	1
03 - CORDOBA	76767	325832	5245454	29381992	2
04 - ALBAIDA (VALENCIA)	527010	143553	41239764	10294890	2
05 - MANZANARES (CIUDAD REAL)	106118	234861	9437757	18699087	2
08 - ALBACETE	69557	161855	5166184	13837628	2
09 - ALGECIRAS (CÁDIZ)	25683	119091	2368198	11634856	2
12 -PELIGROS (GRANADA)	220196	391157	14972057	34196166	2
13 MENORCA	21595	74927	919545	7281850	2
14 - FARO (PORTUGAL)	9401	65834	611170	5011190	2
15 PALMA	103381	390405	17404657	37222054	2
16 IBIZA	9532	94489	659493	11534367	2
17 STA CRUZ DE TENERIFE	10882	46327	1402208	6523246	2
23 - FERROL (LA CORUÑA)	79350	111013	9969195	10768398	2
26 - SAN JUAN DEL PUERTO (HUELVA)	27476	180151	1780650	14806497	2
27 - LEÓN	88701	261570	4715280	22362651	2
35 - HUECAS (TOLEDO)	308952	283248	26240160	25420578	2
41 - CASTELLÓN DE LA PLANA (CASTELLON)	216337	336102	21386035	25652460	2
42 - SANTIAGO DE COMPOSTELA	77162	223991	10022660	20490103	2
49 - JAEN	52984	268345	2374890	22779527	2
51 - ZAMORA	7866	89100	294645	6571556	2
55 - PUERTO DE STA. MARÍA (CÁDIZ)	60816	334143	5299162	28434031	2
62 - LÉRIDA	101722	263484	10930527	18425851	2
65 - SANTANDER	227639	345735	17155994	29093676	2
67 - SALAMANCA	38352	188630	2893390	16416473	2
74 - TARRAGONA	223900	364529	21085614	25880397	2
75 - MÉRIDA (BADAJOZ)	103597	335354	10553565	30579620	2
76 LAS PALMAS	11	3450	894	171159	2
77 LAS PALMAS	16545	51765	919040	6952686	2
79 - CUENCA	6592	64552	446016	5142310	2
81 - BESCANO (GERONA)	310231	412696	31057587	28023095	2
82 - LA CORUÑA	326190	414338	33055472	44102590	2
88 - VILA DO CONDE (OPORTO)	404326	481679	35576690	41512155	2
93 - VALLADOLID	170079	333916	32141884	30525719	2
95 - COIMBRA - (PORTUGAL)	77595	256063	3154792	23101329	2
96 - LOGROÑO	485670	260372	37390824	21725560	2
97 - VILLAFRIA (BURGOS)	166036	222848	24516742	20062250	2
98 ALMERIA	71712	86433	7411239	7848232	2
31 - MÁLAGA	161667	678991	12729621	62590589	2

A continuación se pasará al análisis de los diagramas Box-Plot.

Se pasará al análisis de los diagramas de los Box-Plot. En ellos se ve representado en el eje de abscisas el número del grupo al cual pertenecen, en este caso la nomenclatura ha sido cambiada y se utilizará la misma nomenclatura que en los mapas, es decir, el grupo 1 es el grupo formado por las 4 terminales con más volumen de trafico, el grupo 2 por las 7 terminales con flujo intermedio y el tercero por las 39 terminales con menor trafico. V2 es la variable que denomina el grupo y las ordenadas indican el número de paquetes salientes.

BOX PLOT PARA LA VARIABLE SALIDAS EN KILOS:

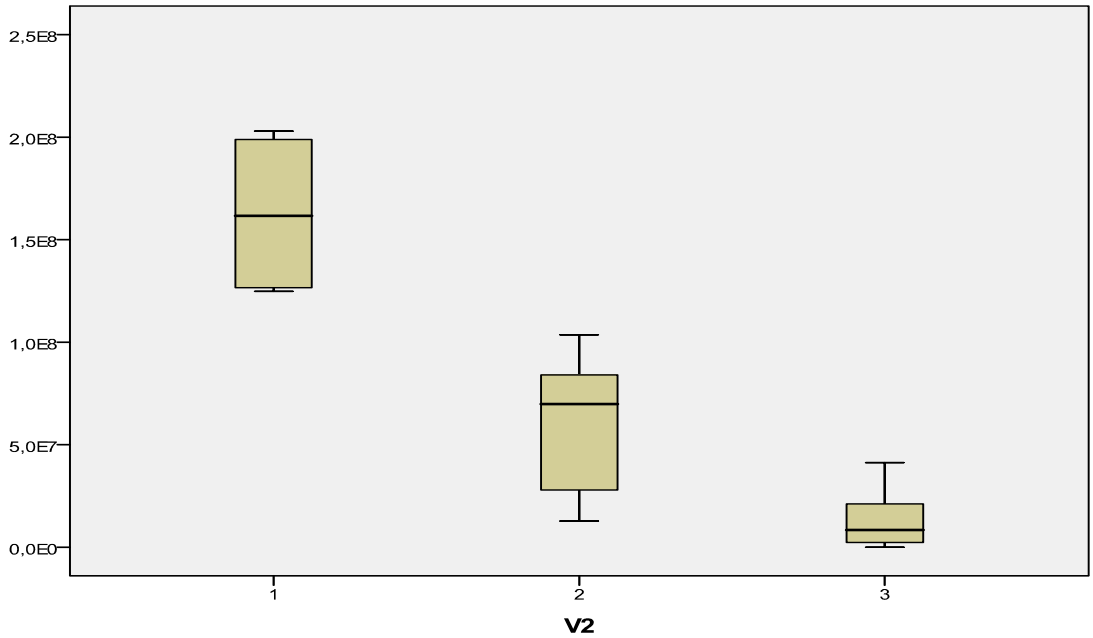


Fig. 4.8- *Gráficos Box Plot para la variable Salidas en kilos*

Se puede observar que respecto a la variable kilos de salida los grupos quedan bien definidos y separados entre si, no existen valores que se superpongan entre los dos primeros grupos pero si lo hay entre el segundo y el tercero. Esta es la causa de que cuando se hicieron los dendogramas la mayoría de los métodos crearán tres grupos en los que los elementos que en este caso están emplazados en el grupo 2 y 3 por los demás métodos estuvieran emplazados en un mismo grupo.

Tabla 4.8. *Estadísticos de cada grupo para la variable salida en kilos:*

Ks	MEDIA	MEDIANA	VARIANZA	D.T	MIN	MAX	ASIMETRÍA
GRUPO 1	1,63E+08	1,62E+08	1,75E+15	4,19E+07	1,25E+08	2,03E+08	2,60E-02
GRUPO 2	5,73E+07	6,98E+07	9,25E+14	3,04E+07	1,27E+07	1,04E+08	-1,20E-01
GRUPO 3	1,23E+07	8,42E+06	1,52E+14	1,23E+07	8,94E+02	4,12E+07	9,65E-01

Los diagramas para los box-plot de las demás variables están adjuntos en el anejo.

BOX PLOT PARA LA VARIABLE NÚMERO PAQUETES SALIENTES

En este caso se puede ver que la diferencia entre el grupo 1 y el grupo 2 ya no es tan marcada como en el caso de la variable anteriormente analizada. Además para esta variable, en el grupo 3 hay un caso que se sale del límite del percentil 75.

Tabla 4.9. *Estadísticos de cada grupo para la variable salida en número de paquetes:*

Ns	MEDIA	MEDIANA	VARIANZA	D.T	MIN	MAX	ASIMETRÍA
GRUPO 1	2421297	2459269	8,48E+11	920638,6	1414019	3352631	-0,119
GRUPO 2	553676,6	44796,5	1,37E+11	369535,7	161667	1384997	1,14
GRUPO 3	133765,2	84025,5	1,84E+10	135452,6	11	527010	1,425

BOX PLOT PARA LA VARIABLE NÚMERO LLEGADAS PAQUETES

En este caso se resuelven a tener un grupo 1 y 2 muy diferenciados entre sí y un grupo 2 y 3 no tan mezclados como en las salidas de número de paquetes pero que a pesar de todo siguen siendo dos grupos separados de forma muy difusa.

Tabla 4.10. *Estadísticos de cada grupo para la variable llegadas en kilos:*

KII	MEDIA	MEDIANA	VARIANZA	D.T	MIN	MAX	ASIMETRÍA
GRUPO 1	1.617.329,50	1.516.003,00	3,14E+11	560.655,12	1.068.925	2.368.387	0,905
GRUPO 2	527.004,86	555.485,50	3,39E+10	184.026,28	216.593	794.056	-0,174
GRUPO 3	228.219,79	237.135,50	1,55E+10	124.373,04	3.450	481.679	0,013

BOX PLOT PARA LA VARIABLE KILOS LLEGADAS

Se puede observar, en este caso la variabilidad del grupo 1 respecto a esta variable y de nuevo los 3 grupos vuelven a tener muchos valores solapados entre si.

Tabla 4.11. *Estadísticos de cada grupo para la variable llegada en número de paquetes:*

NII	MEDIA	MEDIANA	VARIANZA	D.T	MIN	MAX	ASIMETRÍA
GRUPO 1	1,24E+08	1,11E+08	2,71E+15	5,20E+07	75002233	1,97E+08	1,326
GRUPO 2	48467801,8	51961123	2,52E+14	1,59E+07	26455144	83250153	0,465
GRUPO 3	19686872,1	19380668,5	1,16E+14	1,08E+07	171159	44102590	0,295

ANÁLISIS DE LOS GRÁFICOS BOX PLOT

El grupo 1 es el grupo más disperso, esta formado por cuatro terminales localizadas en Barcelona, Madrid, Valencia y Alicante. No tienen relación geográfica, sino que la similitud del volumen de mercancías que mueven podría estar relacionado con la producción y riqueza de la región. Es un grupo formado por terminales muy diversas entre sí, es por ello que los diagramas de Box-Plot son mucho más alargados para este grupo. Son terminales que en un análisis real de comportamiento deberían haberse separado en dos grupos formados por Madrid y Barcelona el primero y Valencia y Alicante el segundo pero que sin embargo, por el objetivo posterior de esta tesina, ha sido necesario agruparlas en un único cluster. La variable para la que presentan una menor variabilidad intragrupo es para la variable salidas de kilos, que por otra parte es la variable más significativa para este estudio. Y la variable en la que presentan una dispersión mayor es en la variable entradas de kilos, y que en este caso se deberá estudiar en el siguiente apartado.

El grupo 2 es en comparación con el primero un grupo mucho menos disperso, a pesar de todo, se puede observar en los diagramas de Box-Plot que los valores de las variables de sus terminales se solapan en algunos casos con los del grupo 3. Esto es debido a que los grupos no tienen una frontera muy marcada, sino más bien una frontera difusa en la que en algunos casos resulta complicado decir con precisión a que grupo pertenecen. Las terminales que forman este grupo son terminales situadas en su mayoría en la zona del País Vasco, en esa región hay una densidad significativa. En este caso puede ser debido a la riqueza de la región pero también al origen de la empresa que se esta analizando. Con lo cual es probable que los motivos sean ajenos al comportamiento de la región. Una de las terminales más aisladas geográficamente de este grupo esta en Cádiz, y la razón puede ser la importancia del transporte marítimo debido al puerto.

El grupo 3 es el grupo más grande con más del 75 % de los casos. Es un grupo muy compacto y con una variabilidad muy pequeña. Los comportamientos de las terminales para las distintas variables son muy similares, presentando más variabilidad para la variable de entradas en kilos. Otro valor significativo es el valor de la variable salidas de paquetes para la terminal de Albaida. Esta es la que presenta una variabilidad mayor,

pudiéndose considerar perteneciente al grupo 2, sin embargo, como respecto a las demás variables está dentro del comportamiento del 50 % se considerará bien clasificada.

4.2 ANÁLISIS DE LA ESTACIONALIDAD DEL SISTEMA DE TERMINALES.

Los resultados empíricos del análisis de la estacionalidad de las terminales son los siguientes. Los resultados se presentan en forma de distribución acumulada de probabilidad de CV analizando las terminales individualmente y en concreto los flujos de salida, tanto de kilos como número de paquetes. La figura a continuación representa dicha distribución.

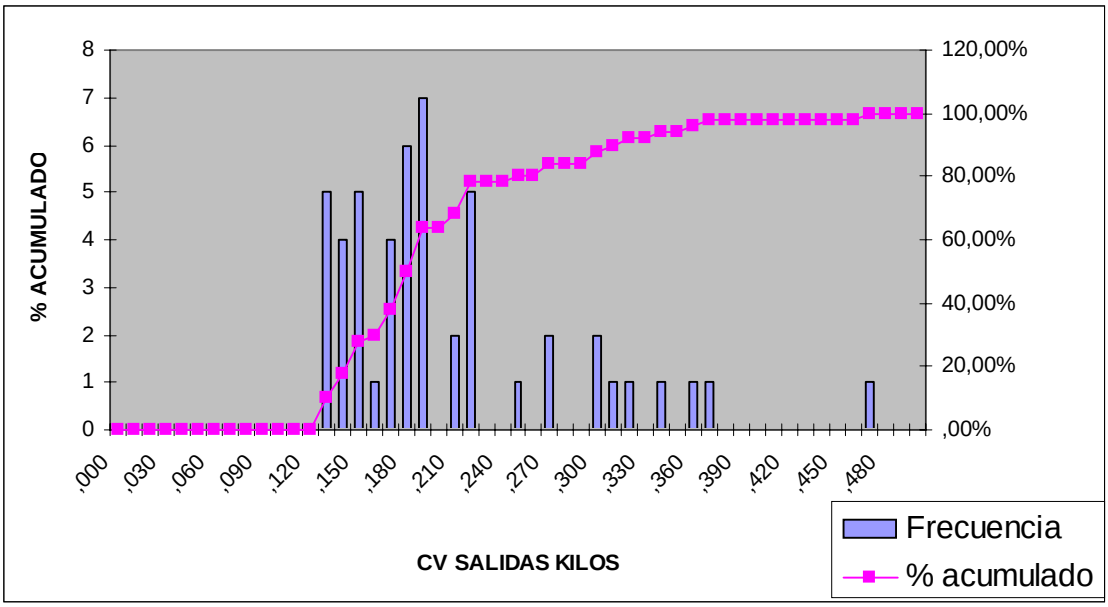


Fig. 4.9- *Probabilidad acumulada de la variable cv de salidas en kilos*

La distribución muestra la proporción de terminales para las cuales el CV esta por debajo del nivel indicado en el eje x. Como se puede apreciar en la grafica, un 85% de las terminales tienen un CV inferior a 0,3 eso significa que las terminales tienen un comportamiento en lo que a salidas de kilos respecta, muy poco variable a lo largo del tiempo. Se compara ahora con la gráfica de distribución acumulada del número de piezas salientes.

