

R E V I S T A

INGENIERIA DE SISTEMAS

Volumen XXXI

Septiembre 2017

- Diseño Optimizado de la Red Eléctrica de un Parque Eólico. 5
Claudio Riso, Eduardo Canale.
- Un enfoque analítico para la confección del ranking FIFA y la realización del sorteo del Mundial de fútbol. 21
Sebastián Cea, Guillermo Durán, Mario Guajardo, Denis Sauré, Gonzalo Zamorano.
- Políticas de desincentivo en el consumo de alcohol y efecto en los accidentes de tránsito 47
Valeria Scapini, Cinthya Vergara, Diego Álvarez.
- Programación entera para el diseño de jornadas laborales de reponedores en la industria del retail. 65
Jaime Miranda, Fernanda Palacios.
- Algoritmos de zonificación para el problema de la recolección de residuos urbanos: El caso de estudio de una ciudad argentina. 81
Marcelo Bianchetti, Guillermo Durán, Ivo Koch, Javier Marengo.

Publicada por el
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

R E V I S T A
INGENIERIA DE SISTEMAS

ISSN 0716 - 1174

EDITOR

Guillermo Durán

*Departamento de Ingeniería Industrial
Universidad de Chile*

EDITOR ASOCIADO

Richard Weber

*Departamento de Ingeniería Industrial
Universidad de Chile*

AYUDANTE DE EDICIÓN

Cinthya Vergara

*Departamento de Ingeniería Industrial
Universidad de Chile*

COMITÉ EDITORIAL

René Caldentey

University of Chicago, Estados Unidos

Héctor Cancela

Universidad de la República, Uruguay

Rafael Epstein

Universidad de Chile, Chile

Luis Llanos

CMPC Celulosa, Chile

Javier Marengo

*Universidad Nacional de
General Sarmiento, Argentina*

Juan de Dios Ortúzar

P. Universidad Católica, Chile

Víctor Parada

Universidad de Santiago, Chile

Oscar Porto

GAPSO, Brasil

Lorena Pradenas

Universidad de Concepción, Chile

Nicolás Stier

Facebook Core Data Science, Estados Unidos

Financiado parcialmente por el Instituto Sistemas Complejos de Ingeniería.

Las opiniones y afirmaciones expuestas representan los puntos de vista de sus autores y no necesariamente coinciden con las del Departamento de Ingeniería Industrial de la Universidad de Chile.

Los artículos sólo pueden ser reproducidos previa autorización del Editor y de los autores.

Representante legal: Fernando Ordóñez
Correo electrónico: ris@dii.uchile.cl
Diagramación: Cinthya Vergara/Macarena Matta

Dirección: Av. Beauchef 851, Santiago, Chile.
Web URL: www.dii.uchile.cl/ris
Diseño Portada: Gabriella Fabbri

Carta Editorial Volumen XXXI

Nos es muy grato presentar este nuevo número de la Revista de Ingeniería de Sistemas (RIS) dedicado a temas de frontera en Investigación de Operaciones, Gestión y Tecnología. Queremos agradecer al Instituto Sistemas Complejos de Ingeniería (ISCI) por su colaboración para hacer posible esta publicación.

Este número contiene artículos de académicos y estudiantes de nuestro Departamento de Ingeniería Industrial, de investigadores del ISCI, y de académicos de universidades de Argentina y Uruguay.

Nuestro objetivo a través de esta publicación es contribuir a la generación y difusión de las tecnologías modernas de gestión y administración. La revista pretende destacar la importancia de generar conocimiento en estas áreas, orientado tanto a problemáticas nacionales como a la realidad de países de características similares de la región.

Estamos seguros de que los artículos publicados en esta oportunidad muestran formas de trabajo innovadoras que serán de gran utilidad e inspiración para todos los lectores, ya sean académicos o profesionales, por lo que esperamos que esta iniciativa tenga la recepción que creemos se merece.