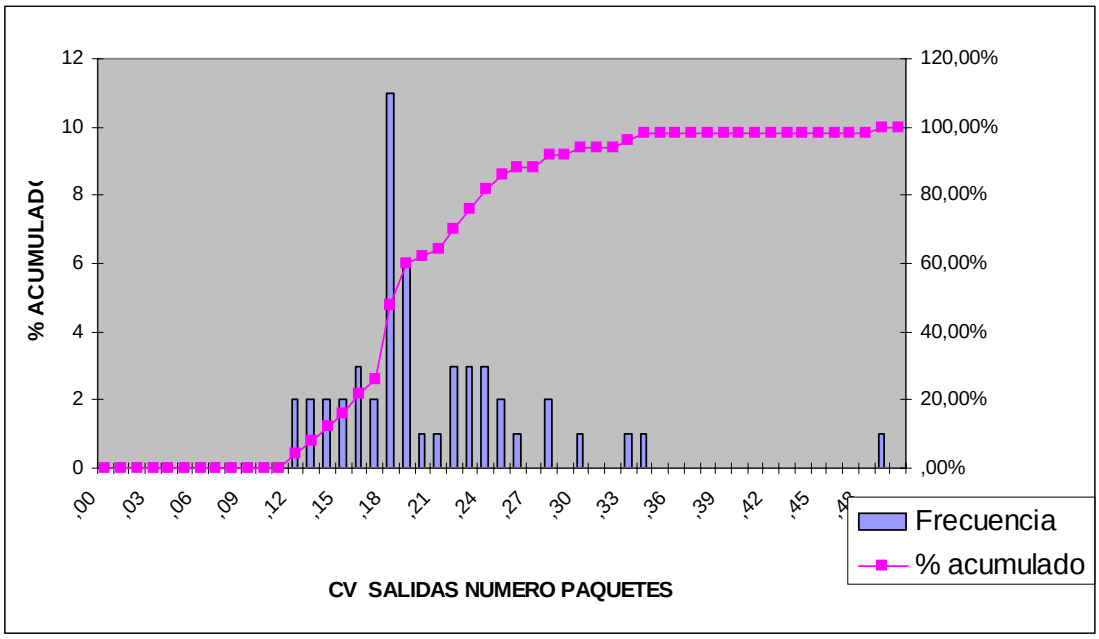


Fig. 4.10. *Probabilidad acumulada de la variable CV de salidas en número de paquetes*

En este caso la variabilidad es ligeramente menor. En este caso un 87% de los casos tiene un CV inferior a 0,3. Se analizará ahora la variabilidad de las salidas de paquetes, en kilos y en piezas.

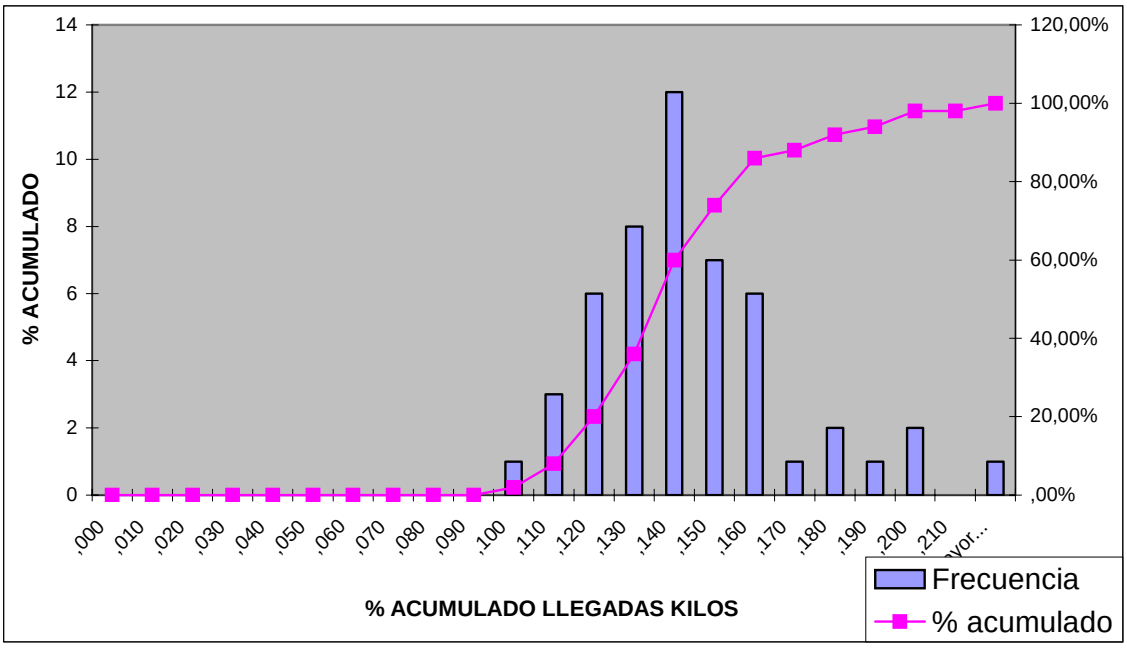


Fig. 4.11. *Probabilidad acumulada de la variable CV de llegadas en kilos*

Se puede observar que la variabilidad de las salidas es también muy pequeña tanto para kilos como para número de paquetes. Sus gráficas son prácticamente idénticas y toman también los mismos valores. En cualquier caso inferiores a 0,2 en prácticamente un 100% de los casos.

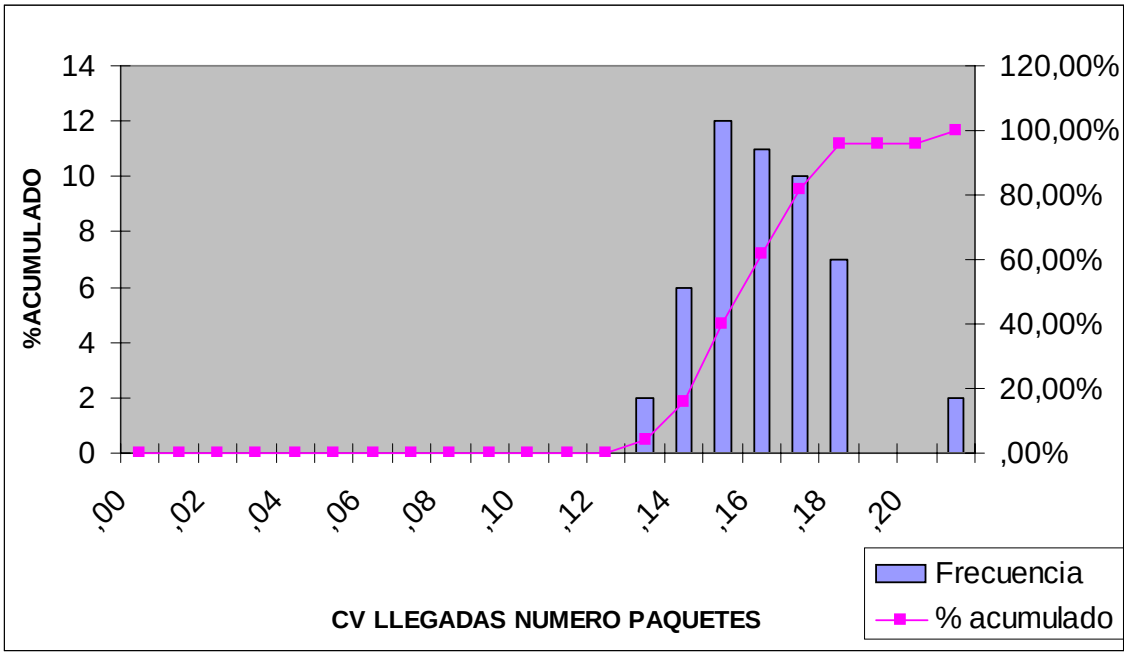


Fig.4.12. *Probabilidad acumulada de la variable cv de llegadas en numero de paquetes*

Esta similitud entre las distintas variables era de esperar ya que existe una cierta correlación, es por ello, que se decide, a partir de este momento, trabajar con una única variable que sea representativa de las dos. Se escoge la variable kilos por ser la más significativa en el ámbito de la paquetería y en concreto se analizarán las salidas y el flujo neto que se definirá a continuación.

Si queremos trabajar con flujo neto en vez de trabajar con entradas y salidas separadamente, se calcula lo que llamaremos *coeficiente de variación de flujo neto* que se definirá como:

$$NCV_j = \sigma_{nj} / [\lambda_j + \mu_j] \qquad (4.1)$$

La gráfica de probabilidad acumulada del coeficiente de variación neto es:

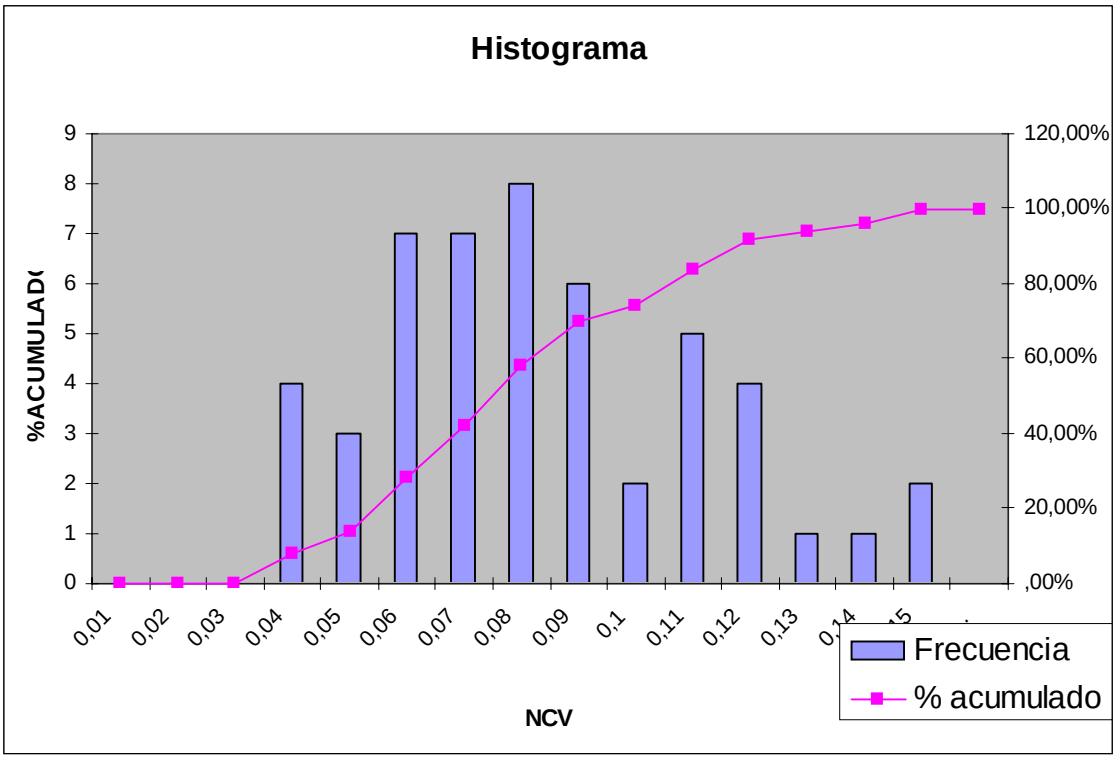


Fig. 4.13. *Probabilidad acumulada de la variable de coeficiente de variación de flujo neto en kilos*

Esta gráfica es significativamente diversa a las anteriores. En este caso se observa que el 100% de las terminales tiene un NCV inferior a 0,15, inferior a si se analizan por separado las salidas y las llegadas.

Se ha hecho el mismo análisis para los 3 grupos de terminales que obtuvimos del análisis cluster, y los resultados son los siguientes:

Tabla 4.12- *Estadísticos de la variable CV*

	Terminales individuales		**Grupos de terminales**	
CV	**Media**	**DS**	**Media**	**DS**
Peso	0,2011	0,0736	0,1265	0,0026

Tabla 4.13- *Estadísticos de la variable NCV*

	Terminales individuales		**Grupos de terminales**	
NCV	**Media**	**DS**	**Media**	**DS**
Peso	0,1464	0,033	0,1219	0,0038

Como era de esperar si se analiza por grupos de terminales, estos presentan una menor variabilidad. Esto es debido a que las irregularidades de las terminales de los grupos se compensan entre ellas, dando como resultados unas curvas mucho mas suaves. También se puede comprobar que analizando el coeficiente de variación neto, se observan los mismos resultados, al suponerlo como la suma de flujos las irregularidades de ambos se compensan. Las terminales con mayores medias tienden a tener una mayor varianza y sin embargo un CV de variación menor. Se quiere ahora obtener una relación para predecir el comportamiento de CV en base al flujo medio.

Antes de realizar el ajuste mediante logaritmos se verá qué aspecto tiene el gráfico de dispersión, éste puede dar una idea a groso modo del tipo de relación que seguirán las variables. Se adjunta a continuación la gráfica de regresión lineal para el conjunto de terminales.