Guillermo Durán
Editor

Richard Weber
Editor Asociado

Llamado a Presentar Trabajos

La Revista Ingeniería de Sistemas (RIS) busca constituir un canal de divulgación de los avances en las áreas de Gestión de Operaciones, Tecnologías de Información e Investigación Operativa, que incluya los mundos académico y empresarial. Son particularmente apropiados artículos orientados a la práctica de estas disciplinas, que estimulen su uso o den cuenta de aplicaciones innovadoras de ellas, especialmente en América Latina.

También son bienvenidos artículos con análisis del estado del arte en un campo particular y de la forma en que los avances en dicho campo se han utilizado en la práctica.

Se espera que los artículos estén escritos de manera que puedan ser leídos por personas no especialistas en el tema tratado. Se recomienda incluir una lista de lecturas sugeridas para que los lectores no especialistas puedan profundizar en el tema.

Formato del Manuscrito

Los autores deben enviar un archivo en formato PDF del manuscrito que desean someter a referato a:

*Comité Editorial Revista Ingeniería de Sistemas,
Departamento de Ingeniería Industrial,
Universidad de Chile.
Santiago, Chile.
Email: ris@dii.uchile.cl*

Los manuscritos deben estar formateados para hojas tamaño carta, a doble espacio, márgenes de 2,5 centímetros en todos los lados, deben incluir un resumen de no más de 150 palabras y su extensión no debe exceder las 20 hojas.

La primera hoja debe contener el título del trabajo, nombre y dirección de los autores (teléfono y correo electrónico del autor de contacto), y un resumen de no más de 150 palabras.

Referencias

Las referencias se deben citar en el cuerpo del texto usando el nombre del autor y el año de publicación, e.g., Morton (1998). Al final del artículo se debe incluir la lista en orden alfabético de las referencias citadas en el texto. Para referencias de revistas científicas el formato es el siguiente: Autor(es), Año de publicación. Título. Nombre completo de la revista , Volumen e.g.:

Kodialam, M. y H. Luss, 1998. Algorithms for Separable Nonlinear Resource Allocation Problems. *Operations Research* , 44(2), 272-284.

Para referencias de libros el formato es el siguiente: autor(es), año de publicación. Título. Editorial, Ciudad; e.g.:

Kleinrock, L., 1975. *Queueing Systems* . John Wiley, New York.

En caso de haber más de una referencia con el mismo autor y año de publicación, se debe usar "a", "b", etc. como sufijo del año de publicación para diferenciarlas.

Detalles en www.dii.uchile.cl/ris

DISEÑO OPTIMIZADO DE LA RED ELÉCTRICA DE UN PARQUE EÓLICO

CLAUDIO RISSO *
EDUARDO CANALE *

Resumen

En este artículo se analiza el problema de diseño a costo mínimo de la red eléctrica de un parque eólico, a saber, la responsable de agregar la potencia de los aerogeneradores para luego entregarla a la red de alta tensión. El artículo presenta el problema de diseño, un modelo exacto simplificado, y explica cómo se usaron algoritmos heurísticos para resolverlo en instancias reales. Los resultados de este trabajo ayudaron a reducir los costos en un proyecto concreto en Uruguay, logrando ahorros en las inversiones de la red en cifras superiores al 30 %.

Palabras Clave: Optimización combinatoria, diseño de redes overlay, redes de distribución, energía eólica

1. Introducción

Este documento presenta algunos desafíos inherentes al despliegue de la red eléctrica de un parque eólico, en particular los que tienen que ver con el *layout de la red*, esto es: con el mapa de conexiones entre puntos y sus recorridos físicos. El trabajo está inspirado en una aplicación real, el *Parque Eólico Las Palomas*, construido por la empresa Teyma-Abengoa en la localidad Palomas, departamento de Salto, Uruguay. El documento describe el modelo de referencia para el despliegue de la red eléctrica de un parque eólico onshore, las instancias sobre las que se realizaron las optimizaciones, así como los métodos y algoritmos usados para resolverlas.