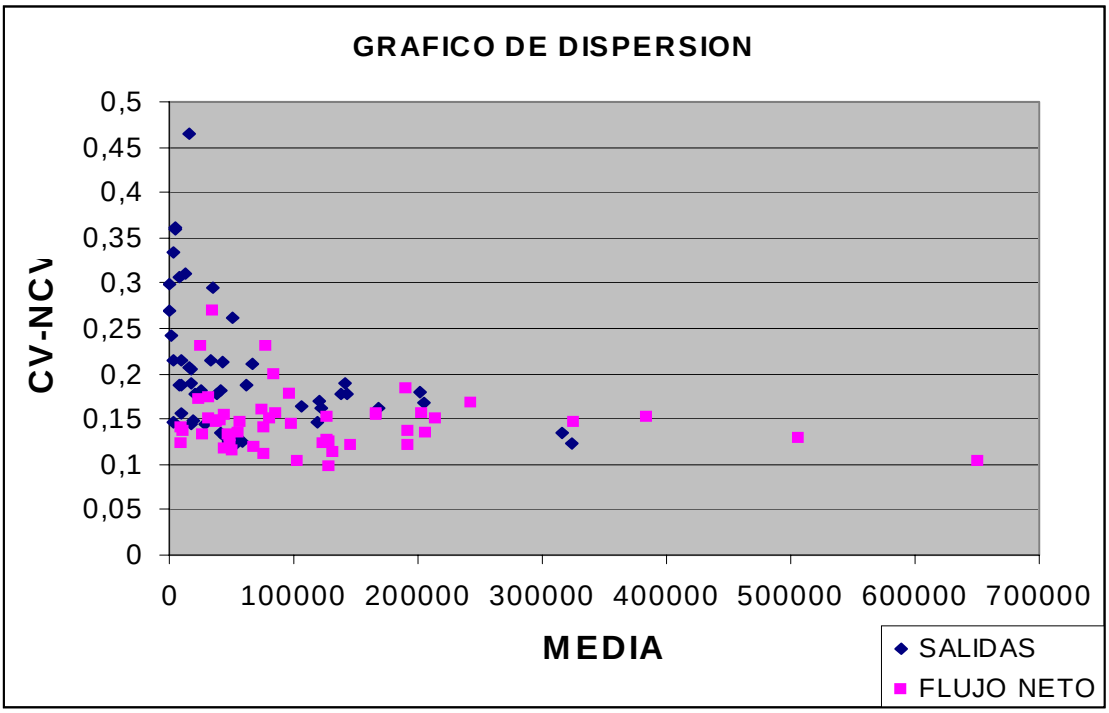


Fig. 4.14. *Diagramas de dispersión para el coeficiente de variación del flujo neto y para el coeficiente de variación de los kilos salidos.*

Se puede observar que la distribución no es una nube de puntos pero tampoco tiene una tendencia clara. También cabe destacar quelas escalas de los ejes son muy dispares. Se prueba en un primer principio con una relación logarítmica como se hace en Hall

(1998). Se hipotizará que la variación del flujo de salida es una exponencial de la media que sería equivalente a decir:

$$\text{Ln}\,[CVj] = a + b\,\ln\,[\,\lambda j] \qquad (4.2)$$

Los resultados obtenidos por medio de la definición de los coeficientes de la recta de regresión vienen adjuntados en la tabla a continuación:

Tabla 4.14- *Coeficientes de la recta de regresión*

b	-0,1155
a	-0,4802
R^2	**0,2959**

La gráfica de dispersión con su respectiva recta de regresión para una relación exponencial sobre el logaritmo de la media es la siguiente:

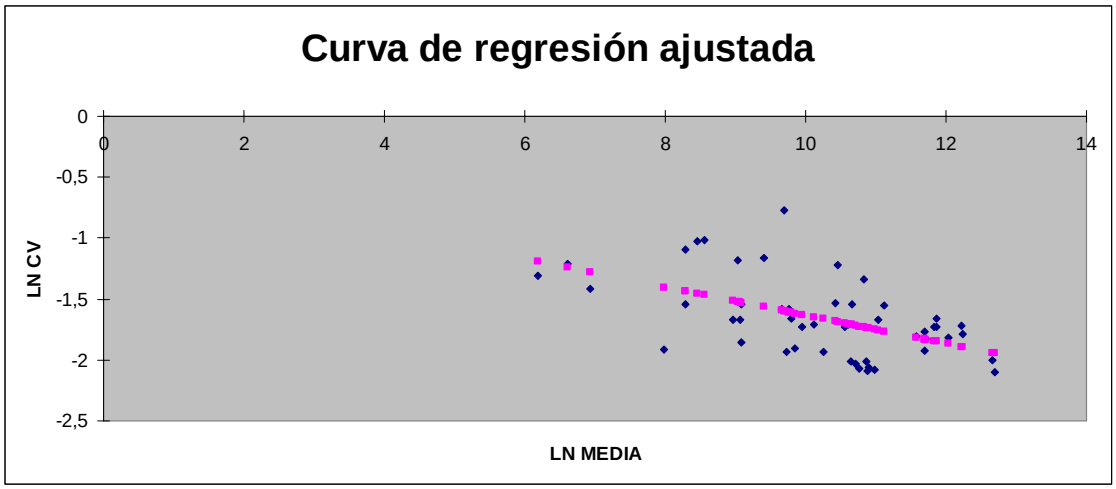


Fig. 4.15. *Recta de regresión para las variables Ln CV y Ln λj*

El coeficiente R^2 es muy pequeño en este caso y no se puede aceptar esta relación como buena. Será menester por lo tanto encontrar otra función que relacione las variables. En el anejo se encuentra la tabla que proporciona los valores obtenidos mediante las distintas funciones y se observa que la relación que proporciona un mejor resultado es la relación polinómica de sexto grado. A pesar de ello el valor es inferior a 0,4. Ese valor tampoco es aceptable como buen ajuste y la razón de mejora no es significativa como para tenerse en cuenta en este estudio. Es por ello que se desestima la

posibilidad de que ambas variables sigan una relación lógica que permita la predicción de su comportamiento.

Se realiza el correspondiente análisis en este caso para el flujo neto. Los resultados obtenidos no son mejores que para el flujo de generaciones.

Tabla 4.14. *Coeficientes de la recta de regresión*

b	-0,0367
a	-1,5279
R^2	**0,0289**

Vemos que el coeficiente de regresión es aun peor que en el caso de generaciones. En este caso la figura representa una nube de puntos que no sigue ninguna tendencia.

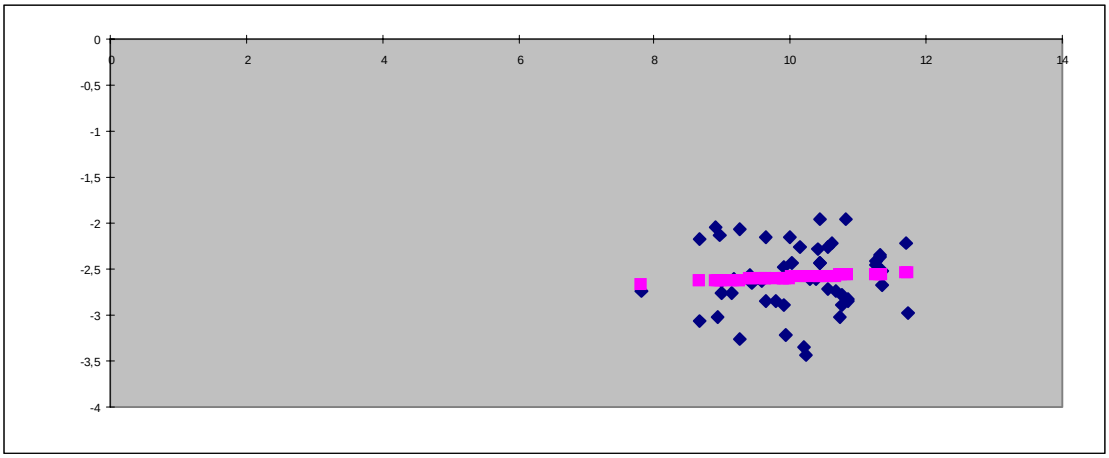


Fig. 4.16. *Recta de regresión para las variables Ln NCV y Ln λj+μj*

Como se puede observar tanto en el gráfico de dispersión como en el análisis de bondad, este ajuste no se puede aceptar como bueno. Después de probar otras relaciones adjuntas en la tabla del anejo 5 se concluyó en que ambas variables no cumplen aparentemente ninguna relación clara.

La razón por la cual se están obteniendo estos resultados tan dispares a los realizados a partir del estudio de Hall pueden ser varias. La primera es que como se ha explicado anteriormente, los datos son muy incompletos. Por una parte se trabaja con 50 terminales frente a las 400 del estudio americano. Y los datos no son diarios sino que se

ha trabajado con una media diaria a partir de un dato mensual. Por otra parte la variabilidad de esta red de terminales es mayor que la variabilidad de la red con la que ha trabajado Hall.

Por último en este estudio de la estacionalidad se proporciona los ajustes del comportamiento de las distintas medidas de flujo a determinados modelos, para poder predecir el comportamiento de la red de distribución. El primero es el modelo de comportamiento del flujo neto entendido como atracciones más generaciones. Según Estrada (2007) la función del flujo neto se ajusta a un modelo de Weibull. El ajuste obtenido es el siguiente:

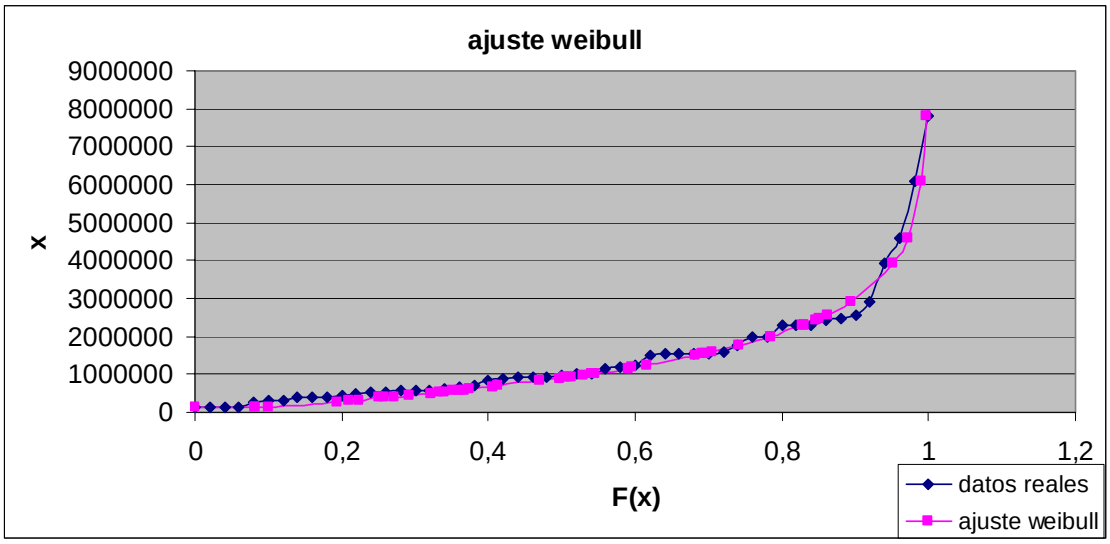


Fig. 4.17. *Ajuste a una distribución de Weibull.*

Los coeficientes del ajuste de Weibull son los siguientes:

Tabla 4.15. *Parámetros de la distribución de Weibull*

θ2	1300000
α2	1

Se analiza la bondad del ajuste por medio de la recta de regresión y se obtiene el siguiente resultado:

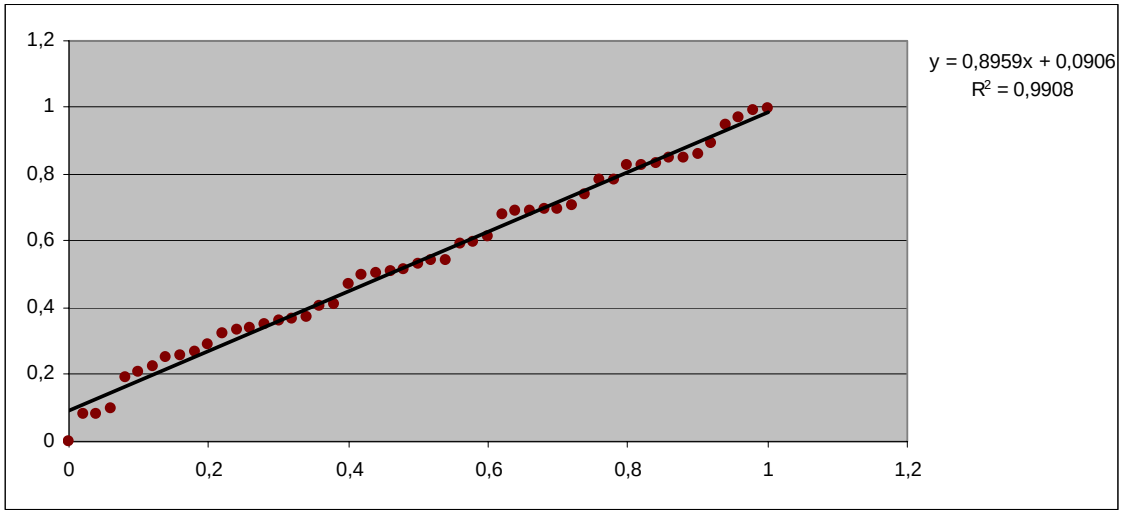


Fig. 4.18. *Análisis de la bondad del ajuste a una distribución de Weibull.*

Se puede comprobar por el coeficiente R^2 que es de 0,99 por lo tanto se puede afirmar que el ajuste es muy bueno.

En segundo lugar se analizarán los resultados obtenidos del modelo para el flujo de salidas. Se parte de los análisis de Hall (1998) en los cuales se ajusta para una empresa de paquetería estadounidense, el flujo de salidas a una distribución Gamma. Partimos pues de la hipótesis que para los datos actuales también seguirá una distribución Gamma y efectivamente el resultado del ajuste es bueno.

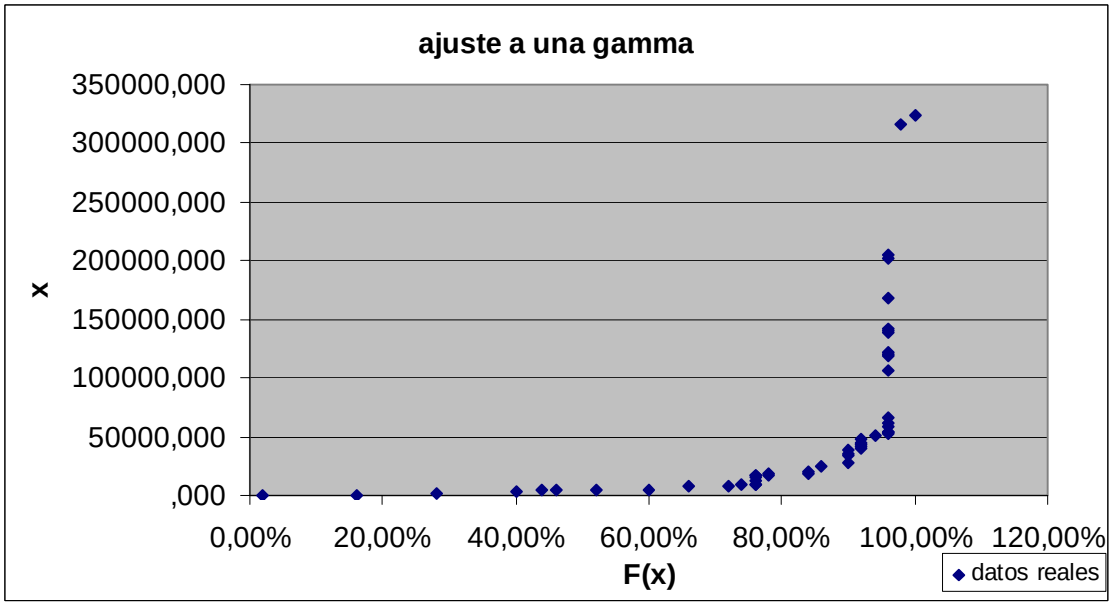


Fig. 4.19. *Ajuste a una distribución de Gamma.*

La tabla 4.13 indica los parámetros para la distribución.

Tabla 4.16. *Parámetros estimados de la distribución*

Parámetros de distribución estimados		
		MEDIA s
Distribución Gamma	Forma	0,644
	Escala	0

En la figura 4.22 se puede observar la relación que anteriormente se ha obtenido para el flujo neto mediante la herramienta de análisis de datos de excel, ahora mediante SPSS se puede observar que efectivamente los valores siguen una recta, definiendo así la distribución gamma. La figura esta adjunta en el anejo.

Por último se estudiará el comportamiento del cociente entre flujo de entradas y salidas. Se analiza el cociente para cada una de las terminales para ver que terminales sufren de desequilibrios. Al observar los resultados obtenidos se puede ver que la gran mayoría los sufre. (Obsérvese la tabla 4.14)

Tabla 4.16 *Tabla del cociente entre flujo de entradas y salidas para cada una de las terminales.*

	ENTRADAS	SALIDAS	COCIENTE
01 - LAZKAO (GUIPÚZCOA)	52472,11	138337,79	**0,379**
03 - CORDOBA	49087,68	8735,064	**5,62**
04 - ALBAIDA (VALENCIA)	17159,66	67045,462	**0,256**
05 - MANZANARES (CIUDAD REAL)	31211,43	15685,192	**1,99**
08 - ALBACETE	22996,07	8339,781	**2,757**
09 - ALGECIRAS (CÁDIZ)	19434,94	3952,273	**4,917**
10 - STA. PERPETUA - (BCN)	190492,4	315482,31	**0,604**
12 -PELIGROS (GRANADA)	57093,15	24628,595	**2,318**
14 - FARO (PORTUGAL)	8375,85	1017,533	**8,232**
20 - DERIO (BILBAO)	102683,92	140954,66	**0,728**
23 - FERROL (LA CORUÑA)	17990,23	16241,732	**1,108**
26 - SAN JUAN DEL PUERTO (HUELVA)	24590,5	2952,263	**8,329**
27 - LEÓN	37369,05	7795,273	**4,794**
29 - COSLADA (MADRID)	327287,43	323670,74	**1,011**
31 - MÁLAGA	103390,81	20787,654	**4,974**
35 - HUECAS (TOLEDO)	42401,42	42941,709	**0,987**
40 - HERNANI (SAN SEBASTIAN)	59958,4	106545,7	**0,563**
41 - CASTELLÓN DE LA PLANA (CASTELLON)	42726,12	35103,463	**1,217**
42 - SANTIAGO DE COMPOSTELA	34233,37	16688,881	**2,051**
49 - JAEN	38074,9	3951,831	**9,635**
51 - ZAMORA	10978,69	483,169	**22,722**
52 - RIBAROJA DEL TURIA (VALENCIA)	178673,72	205240,16	**0,871**
55 - PUERTO DE STA. MARÍA (CÁDIZ)	47488,97	8824,901	**5,381**
60 - PAMPLONA	60815,69	142331,19	**0,427**
61 - COBATILLAS (MURCIA)	89209,86	38537,53	**2,315**

	ENTRADAS	SALIDAS	COCIENTE
63 - IGUALADA (BCN)	46351,3	168518,13	**0,275**
65 - SANTANDER	48592,79	28223,965	**1,722**
67 - SALAMANCA	27401,6	4731,779	**5,791**
70 - VITORIA	44166,57	122061,24	**0,362**
73 - SEVILLA	139071,08	53320,895	**2,608**
74 - TARRAGONA	42687,41	33504,737	**1,274**
75 - MÉRIDA (BADAJOZ)	51086,68	17504,881	**2,918**
79 - CUENCA	8590,93	743,099	**11,561**
80 - EL PRAT (BCN)	72771,45	119210,03	**0,61**
81 - BESCANO (GERONA)	46026,85	50766,356	**0,907**
82 - LA CORUÑA	73642,59	53849,054	**1,368**
83 - VIGO	87680,65	44832,749	**1,956**
84 - SAN JUAN DE MOZARRIFAR	85880,37	120297,69	**0,714**
85 - LUGO	28685,9	8867,201	**3,235**
86 - SAN CIPRIÁN DE VIÑAS (ORENSE)	30786,37	18808,597	**1,637**
87 - TORRELLANO (ALICANTE)	123842,33	201698,86	**0,614**
88 - VILA DO CONDE (OPORTO)	69347,91	58916,201	**1,177**
89 - ALVERCA DO RIBATEJO (LISBOA)	98610,42	47624,161	**2,071**
90 - SILVOTA (ASTURIAS)	87528,28	41615,577	**2,103**
93 - VALLADOLID	50978,86	52701,809	**0,967**
95 - COIMBRA - (PORTUGAL)	38517,13	5238,313	**7,353**
98 - ALMERÍA	13118	12247,706	**1,071**

Menos de un 50% de las terminales tienen un cociente entre 0,5 y 2 eso significa que la mayor parte de las terminales están desequilibradas normalmente. La repercusión de ello será que continuamente necesitarán rebalanceos para ellas.

Si se analiza el comportamiento de dicho estadístico, en este caso para grupos de terminales, se puede comprobar que efectivamente, la agrupación de terminales reduce los extremos.

La hipótesis inicial es que si una terminal, analizando flujos medios, está desequilibrada, ya sea por déficit o por exceso entonces se considerará continuamente desequilibrada.

Al realizar el estudio del comportamiento se observó que como en el estudio previo de Hall el modelo que seguía el cociente era un modelo Lognormal.

A continuación en la figura 4.24 se ve representada la función de distribución acumulada.

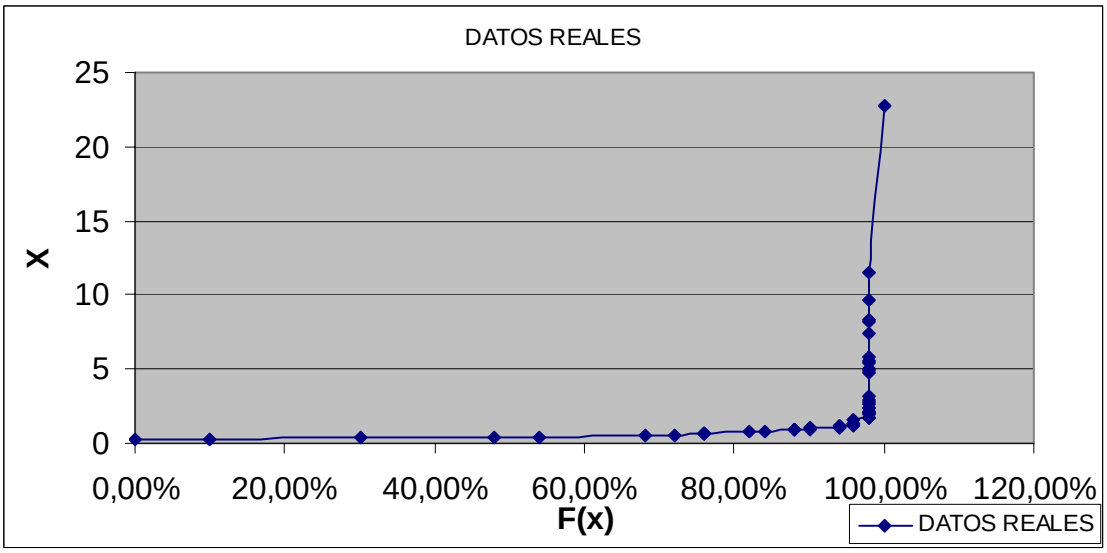


Fig. 4.20. *Función de distribución acumulada para el cociente entre kilos salientes y kilos entrantes*.

De nuevo se ha utilizado la herramienta de P-P para el análisis obteniéndose los siguientes resultados adjuntos en el anejo.

Tabla 4.17. *Parámetros del ajuste a una distribución Lognormal.*

Parámetros de distribución estimados		
		cociente
Distribución Lognormal	Escala	1,712
	Forma	1,041

En la figura 4.25 podemos observar que efectivamente el cociente entre llegadas y salidas sigue una línea recta por lo tanto se acepta como ajuste a una distribución lognormal. A pesar de todo y como ya se explicó en el apartado de metodología los tres métodos se someterán al análisis de Kolmogorov-Smirnov para el análisis de bondad de dichos ajustes.

En este punto, se considera de relevancia la distribución geográfica de las terminales en función de su flujo neto. Como en este caso trabajamos con una única variable no resulta necesaria la clasificación mediante el análisis cluster. La clasificación en función del flujo neto se adjunta a continuación.

Fig. 4.21 – *Distribución geográfica de las terminales en función del flujo neto, las unidades son de masa.*

La figura 4.26 representa en degradación de colores y según el tamaño de la marca las terminales que tienen más salidas que entradas y por lo tanto en las que habrá un déficit de camiones. La representación a partir del mapa ayudará a la hora de hacer el balance de las rutas para minimizar los movimientos en vacío.

Por último se analizará el problema de movimientos en vacío. Como se dijo anteriormente, el flujo total de camiones vacíos está limitado inferiormente por $\sum_j |e_j| / 2$. Este estadístico iguala el 13,3% del flujo total de cargas en la red de terminales individuales, pero solo el 8,2% en la red de grupos de terminales. Estos valores indican que a largo término los desequilibrios son mayores en una red local que regional, a pesar de que en la regional continúan siendo muy significativos. Nos referimos como red regional a la red de grupos de terminales.

El exceso de camiones vacíos necesarios fue calculado para períodos diarios, en función de la variable peso, para terminales individuales y para grupos de terminales. Los

excesos obtenidos son un 60% mayores para terminales individuales que para grupos de terminales.

4.3 OPTIMIZACIÓN DE LA RED DE DISTRIBUCIÓN

Se estudian ahora los resultados obtenidos de la aplicación de los algoritmos para optimizar la asignación de la flota y las rutas. Las variables que definen el problema para cada caso son: el número de terminales y su localización, el número de *hubs* y su localización (en el caso de que la red no tenga ya un número de *hubs* definidos se tomará un número igual a la raíz del número de terminales del sistema), el horario de apertura y cierre de las terminales, la velocidad media de los transportes y el número de iteraciones no factibles que se le permiten al programa.

La batería de problemas realizada para el primer algoritmo consiste en un conjunto de 12 problemas. Primero se trabajó con los problemas de menor número de terminales. En concreto, el primero fue realizado con las terminales de la región de levante de la península. Es un conjunto formado por un total de 24 terminales. Se aplicó el algoritmo para 200 iteraciones, para 1000 iteraciones y para 4000 iteraciones obteniéndose la siguiente tabla de costes:

Tabla 4.18. *Tabla de costes según el número de iteraciones para la zona de Levante.*

ITERACIONES	COSTE COMPUTACIONAL segundos	COSTE FINAL euros
200	19	22235
500	35	21013
1000	66	20499
2000	140	20210
3000	215	20107
4000	291	19964

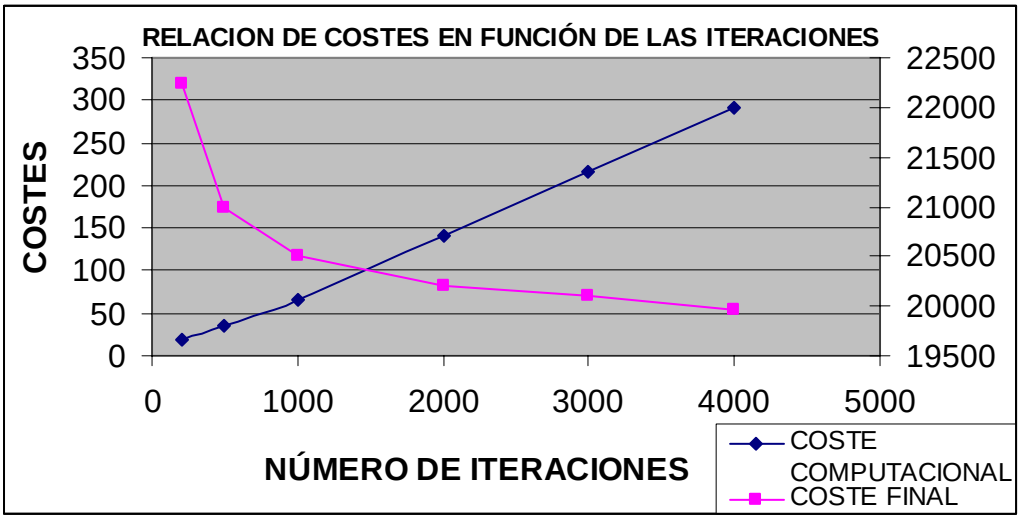


Fig 4.22 – *Relación entre el coste computacional y el coste de la red en función del número de iteraciones.*

Como se puede observar en la figura 4.27 a medida que se aumenta el número de iteraciones el coste computacional aumenta de manera lineal mientras que el coste computacional disminuye de manera exponencial.