* Facultad de Ingeniería, Universidad de la República, Uruguay

Siguiendo los lineamientos del reporte [3] para el esquema de componentes del layout de un parque eólico, el diagrama en bloques es el de Figura 1. Los *aerogeneradores* (*wind turbines*, WT), entregan su potencia a través de una *red local* (*local wind turbine grid*, LWTG) hacia un *punto de agregación* (*Collecting Point*, CP o PT), que conecta a su vez con la red de alta tensión en un PCC (*point of common connection*) mediante un sistema de transmisión. La interface del parque eólico con la red (*wind farm grid interface*), es la responsable de adaptar: voltage, frecuencia y potencia reactiva para adecuarlas a las del PCC. Como veremos, este trabajo se centró en optimizar la LWTG.

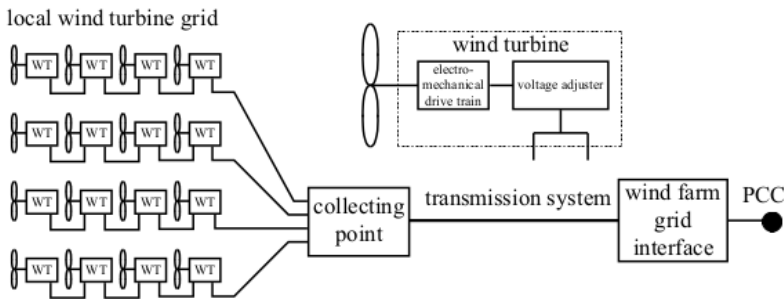


Figura 1: Layout general de un parque eólico

Como referencia (ver [3]), en un parque de una potencia instalada de unos 60MW, los costos de las WT constituyen el 59.9% del total, el sistema de transmisión un 21.6%, mientras que 11.1% corresponden a pérdidas eléctricas de distinto tipo, y 7.4% al despliegue del LWTG. El caso de estudio sobre el que este trabajo se basa (el Parque Eólico Las Palomas) tiene una potencia instalada es de 70MW (35 WT de 2MW). Se implementó mediante un “leasing operativo”, donde se licitó el desarrollo, la ingeniería y la construcción, así como su posterior operación y mantenimiento por un periodo de veinte años. Los datos con los que los oferentes contaban eran: las posiciones y potencia de las WT a instalar, la caminería interna a construir para conectar esos puntos, más un conjunto de restricciones técnico ambientales a cumplir.

A efectos de este artículo se destaca: se busca mantener bajo control la caída de tensión (no las pérdidas eléctricas en la red), los circuitos eléctricos deben ser desplegados subterráneamente, en zanjas que acompañen el trazado de la red caminería. Los oferentes debían proponer un diseño para la red eléctrica a construir, compatible con las restricciones. El adjudicatario resultó la empresa Teyma-Abengoa. Una vez adjudicado, el pliego preveía unos meses para que el contratista sugiriera mejoras al proyecto. La participación académica de los autores de este artículo se dio en ese período.

La restricción de acompañar la caminería, que tiene su raíz en las nor-

mas de cuidado ambiental nacionales, cambia sustancialmente la naturaleza del problema. Cuando se permite cualquier tendido para los circuitos que conectan los puntos, la red se diseña en una única capa. Se recomienda ver [2], donde se trata un problema así para un parque offshore. Acompañar la caminería implica coordinar el diseño entre dos redes, al transformar a la red eléctrica (LWTG, el objeto de estudio este trabajo) en una red sobre otra red (*overlay*) de caminos existente. Como resultado de estos cambios, los costos típicos para un proyecto de este tipo, sin considerar las pérdidas eléctricas ni los componentes de transmisión son: 65 % para las WTs, 20 % para la caminería y 15 % para la LWTG. No obstante, la red local era una de las pocas variables de decisión donde el contratista podía optimizar, y aunque bajo en términos relativos, el despliegue del LWTG cuesta varios millones de dólares, y su optimización presenta interés tanto académico como económico.