A continuación se aplicó el mismo algoritmo para un sistema de 16 terminales con dos tipos de demanda distinta. De nuevo para un problema de 45 terminales con dos demandas distintas, para la región norte de la península y finalmente para el sistema entero de terminales con el que se trabaja. Las tablas de costes proporcionadas a continuación son los costes tanto computacionales como de organización de la red para 1000 iteraciones.

Tabla 4.19. *Tabla de resultados de la aplicación del algoritmo de un día para la batería de problemas.*

NUM. TERMINALES	COSTE COMPUTACIONAL segundos	COSTE FINAL euros
16 a	28	55135,3
16 b	15	19747,46
levante	66	20498,8
norte	448	109147,2
45 a	394	138949,2
45 b	347	80679,76
88 a	696	118966,2
sistema completo a	745	559657,2
sistema completo b	729	536038,9

La indicación a y b se refiere a sistemas de terminales con distintas demandas para las mismas terminales. Así *16 a* significa que el sistema consta de 16 terminales a las cuales se les ha asignado la demanda a correspondiente al día *D* y *16 b*, sería el mismo sistema de terminales con una demanda b correspondiente al día *D+1*. Esta distinción se hace porque a continuación se aplicará el algoritmo para dos días consecutivos con demanda distinta entre los dos días.

En el caso de que no se aplicara el segundo algoritmo los costes añadidos a la suma de costes de la organización del día 1 y del día 2 estaría marcada por los camiones contratados que deberían hacer a la vuelta un desplazamiento en vacío. Por lo tanto se analizará el beneficio que me aporta la aplicación del segundo algoritmo. En este caso se trabajará con los problemas de 16 terminales y de 45 terminales para una demanda constante y para una demanda variable. Los resultados se adjuntan en la tabla 4.18.

Tabla 4.20. *Costes de la resolución del problema de backhauling sin la aplicación del algoritmo de optimización*

NUM. TERMINALES	COSTE 1 euros	COSTE 2 euros	INCREMENTO COSTE euros	COSTE FINAL SIN ALGORIMTO
16	55136	19748	17000	91884
16 CTE	54781	54781	6750	116312
45	155126	89643	36000	280769
LEVANTE	62234	64629	8250	135113
NORTE	109147	114757	10000	233904

Tabla 4.21. *Tabla de resumen del beneficio económico que implica la aplicación del algoritmo.*

NUM. TERMINALES	COSTE FINAL SIN ALGORIMTO	CON ALGORITMO DE DOS DÍAS	DIFERENCIA euros	MEJORA %
16	91884	84682	7202	8.5
16 CTE	116312	39623	76689	65,9
45	280769	251712	29057	11.2
LEVANTE	135113	82347	52766	39,1
NORTE	233904	149472	84432	36,1

Como se puede observar en las tablas adjuntas, el porcentaje mejorado es muy significativo consiguiéndose una reducción de costes en general muy elevada. A pesar de todo, los porcentajes obtenidos son muy dispares, la razón de que sea así es que para los casos teóricos nombrados como 16 y 45 la variabilidad de la demanda entre ambos días era muy elevada. Se debe tener en cuenta que eran los casos teóricos para la calibración del modelo y por lo tanto distaban de ser casos reales. Si se leen esos casos con detenimiento se puede constatar que el porcentaje de camiones que se han debido parar es muy elevado, es por ello que ambos resultados son muy similares. En el caso de nombrado como 16 CTE se tomó exactamente la misma demanda para ambos días, de ahí que el coste 1 sea igual al coste 2. Por esa razón el porcentaje de mejora es mucho más elevado. El número de camiones que se deben parar es muy pequeño ya que al ser la misma demanda se necesita el mismo número de camiones. Sólo son los camiones que estaban realizando rutas que pertenecientes a zonas con flujos muy orientados los únicos que no se podrían recolocar fácilmente. Por lo que respecta a los resultados del análisis de la zona de Levante y de Norte los resultados son también muy buenos, eso es debido a que al ser casos reales la demanda entre un mes u otro no varía de forma excesiva. En este caso se tomó la demanda del mes de enero y del de febrero, sin embargo si se hubiera tomado la del mes de abril y agosto se hubiera notado un decrecimiento de la mejora mucho más elevado ya que en el mes de agosto la demanda decrece considerablemente.

CAPÍTULO 5

CONCLUSIONES

La optimización de las flotas de camiones ha sido un tema principal de investigación en lo que a temas de investigación en organización de recursos respecta. En concreto, lo ha sido el problema de los movimientos en vacío. En cualquier caso, el elemento estocástico más importante en el estudio de paquetería es la demanda de equipamiento, La naturaleza estocástica de la variable está incorporada en el modelo como el valor esperado y su varianza entre unos orígenes y destinos dados. Esta naturaleza estocástica se refleja físicamente en el número de vehículos disponibles en un determinado punto. Se ha podido comprobar que los efectos de dicha aleatoriedad están mitigados mediante la agrupación de las terminales y a lo largo del tiempo. Estas agrupaciones pueden ser muy diversas dando en consecuencia resultados distintos, que lejos de ser contradictorios proporcionan una visión más amplia del comportamiento de las terminales.

En este caso se realizó una clasificación en función de 4 variables mediante un análisis cluster. En ese estudio se comprobó que esta forma de clasificación de las terminales no resultaba la más adecuada para el análisis de su comportamiento salvo en el caso de disponer de un número de casos elevados. En especial en el caso de que dicha agrupación sea el objeto de un futuro análisis. Para ese propósito sería más adecuado,

por la naturaleza de los datos, una agrupación geográfica en la que el número de casos por grupo sea más homogéneo o más significativo para el análisis posterior.

El análisis se realizó mediante el método de Ward utilizando la distancia euclidea. Este método no proporciona un resultado físicamente más realista sin embargo proporcionaba una distribución más compacta de las terminales. De dicho análisis cluster resultaron 3 grupos en función del flujo de mercancías recibidas y enviadas.

Se han podido analizar correctamente dos de los tres clusters obtenidos, obteniéndose así unos resultados de predecibilidad similares a los obtenidos por Hall.

Se observó que el 85% de las terminales tenía un CV inferior a 0,3 eso significa que la variabilidad es muy pequeña. Se analizó el comportamiento de este estadístico tanto para terminales individuales como para grupos de terminales y también para una variable nueva llamada flujo neto que es la suma de las entradas y las salidas. Resultó que el coeficiente de desviación estándar para las terminales era un orden de magnitud mayor que en el caso de analizarlo mediante la agrupación de las terminales y en el caso del coeficiente de variación neto sucedía de igual manera.

Se puede observar que la variabilidad de las salidas es también muy pequeña tanto para kilos como para número de paquetes. Sus graficas son prácticamente idénticas y toman también los mismos valores. En cualquier caso inferiores a 0,2 en prácticamente un 100% de los casos.

Por otra parte se observa que el 100% de las terminales tiene un NCV menor de 0,15, inferior a si se analizan por separado las salidas y las llegadas. Para el coeficiente de variación neto, se observan los mismos resultados, al suponerlo como la suma de flujos las irregularidades de ambos se compensan. Las terminales con mayores medias tienden a tener una mayor varianza y sin embargo un CV de variación menor.

Los resultados, son por lo tanto muy significativos, se concluye con que la variabilidad del sistema de terminales es menor si se consideran grupos de terminales a si se consideran las terminales individualmente. Esto es debido a que las irregularidades de

las terminales de los grupos se compensan entre ellas, dando como resultados unas curvas mucho más suaves.

Por lo que a los modelos respecta, a diferencia de lo que se esperaba, el coeficiente de variación no sigue una relación respecto de su media y sin embargo no se ha podido hallar ninguna relación asimilable. La razón por la cual se están obteniendo estos resultados tan dispares a los realizados a partir del estudio de Hall pueden ser varias. La primera es que como se ha explicado anteriormente, los datos son muy incompletos. Por una parte se trabaja con 50 terminales frente a las 400 del estudio americano. Y los datos no son diarios sino que se ha trabajado con una media diaria a partir de un dato mensual. Se analizan las funciones de distribución acumuladas y se concluye que la de la variable del flujo neto se ajusta a un modelo Weibull, se comprueba obteniendo un coeficiente R^2 de 0,99.

La función de distribución acumulada de la variable salidas en kilos se ajusta a un modelo Gamma y la variable cociente entre generaciones y atracciones se ajusta a un modelo Log normal.

En la optimización de la red se obtiene para 1000 iteraciones una solución optimizada muy buena y en la aplicación del segundo algoritmo para los problemas teóricos en los que la demanda varia en un 50% se consigue ahorrar alrededor de un 9% y para los problemas reales alrededor de un 32%. En consecuencia si se mantiene la demanda entre los dos días la mejora por la aplicación del segundo algoritmo alcanza el 65%. Se concluye pues, que este algoritmo puede aumentar significativamente los beneficios de las empresas de transporte.

REFERENCIAS BIBLIOGRÁFICAS

Canavos, G. (1988), *Probabilidad y Estadística; aplicaciones y métodos*. Ed. McGraw Hill, México.

Cuadras, C. M. (1981), *Métodos de análisis multivariante*. Ed. Eunibar, Barcelona.

Compés, L., 2005 From the Firm to the Market. The Vertical Desintegration of Land Transportation in Spain. Revista de Economía Institucional, vol.7 num.13

Du, Y.F., Hall, R,. 1997. Fleet sizing and empty equipment redistribution for center-terminal transportation networks. Management Science 43. 145-157

Estrada, M. 2007 *Análisis de estrategias eficientes en la logística de distribución de paquetería.*

Everitt, B. (1977), *Cluster analysis*. Heinemann Educational Books, Londres.

Ferrán, M. (1996), *SPSS para Windows. Programación y análisis estadístico*. Ed. McGraw Hill / Interamericana de España, S.A.

Fondo Europeo, (2005). *Informática aplicada al análisis económico*. Universidad Autónoma de Madrid.

Hall, R. 1998 Stochastic freight flow patterns: implications for fleet optimization. *Transportation Research Part A, 449-465*

Hall, R. ,Zhong, H.S. 2002 Decentralized inventory control policies for equipment management in a many-to-many network. *Transportation Research Part A 849-865*

Ortego, M. *Anàlisi de dades. Universitat Politècnica de Catalunya*

Palomera, R., Pawlowsky, V. (2001), *Tècniques estadístiques aplicades a la classificació de basalts*. Tesina d'especialitat ETSECCPB-UPC, 2001.

Ruszczynski, A., Shapiro, A., 2003.Stochastic Programming Models. Elsevier Science

Salvador, F. *Análisis de conglomerados o cluster.* Universidad de Zaragoza.

Vindel, D. 2004 *Anàlisi i classificació estacional de les carreteres catalanes.*

Visauta, B. *Análisis Estadístico*. Volumen I

Visauta, B., Sartori, J.C. *Análisis Estadístico*. Volumen II

- www.norusis.com/pdf/SPC_v13.pdf
- www.SPSStools.net/spss.htm *(programa SPSS)*
- Europa.eu/scadplus/legles/lvb/l24007.htm *(Fondo Europeo)*

ANEXO A1

DENDOGRAMAS SEGÚN LOS DISTINTOS MÉTODOS

DENDOGRAMA UTILIZANDO EL *SINGLE LINKAGE*

```
  C A S E      0         5        10        15        20        25
Label     Num  +---------+---------+---------+---------+---------+

Case 10    10   -+
Case 40    40   -+
Case 13    13   -+
Case 39    39   -+-+
Case 9      9   -+ |
Case 25    25   -+ |
Case 6      6   -+ |
Case 12    12   -+ |
Case 4      4   -+ |
Case 23    23   -+ |
Case 30    30   -+ |
Case 47    47   -+ |
Case 16    16   -+-+
Case 33    33   -+ |
Case 5      5   -+ |
Case 46    46   -+ |
Case 17    17   -+ |
Case 53    53   -+-+
Case 24    24   -+ |
Case 15    15   -+-+
Case 56    56   -+ |
Case 38    38   ---+
Case 2      2   -+-+
Case 27    27   -+ |
Case 37    37   ---+-+
Case 22    22   -+ | |
Case 36    36   -+-+ |
Case 32    32   -+ | |
Case 8      8   ---+ |
Case 11    11   ---+ +-+
Case 20    20   ---+ | |
Case 55    55   -----+ |
Case 42    42   -----+ |
Case 52    52   -----+ |
Case 54    54   -------+
Case 3      3   -------+
Case 43    43   ---+-+ +---+
Case 49    49   ---+ | |   |
Case 29    29   ---+-+ |   |
Case 51    51   ---+ | |   |
Case 44    44   -----+-+   +-+
Case 50    50   -----+     | |
Case 19    19   -----+     | |
Case 1      1   -----+-+   | |
Case 34    34   -----+ |   | |
Case 28    28   -------+---+ |
Case 21    21   -------+     +-+
Case 35    35   -------------+ |
Case 41    41   -----+---+   | +---------------+
Case 45    45   -----+   +---+ |               |
Case 14    14   ---------+     |               +-----------------+
Case 31    31   ---------------+               |                 |
Case 26    26   -------------------+-----------+                 |
Case 48    48   -------------------+                             |
Case 7      7   ----------------------------------------+-------+
Case 18    18   -----------------------------------------+
```

Fig. A1.1- *Dendograma según el método de single linkage.*

DENDOGRAMA UTILIZANDO EL *COMPLETE LINKAGE*

```
  C A S E      0         5        10        15        20        25
Label     Num  +---------+---------+---------+---------+---------+

Case 10    10   -+
Case 40    40   -+
Case 13    13   -+
Case 39    39   -+
Case 9      9   -+
Case 25    25   -+
Case 6      6   -+-+
Case 12    12   -+ |
Case 15    15   -+ +-+
Case 56    56   -+ | |
Case 38    38   ---+ |
Case 16    16   -+   +---+
Case 33    33   -+   |   |
Case 5      5   -+   |   |
Case 46    46   -+   |   |
Case 17    17   -+---+   |
Case 53    53   -+       |
Case 24    24   -+       |
Case 4      4   -+       +-----+
Case 23    23   -+       |     |
Case 30    30   -+       |     |
Case 47    47   -+       |     |
Case 3      3   -+-----+ |     |
Case 54    54   -+     | |     |
Case 43    43   -+---+ | |     |
Case 49    49   -+   | +-+     |
Case 2      2   -+   | |       |
Case 27    27   -+-+ | |       |
Case 37    37   -+ | +-+       |
Case 22    22   -+ | |         +-+
Case 36    36   -+ | |         | |
Case 32    32   -+-+ |         | |
Case 8      8   -+ +-+         | |
Case 11    11   -+ |           | |
Case 42    42   -+-+           | |
Case 52    52   -+ |           | |
Case 20    20   -+-+           | |
Case 55    55   -+             | |
Case 29    29   -+-+           | |
Case 51    51   -+ |           | +-------------------------------+
Case 44    44   -+-+-+         | |                               |
Case 50    50   -+ | +---------+ |                               |
Case 19    19   ---+ |           |                               |
Case 35    35   -----+           |                               |
Case 1      1   -+-+             |                               |
Case 34    34   -+ |             |                               |
Case 28    28   ---+---+         |                               |
Case 21    21   ---+   +-+       |                               |
Case 41    41   -+---+ | |       |                               |
Case 45    45   -+   +-+ +-------+                               |
Case 14    14   -----+   |                                       |
Case 31    31   ---------+                                       |
Case 26    26   -------+-----------------+                       |
Case 48    48   -------+                 +-----------------------+
Case 7      7   -------------+-----------+
Case 18    18   -------------+
```

Fig. A1.2 - *Dendograma según el método de complete linkage.*

DENDOGRAMA UTILIZANDO EL *CENTROID METHOD*

```
Dendrogram using Centroid Method

    C A S E      0         5        10        15        20        25
  Label     Num  +---------+---------+---------+---------+---------+

  Case 10    10   -+
  Case 40    40   -+
  Case 13    13   -+
  Case 39    39   -+
  Case 9      9   -+
  Case 25    25   -+
  Case 6      6   -+
  Case 12    12   -+
  Case 15    15   -+
  Case 56    56   -+
  Case 38    38   -+
  Case 16    16   -+
  Case 33    33   -+
  Case 5      5   -+
  Case 46    46   -+
  Case 17    17   -+
  Case 53    53   -+
  Case 24    24   -+
  Case 4      4   -+
  Case 23    23   -+
  Case 30    30   -+
  Case 47    47   -+
  Case 2      2   -+
  Case 27    27   -+
  Case 37    37   -+
  Case 8      8   -+
  Case 11    11   -+-+
  Case 22    22   -+ |
  Case 36    36   -+ |
  Case 32    32   -+ |
  Case 20    20   -+ |
  Case 42    42   -+ |
  Case 52    52   -+ |
  Case 55    55   -+ +-+
  Case 3      3   -+ | |
  Case 54    54   -+ | |
  Case 43    43   -+ | |
  Case 49    49   -+ | |
  Case 29    29   -+ | |
  Case 51    51   -+ | |
  Case 50    50   -+-+ |
  Case 44    44   -+   +-------------------------------------------+
  Case 19    19   -+   |                                           |
  Case 35    35   -+   |                                           |
  Case 1      1   -+   |                                           |
  Case 34    34   -+   |                                           |
  Case 28    28   -+   |                                           |
  Case 21    21   -+-+ |                                           |
  Case 41    41   -+ | |                                           |
  Case 45    45   -+ +-+                                           |
  Case 14    14   -+ |                                             |
  Case 31    31   ---+                                             |
  Case 26    26   ---+---------------+                             |
  Case 48    48   ---+               +-----------------------------+
  Case 7      7   ---------+---------+
  Case 18    18   ---------+
```

Fig. A1.3 - *Dendograma según el método del centroide.*

DENDOGRAMA UTILIZANDO EL *MEDIAN METHOD*

```
Dendrogram using Median Method

    C A S E      0         5        10        15        20        25
  Label     Num  +---------+---------+---------+---------+---------+

  Case 10    10   -+
  Case 40    40   -+
  Case 13    13   -+
  Case 39    39   -+
  Case 9      9   -+
  Case 25    25   -+
  Case 6      6   -+-+
  Case 12    12   -+ |
  Case 15    15   -+ +-----+
  Case 56    56   -+ |     |
  Case 38    38   ---+     |
  Case 17    17   -+       |
  Case 53    53   -+       |
  Case 24    24   -+       |
  Case 4      4   -+       |
  Case 23    23   -+-+     |
  Case 30    30   -+ |     |
  Case 47    47   -+ |     |
  Case 16    16   -+ |     |
  Case 33    33   -+ |     +-------+
  Case 5      5   -+ |     |       |
  Case 46    46   -+ |     |       |
  Case 2      2   -+ |     |       |
  Case 27    27   -+-+     |       |
  Case 37    37   -+ |     |       |
  Case 22    22   -+ |     |       |
  Case 36    36   -+-+---+ |       |
  Case 32    32   -+ |   | |       |
  Case 8      8   ---+   | |       |
  Case 11    11   ---+   | |       |
  Case 42    42   ---+   +-+       |
  Case 52    52   ---+   |         |
  Case 20    20   ---+   |         |
  Case 55    55   ---+   |         |
  Case 3      3   -----+-+         |
  Case 54    54   -----+           |
  Case 43    43   ---+-+           |
  Case 49    49   ---+ |           |
  Case 29    29   ---+-+-+         |
  Case 51    51   ---+ | |         +-----------+
  Case 50    50   -----+ +-+       |           |
  Case 44    44   -----+ | +-------+           |
  Case 19    19   -------+ |       |           |
  Case 35    35   ---------+       |           |
  Case 1      1   -----+           |           |
  Case 34    34   -----+           |           +-------------------+
  Case 28    28   -----+-----+     |           |                   |
  Case 21    21   -----+     +---+ |           |                   |
  Case 41    41   -----+---+ |   | |           |                   |
  Case 45    45   -----+   +-+   +-+           |                   |
  Case 14    14   ---------+     |             |                   |
  Case 31    31   ---------------+             |                   |
  Case 26    26   -----------------+-----------+                   |
  Case 48    48   -----------------+                               |
  Case 7      7   -----------------------------------+-------------+
  Case 18    18   -----------------------------------+
```

Fig. A1.4 - *Dendograma según el método de las medianas.*

DENDOGRAMA UTILIZANDO EL *BETWEEN GROUPS*

```
  C A S E      0         5        10        15        20        25
Label     Num  +---------+---------+---------+---------+---------+
Case 10    10   -+
Case 40    40   -+
Case 13    13   -+
Case 39    39   -+
Case 9      9   -+
Case 25    25   -+
Case 6      6   -+-+
Case 12    12   -+ |
Case 15    15   -+ |
Case 56    56   -+ |
Case 38    38   ---+
Case 16    16   -+ +---+
Case 33    33   -+ |   |
Case 5      5   -+ |   |
Case 46    46   -+ |   |
Case 17    17   -+-+   |
Case 53    53   -+     |
Case 24    24   -+     |
Case 4      4   -+     |
Case 23    23   -+     +-+
Case 30    30   -+     | |
Case 47    47   -+     | |
Case 2      2   -+     | |
Case 27    27   -+-+   | |
Case 37    37   -+ |   | |
Case 8      8   -+-+   | |
Case 11    11   -+ |   | +-+
Case 22    22   -+ +---+ | |
Case 36    36   -+-+     | |
Case 32    32   -+ |     | |
Case 20    20   ---+     | |
Case 42    42   ---+     | |
Case 52    52   ---+     | |
Case 55    55   ---+     | +-----+
Case 3      3   ---+-----+ |     |
Case 54    54   ---+       |     |
Case 43    43   -+---+     |     |
Case 49    49   -+   |     |     |
Case 29    29   -+-+ +-+   |     |
Case 51    51   -+ | | |   |     |
Case 50    50   ---+ | +---+     |
Case 44    44   ---+-+ |         +-------------------------------+
Case 19    19   ---+   |         |                               |
Case 35    35   -------+         |                               |
Case 1      1   ---+             |                               |
Case 34    34   ---+-+           |                               |
Case 28    28   ---+ +---+       |                               |
Case 21    21   -----+   +-+     |                               |
Case 41    41   ---+---+ | |     |                               |
Case 45    45   ---+   +-+ +-----+                               |
Case 14    14   -------+   |                                     |
Case 31    31   -----------+                                     |
Case 26    26   -----------+---------------------+               |
Case 48    48   -----------+                     +---------------+
Case 7      7   ---------------------+-----------+
Case 18    18   ---------------------+
```

Fig. A1.5 - *Dendograma según el método de las distancias entregrupos.*

DENDOGRAMA UTILIZANDO EL *WITHIN GROUP*

```
  C A S E      0         5        10        15        20        25
Label     Num  +---------+---------+---------+---------+---------+

Case 10    10   -+
Case 40    40   -+
Case 13    13   -+-+
Case 39    39   -+ |
Case 9      9   -+ |
Case 25    25   -+ |
Case 56    56   ---+
Case 6      6   -+-+---+
Case 12    12   -+ |   +-----+
Case 38    38   ---+   |     |
Case 55    55   -------+     |
Case 17    17   -+           |
Case 53    53   -+---+       |
Case 24    24   -+   |       |
Case 4      4   -+-+ |       |
Case 23    23   -+ | |       |
Case 30    30   ---+-+       +---+
Case 47    47   ---+ +---+   |   |
Case 16    16   -+-+ |   |   |   |
Case 33    33   -+ | |   |   |   |
Case 5      5   ---+-+   |   |   |
Case 46    46   ---+ |   +-+ |   |
Case 15    15   -----+   | | |   +-------+
Case 2      2   -+-+     | | |   |       |
Case 27    27   -+ +---+ | +-+   |       |
Case 37    37   ---+   +-+ |     |       +---+
Case 11    11   -------+   |     |       |   |
Case 3      3   -----------+     |       |   |
Case 31    31   -----------------+       |   |
Case 48    48   -------------------------+   |
Case 43    43   -----+---+                   |
Case 49    49   -----+   +---+               |
Case 44    44   ---------+   |               +---+
Case 22    22   -+-+         |               |   |
Case 36    36   -+ +-+       |               |   |
Case 32    32   ---+ |       +-------+       |   |
Case 8      8   -----+-+     |       |       |   |
Case 20    20   -----+ |     |       |       |   |
Case 52    52   -------+-+   |       |       |   |
Case 42    42   -------+ +---+       +-------+   +---+
Case 54    54   ---------+           |           |   |
Case 29    29   -----+---+           |           |   |
Case 51    51   -----+   |           |           |   |
Case 50    50   ---------+-----+     |           |   |
Case 19    19   ---------+     +-----+           |   +-----+
Case 35    35   ---------------+                 |   |     |
Case 26    26   ---------------------------------+   |     |
Case 1      1   ---------+-+                         |     |
Case 34    34   ---------+ +-+                       |     |
Case 28    28   -----------+ +---------+             |     +-----+
Case 21    21   -------------+         +-------------+     |     |
Case 41    41   -----------+-----+     |                   |     |
Case 45    45   -----------+     +-----+                   |     |
Case 14    14   -----------------+                         |     |
Case 7      7   -----------------------------------------+     |
Case 18    18   -------------------------------------------------+
```

Fig. A1.6 - *Dendograma según el método de las distancias intragrupos.*

DENDOGRAMA UTILIZANDO EL MÉTODO DE *WARD*

```
Dendrogram using Ward Method
   C A S E      0         5        10        15        20        25
 Label     Num  +---------+---------+---------+---------+---------+

 Case 10    10   -+
 Case 40    40   -+
 Case 13    13   -+
 Case 39    39   -+
 Case 9      9   -+
 Case 25    25   -+
 Case 38    38   -+
 Case 16    16   -+-+
 Case 33    33   -+ |
 Case 5      5   -+ |
 Case 46    46   -+ |
 Case 6      6   -+ |
 Case 12    12   -+ |
 Case 15    15   -+ |
 Case 56    56   -+ |
 Case 17    17   -+ |
 Case 53    53   -+ +-----+
 Case 24    24   -+ |     |
 Case 4      4   -+ |     |
 Case 23    23   -+ |     |
 Case 30    30   -+ |     |
 Case 47    47   -+ |     |
 Case 2      2   -+ |     |
 Case 27    27   -+ |     |
 Case 37    37   -+ |     |
 Case 8      8   -+-+     |
 Case 11    11   -+       |
 Case 3      3   -+       |
 Case 54    54   -+       |
 Case 43    43   -+       |
 Case 49    49   -+       +-------------------------------------+
 Case 42    42   -+       |                                     |
 Case 52    52   -+       |                                     |
 Case 22    22   -+       |                                     |
 Case 36    36   -+       |                                     |
 Case 32    32   -+       |                                     |
 Case 20    20   -+       |                                     |
 Case 55    55   -+       |                                     |
 Case 29    29   -+       |                                     |
 Case 51    51   -+       |                                     |
 Case 44    44   -+       |                                     |
 Case 50    50   -+-+     |                                     |
 Case 19    19   -+ |     |                                     |
 Case 35    35   -+ |     |                                     |
 Case 1      1   -+ +-----+                                     |
 Case 34    34   -+ |                                           |
 Case 28    28   -+ |                                           |
 Case 21    21   -+ |                                           |
 Case 41    41   -+-+                                           |
 Case 45    45   -+                                             |
 Case 14    14   -+                                             |
 Case 31    31   -+                                             |
 Case 26    26   -+---+                                         |
 Case 48    48   -+   +-----------------------------------------+
 Case 7      7   -+---+
 Case 18    18   -+
```

Fig. A1.7 - *Dendograma según el métodode Ward.*

ANEJO A2

TABLAS Y ESTADÍSTICOS

RELACIÓN EXPONENCIAL ENTRE EL CV Y LA MEDIA.