La Sección 2 presenta el detalle general del problema, y las características de una instancia basada en un problema real. En Sección 3 se presenta un modelo combinatorio del problema que, bajo ciertas premisas de las instancias, permite encontrar soluciones. La Sección 4 explica las características de las heurísticas usadas para encontrar soluciones en otras instancias de valor práctico, no ajustadas a las premisas anteriores. Finalmente, Sección 5 resume los principales resultados y conclusiones, así como líneas de trabajo futuro.

2. Modelo para el LWTG

El problema de referencia es el siguiente. Se busca un esquema para interconectar las 35 WTs (wind-turbines) del parque eólico en cuestión con el CP (collecting-point). Las posiciones de esos 36 puntos son conocidas y la topología de interconexión debe ser un árbol, donde CP corresponde a la raíz. Cada WT puede estar conectado hacia el CP, o hacia otro WT en su camino al mismo. Llamaremos “nivel lógico” al esquema de conexiones eléctricas (a.k.a. circuitos) entre WTs o hacia el CP. En todo árbol el camino de la potencia generada por los WTs hacia el CP es único. El tipo de cable a usar en cada arista del árbol está determinado por el número de WTs que entregan su potencia a través de la misma. Como característica, cada tipo de cable tiene un costo y caída de tensión por metro conocida para toda cantidad de wind-turbines.

Todos los cables van dentro de zanjas, que solamente pueden ser desplegadas sobre trazados existentes en un mapa conocido. Llamaremos “nivel físico” al esquema de caminos seguido por los circuitos eléctricos. A las posiciones físicas correspondientes al CP y a los WTs, se suman otros 20 “puntos de bifurcación” (cruces entre caminos de la red vial). Del despliegue físico elegido

para cada circuito surge un metraje a usar de cada tipo de cable. Los tramos físicos son de largo conocido y el tipo de zanja a usar en cada tramo surge del número de circuitos a desplegar en ella.

El costo total del LWTG es la suma de los costos de: los cables de los circuitos, las zanjas y de la cantidad de circuitos de entrada al CP. El costo de los cables es la suma de los productos entre los metros a usar de cada tipo de cable y el costo por metro de cada uno. El costo de las zanjas es proporcional al largo de cada tramo (conocido), y a un costo por metro que depende del número de circuitos a enterrar en cada tramo. El costo de la cantidad de circuitos de entrada al CP es proporcional a un valor conocido.

Como elemento adicional de diseño, existen umbrales a respetar para la caída de tensión asociada al layout elegido para el LWTG. La caída total desde cada WT en su camino al collecting-point no puede superar el 3%. Además, un WT puede ser extremo de a lo sumo tres circuitos. El objetivo de este trabajo es encontrar una configuración de conexiones para el parque, que minimice las inversiones en: cableados, zanjas y circuitos de entrada al collecting-point, necesarias para un layout factible, esto es, cumpliendo con todas las restricciones de diseño antes mencionadas.

2.1. Instancia Base

Elaboraremos a continuación en los detalles de la instancia *baseline* sobre la que se realizaron las primeras optimizaciones del proyecto.

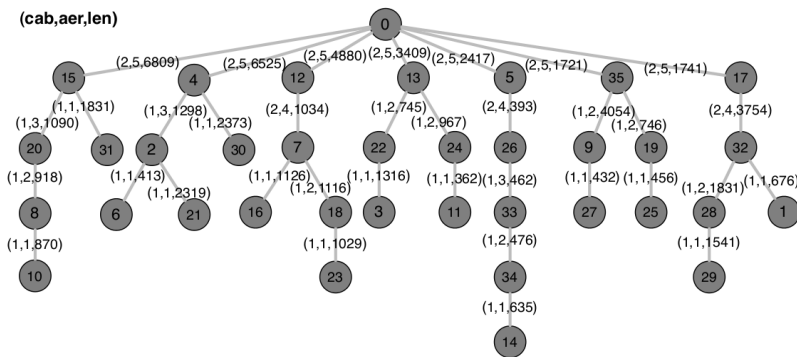


Figura 2: Solución *baseline* para la interconexión eléctrica. Las etiquetas en los circuitos detallan: tipo de cable, número de generadores y distancia física