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,543944818
Coeficiente de determinación R^2	0,295875965
R^2 ajustado	0,281206714
Error típico	0,272469837
Observaciones	50

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*
Regresión	1	1,49740272	1,49740272	20,1698076
Residuos	48	3,56351097	0,074239812	
Total	49	5,06091369		

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*
Intercepción	-0,480242177	0,265214686	-1,810767661	0,07643792
Variable X 1	-0,115476544	0,025712417	-4,491080892	4,45E-05

Cálculos del comportamiento de las variables para el grupo 1

SALIDAS	MEDIA	DES.TIP	CV	ln CV	ln MEDIA	xy	xx	recta regresión	error^2	y-ymedio^2
10 - STA. PERPETUA	315482	42525	0,135	-2,004	12,662	-25,4	160,3	-1,942	0,004	0,119
29 - COSLADA	323671	39799	0,123	-2,096	12,687	-26,6	161	-1,945	0,023	0,191
52 - RIBAROJA (VALENCIA)	205240	34395	0,168	-1,786	12,232	-21,9	149,6	-1,893	0,011	0,016
87 - TORRELLANO	201699	36208	0,18	-1,718	12,215	-21	149,2	-1,891	0,03	0,003

RESULTADOS POR AMBOS MÉTODOS				
CV-MEDIA		**COMPARACION**	***LN CV-LN MEDIA***	
a	0,1260484		**a**	-3,789055335
b	3,14E-07		**b**	0,169546532
R^2	0,377261163		**R^2**	0,364067457

Cálculos del comportamiento de las variables para el grupo 2

SALIDAS	MEDIA	DES.TIP	CV	ln CV	ln MEDIA	xy	xx	recta regresion	error^2	y-ymedio^2
01 - LAZKAO (GUIPÚZCOA)	138338	24464	0,177	-1,733	11,837	-20,5	140,1	-1,847	0,013	0,005
20 - DERIO (BILBAO)	140955	26733	0,19	-1,663	11,856	-19,7	140,6	-1,849	0,035	0
40 - HERNANI	106546	17433	0,164	-1,81	11,576	-21	134	-1,817	0	0,023
60 - PAMPLONA	142331	25204	0,177	-1,731	11,866	-20,5	140,8	-1,85	0,014	0,005
61 - COBATILLAS	38538	6865	0,178	-1,725	10,559	-18,2	111,5	-1,7	0,001	0,004
63 - IGUALADA (BCN)	168518	27320	0,162	-1,819	12,035	-21,9	144,8	-1,87	0,003	0,026
70 - VITORIA	122061	19750	0,162	-1,821	11,712	-21,3	137,2	-1,833	0	0,026
73 - SEVILLA	53321	6612	0,124	-2,087	10,884	-22,7	118,5	-1,737	0,123	0,184
80 - EL PRAT (BCN)	119210	17457	0,146	-1,921	11,689	-22,5	136,6	-1,83	0,008	0,069
83 - VIGO	44833	5874	0,131	-2,032	10,711	-21,8	114,7	-1,717	0,099	0,14
84 - ZARAGOZA	120298	20553	0,171	-1,767	11,698	-20,7	136,8	-1,831	0,004	0,012
89 - ALVERCA DO RIBATEJO	47624	5986	0,126	-2,074	10,771	-22,3	116	-1,724	0,122	0,172
90 - SILVOTA (ASTURIAS)	41616	5580	0,134	-2,009	10,636	-21,4	113,1	-1,708	0,09	0,123

RESULTADOS POR AMBOS MÉTODOS				
CV-MEDIA		**COMPARACION**	***LN CV-LN MEDIA***	
a	0,1260484		**a**	-3,78905533
b	3,14E-07		**b**	0,16954653
R^2	0,37726116		**R^2**	0,36406746

Cálculos del comportamiento de las variables para el grupo 3

SALIDAS	MEDIAs	DES.TIP	CV	ln CV	ln MEDIA	xy	xx	recta regresion	error^2	y-ymedio^2
03 - CORDOBA	8735	1642	0,188	-1,672	9,075	-15,2	82,36	-1,528	0,021	0,0149
04 - ALBAIDA (VALENCIA)	67045	14205	0,212	-1,552	11,113	-17,2	123,5	-1,764	0,045	2,4081
05 - MANZANARES	15685	3232	0,206	-1,58	9,66	-15,3	93,32	-1,596	0	2,4951
08 - ALBACETE	8340	2562	0,307	-1,18	9,029	-10,7	81,52	-1,523	0,117	1,3926
09 - ALGECIRAS (CÁDIZ)	3952	846	0,214	-1,542	8,282	-12,8	68,59	-1,437	0,011	2,3772
12 -PELIGROS (GRANADA)	24629	4469	0,181	-1,707	10,112	-17,3	102,2	-1,648	0,003	2,9127
14 - FARO (PORTUGAL)	1018	247	0,243	-1,414	6,925	-9,8	47,96	-1,28	0,018	2,0006
23 - FERROL (LA CORUÑA)	16242	7536	0,464	-0,768	9,695	-7,45	94	-1,6	0,692	0,5897
26 - SAN JUAN DEL PUERTO	2952	435	0,147	-1,915	7,99	-15,3	63,85	-1,403	0,262	3,6669
27 - LEÓN	7795	1463	0,188	-1,673	8,961	-15	80,3	-1,515	0,025	2,7988
31 - MÁLAGA	20788	3678	0,177	-1,732	9,942	-17,2	98,85	-1,628	0,011	3
35 - HUECAS (TOLEDO)	42942	9169	0,214	-1,544	10,668	-16,5	113,8	-1,712	0,028	2,384
41 - CASTELLÓN	35103	10336	0,294	-1,223	10,466	-12,8	109,5	-1,689	0,217	1,4949
42 - SANTIAGO	16689	2413	0,145	-1,934	9,722	-18,8	94,53	-1,603	0,11	3,7406
49 - JAEN	3952	1319	0,334	-1,097	8,282	-9,09	68,59	-1,437	0,115	1,2036
51 - ZAMORA	483	130	0,27	-1,31	6,18	-8,1	38,2	-1,194	0,014	1,7164
55 - PUERTO DE STA. MARÍA	8825	1383	0,157	-1,853	9,085	-16,8	82,54	-1,529	0,105	3,435
62 - LÉRIDA	17936	3403	0,19	-1,662	9,795	-16,3	95,93	-1,611	0,003	2,7623
65 - SANTANDER	28224	4077	0,144	-1,935	10,248	-19,8	105	-1,664	0,074	3,7436
67 - SALAMANCA	4732	1702	0,36	-1,023	8,462	-8,65	71,61	-1,457	0,189	1,0459
74 - TARRAGONA	33505	7226	0,216	-1,534	10,419	-16	108,6	-1,683	0,022	2,3533

75 - MÉRIDA (BADAJOZ)	17505	3600	0,206	-1,582	9,77	-15,5	95,46	-1,608	0,001	2,5013
79 - CUENCA	743	221	0,298	-1,211	6,611	-8	43,7	-1,244	0,001	1,4656
81 - BESCANO (GERONA)	50766	13316	0,262	-1,338	10,835	-14,5	117,4	-1,731	0,155	1,7909
82 - LA CORUÑA	53849	6832	0,127	-2,065	10,894	-22,5	118,7	-1,738	0,106	4,2623
85 - LUGO	8867	1904	0,215	-1,538	9,09	-14	82,63	-1,53	0	2,3666
86 - (ORENSE)	18809	2793	0,149	-1,907	9,842	-18,8	96,87	-1,617	0,084	3,637
88 - (OPORTO)	58916	7369	0,125	-2,079	10,984	-22,8	120,6	-1,749	0,109	4,3213
93 - VALLADOLID	52702	7081	0,134	-2,007	10,872	-21,8	118,2	-1,736	0,074	4,0289
95 - COIMBRA - (PORTUGAL)	5238	1893	0,361	-1,018	8,564	-8,72	73,34	-1,469	0,204	1,0358
96 - LOGROÑO	61595	11554	0,188	-1,674	11,028	-18,5	121,6	-1,754	0,006	2,8008
97 - VILLAFRIA (BURGOS)	40819	7420	0,182	-1,705	10,617	-18,1	112,7	-1,706	0	2,907
98 - ALMERÍA	12248	3814	0,311	-1,167	9,413	-11	88,61	-1,567	0,16	1,361
			sumatorio	-51,14	312,634	-490	3015		2,983	80,015
			promedio	-1,55	9,474					

RESULTADOS POR AMBOS MÉTODOS				
CV-MEDIA		COMPARACION	*LN CV-LN MEDIA*	
a	0,258630921		**a**	**-0,548066728**
b	-1,50E-06		**b**	**-0,105720504**
R^2	**0,11536757**	**<==========>**	**R^2**	**0,139647022**

Tabla de relaciones entre CV y su media

MEDIA s	CV	COSH M	SENH CV	TANH CV	LN CV	LN M	EXP CV	100000^CV	gamma ln m	gamma ln cv
138338	0,177	12,531	0,178	0,175	-1,733	11,8	1,193	7,659	1E+06	1,65
8735	0,188	9,768	0,189	0,186	-1,672	9,08	1,207	8,706	70533	1,59
67045	0,212	11,806	0,213	0,209	-1,552	11,1	1,236	11,46	7E+05	1,46
15685	0,206	10,354	0,208	0,203	-1,58	9,66	1,229	10,72	1E+05	1,49
8340	0,307	9,722	0,312	0,298	-1,18	9,03	1,36	34,38	66955	1,07
3952	0,214	8,975	0,216	0,211	-1,542	8,28	1,239	11,75	28777	1,45
315482	0,135	13,355	0,135	0,134	-2,004	12,7	1,144	4,72	4E+06	1,94
24629	0,181	10,805	0,182	0,18	-1,707	10,1	1,199	8,079	2E+05	1,63
1018	0,243	7,618	0,245	0,238	-1,414	6,93	1,275	16,42	6026	1,32
140955	0,19	12,549	0,191	0,187	-1,663	11,9	1,209	8,877	2E+06	1,58
16242	0,464	10,388	0,481	0,433	-0,768	9,7	1,59	208,9	1E+05	0,65
2952	0,147	8,683	0,148	0,146	-1,915	7,99	1,159	5,455	20634	1,85
7795	0,188	9,654	0,189	0,186	-1,673	8,96	1,206	8,679	62057	1,59
323671	0,123	13,381	0,123	0,122	-2,096	12,7	1,131	4,119	4E+06	2,04
20788	0,177	10,635	0,178	0,175	-1,732	9,94	1,194	7,667	2E+05	1,65
42942	0,214	11,361	0,215	0,21	-1,544	10,7	1,238	11,68	4E+05	1,45
106546	0,164	12,269	0,164	0,162	-1,81	11,6	1,178	6,578	1E+06	1,74
35103	0,294	11,159	0,299	0,286	-1,223	10,5	1,342	29,66	3E+05	1,12
16689	0,145	10,416	0,145	0,144	-1,934	9,72	1,156	5,282	1E+05	1,87
3952	0,334	8,975	0,34	0,322	-1,097	8,28	1,396	46,69	28774	0,98
483	0,27	6,874	0,273	0,263	-1,31	6,18	1,31	22,33	2501	1,21
205240	0,168	12,925	0,168	0,166	-1,786	12,2	1,182	6,885	2E+06	1,71
8825	0,157	9,778	0,157	0,155	-1,853	9,09	1,17	6,075	71349	1,78
142331	0,177	12,559	0,178	0,175	-1,731	11,9	1,194	7,681	2E+06	1,65
38538	0,178	11,253	0,179	0,176	-1,725	10,6	1,195	7,776	4E+05	1,65
17936	0,19	10,488	0,191	0,188	-1,662	9,8	1,209	8,887	2E+05	1,58
168518	0,162	12,728	0,163	0,161	-1,819	12	1,176	6,465	2E+06	1,75
28224	0,144	10,941	0,145	0,143	-1,935	10,2	1,155	5,275	3E+05	1,87
4732	0,36	9,155	0,367	0,345	-1,023	8,46	1,433	62,82	35305	0,91
122061	0,162	12,405	0,163	0,16	-1,821	11,7	1,176	6,442	1E+06	1,75
53321	0,124	11,577	0,124	0,123	-2,087	10,9	1,132	4,169	5E+05	2,03
33505	0,216	11,113	0,217	0,212	-1,534	10,4	1,241	11,98	3E+05	1,44
17505	0,206	10,463	0,207	0,203	-1,582	9,77	1,228	10,67	2E+05	1,49
743	0,298	7,304	0,302	0,289	-1,211	6,61	1,347	30,91	4167	1,1
119210	0,146	12,382	0,147	0,145	-1,921	11,7	1,158	5,398	1E+06	1,85
50766	0,262	11,528	0,265	0,256	-1,338	10,8	1,3	20,49	5E+05	1,24
53849	0,127	11,587	0,127	0,126	-2,065	10,9	1,135	4,309	5E+05	2
44833	0,131	11,404	0,131	0,13	-2,032	10,7	1,14	4,52	4E+05	1,97
120298	0,171	12,391	0,172	0,169	-1,767	11,7	1,186	7,149	1E+06	1,69
8867	0,215	9,783	0,216	0,211	-1,538	9,09	1,24	11,85	71733	1,45
18809	0,149	10,535	0,149	0,147	-1,907	9,84	1,16	5,528	2E+05	1,84
201699	0,18	12,908	0,18	0,178	-1,718	12,2	1,197	7,899	2E+06	1,64
58916	0,125	11,677	0,125	0,124	-2,079	11	1,133	4,221	6E+05	2,02
47624	0,126	11,464	0,126	0,125	-2,074	10,8	1,134	4,251	5E+05	2,01
41616	0,134	11,329	0,134	0,133	-2,009	10,6	1,143	4,682	4E+05	1,95
52702	0,134	11,566	0,135	0,134	-2,007	10,9	1,144	4,697	5E+05	1,94
5238	0,361	9,257	0,369	0,346	-1,018	8,56	1,435	64,12	39618	0,9
61595	0,188	11,721	0,189	0,185	-1,674	11	1,206	8,667	6E+05	1,59
40819	0,182	11,31	0,183	0,18	-1,705	10,6	1,199	8,107	4E+05	1,63
12248	0,311	10,106	0,316	0,302	-1,167	9,41	1,365	36,06	1E+05	1,06

ANEXO A3

BONDAD DE AJUSTE DE LOS DISTINTOS MODELOS DE PREDICCIÓN

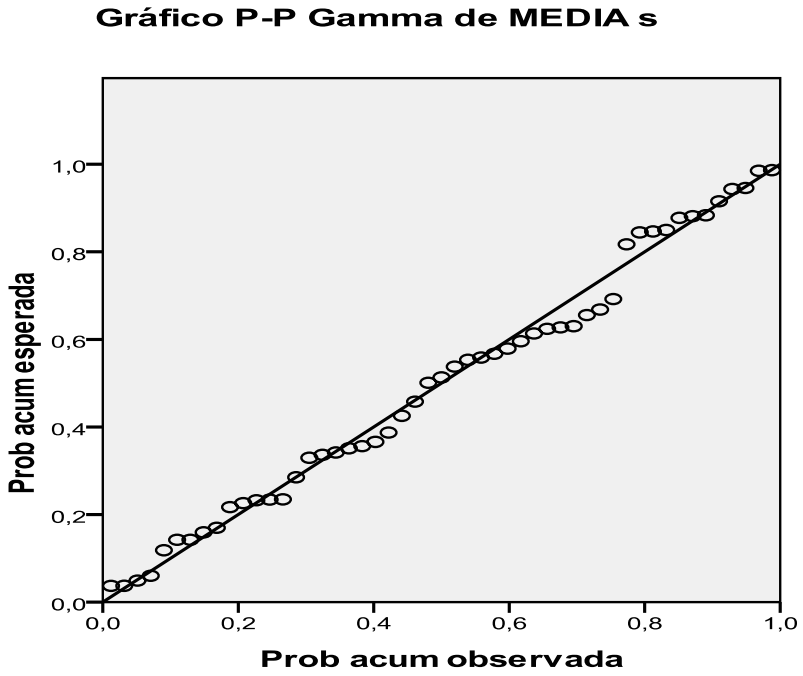


Fig. A3.1.- *Grafico P-P del ajuste a una distribución Gamma.*

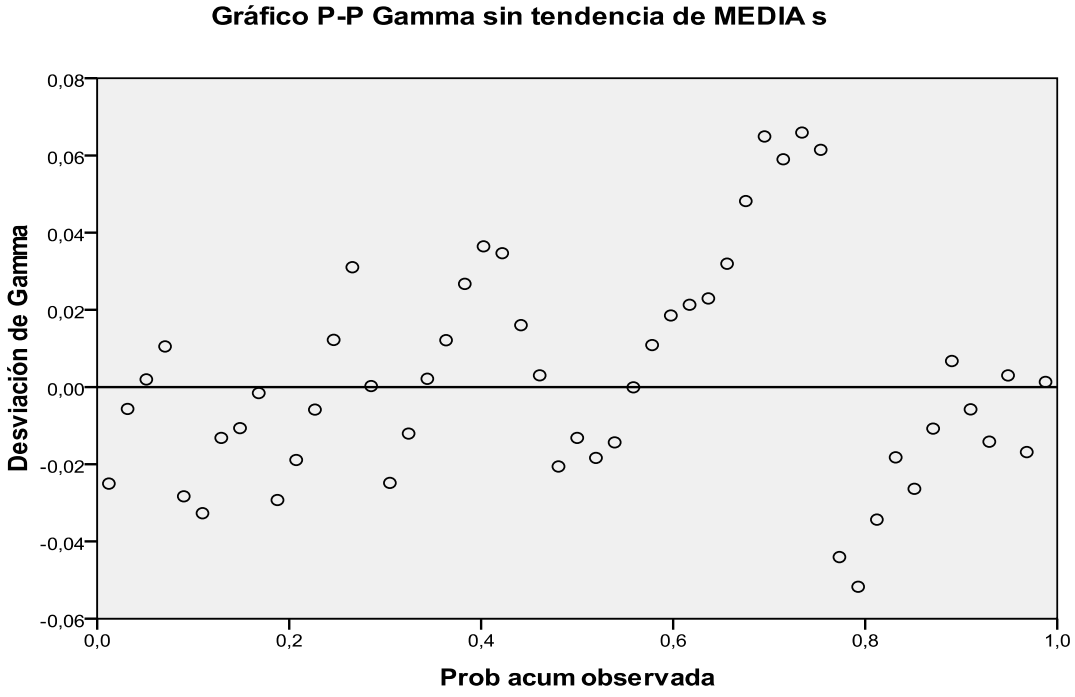


Fig. A3.2.- *Grafico P-P del ajuste a una distribución Gamma.*

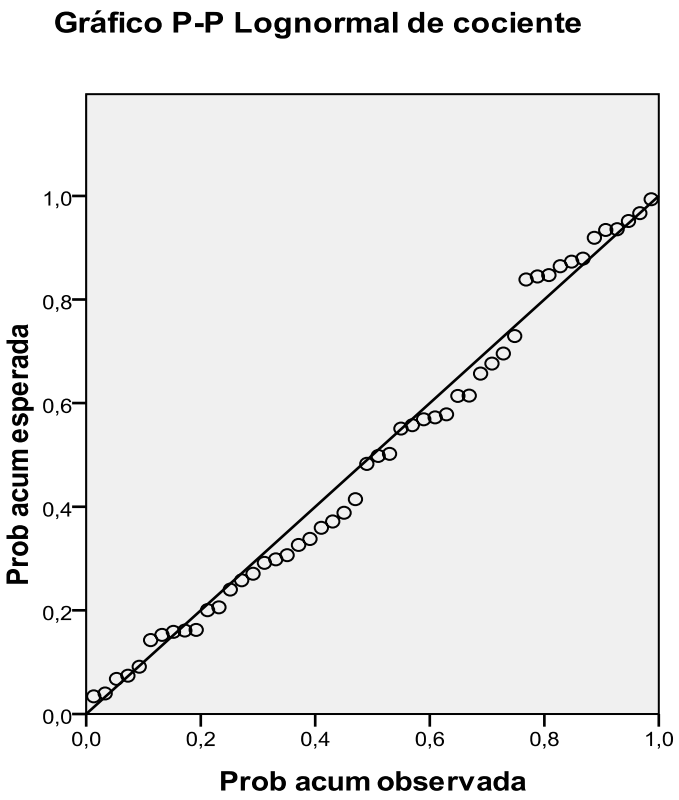


Fig. A3.3.- *Diagrama P-P para el ajuste a una distribución Lognormal.*

ANEXO A4

DIAGRAMAS BOX-PLOT

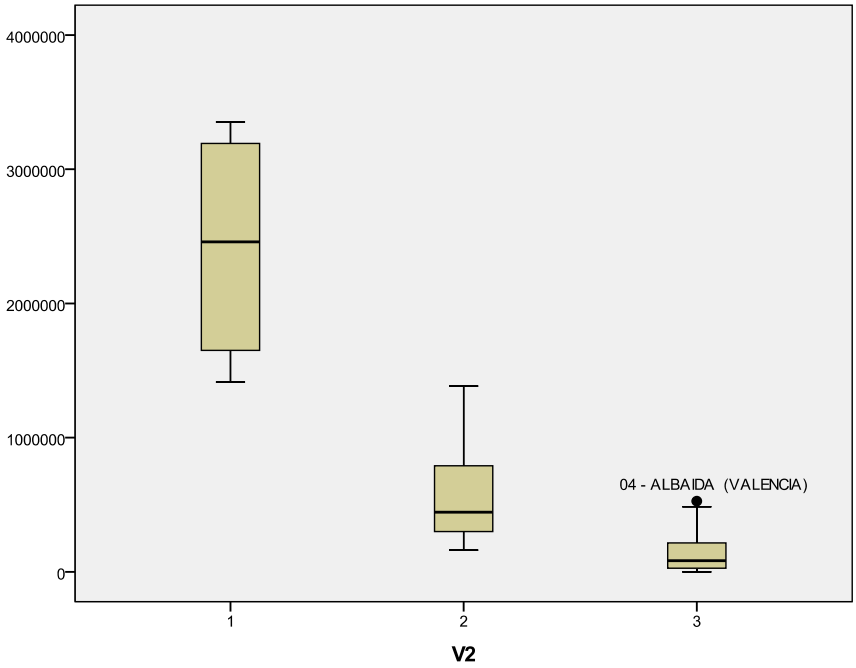


Fig. A4.1. *Gráficos Box Plot para la variable numero paquetes salientes*

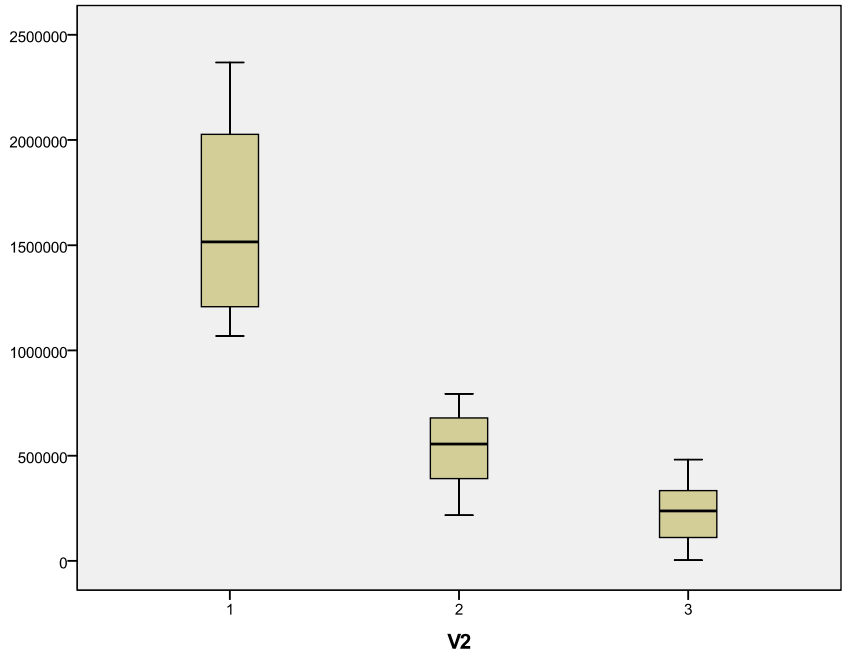


Fig. A4.2. *Gráficos Box Plot para la variable numero llegadas paquetes*

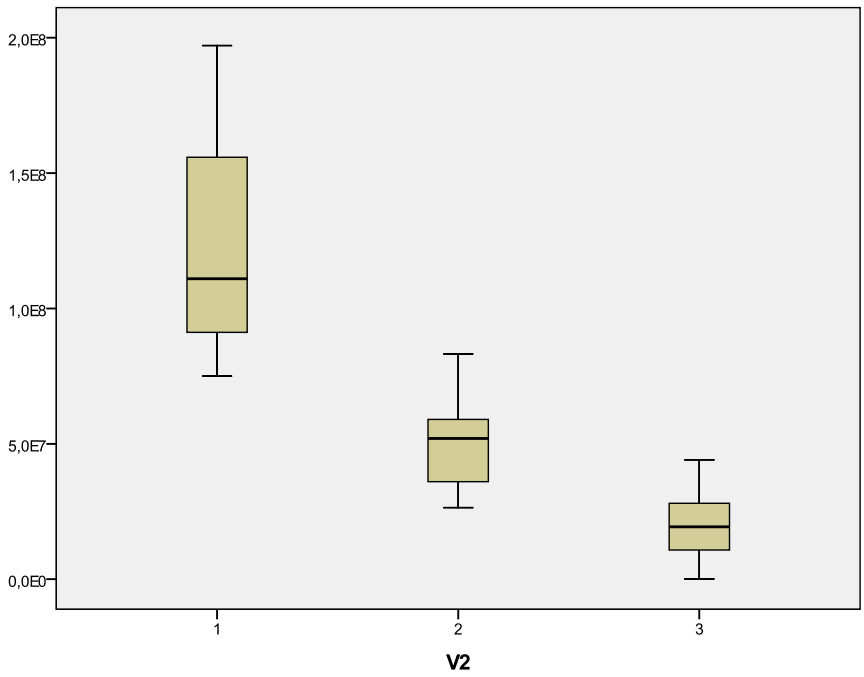


Fig. A4.3. Gráficos Box Plot para la variable kilos llegadas