La empresa (Teyma-Abengoa) había sido adjudicada con el diseño para el LWTG esquematizado en Figura 2. La empresa llegó a esta configuración manualmente, explorando una decena de diseños elegidos cualitativamente, apoyados en la intuición de los técnicos. En el esquema anterior, el nodo 0

representa al CP y los numerados de 1 a 35 corresponden a los WTs. En cada arista del árbol se ha incluido la siguiente información: índice del tipo de cable usado en el circuito, cantidad de wind-turbines que entregan su potencia a través del mismo y recorrido en metros de ese circuito sobre el plano físico. El diseño anterior usó el siguiente conjunto de tipos de cable como referencia.

Tabla 1: Tipo y costo de cables en función del número de wind-turbines

Número de WTs (n)	1	2	3	4	5	6	7	8	9	10
Tipo de cable (k)	1	1	1	2	2	3	4	4	5	6
Máximo de WTs (m _k)	3	3	3	5	5	6	8	8	9	10
Costo por metro (c _k)	620	620	620	968	968	1,209	2,668	2,668	3,150	4,931
Caída x 1000m (%)	0.095	0.190	0.285	0.173	0.216	0.182	0.151	0.173	0.136	0.101

A modo de ejemplo sobre cómo interpretar los datos de Tabla 1. Los circuitos implementados con el tipo de cable 1, soportan la potencia de hasta 3 wind-turbines y tienen un costo por metro de 620. La caída de tensión cada 1000mts depende del número de WTs, siendo: 0.095 %, 0.190 % y 0.285 % para 1, 2 y 3 WTs respectivamente, para el tipo de cable 1.

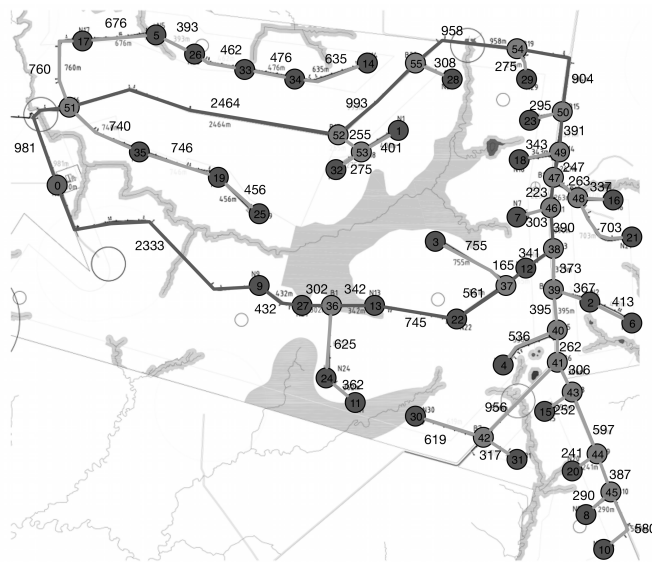


Figura 3: Estructura física de la red

El diseño baseline se hizo sobre el mapa de caminos de Figura 3. El mapa está a un arco de ser conexo minimal, ya que de hecho embebe una única componente “dos conexas”, en la figura de un ciclo destacado en magenta. En el mismo están marcadas las posiciones del collecting-point (índice 0), de las wind-turbines (índices 1 a 35) y de los restantes 20 puntos de bifurcación. La

figura también incluye las distancias físicas en metros de cada tramo.

A efectos de la optimización, la caminería constituye un dato del problema. No obstante, veremos posteriormente que algunas de las optimizaciones más importantes sí tuvieron asociados cambios, manualmente introducidos al mapa físico. Por separado, las figuras 2 y 3 resultan simples, pero el esquema luego de superponer una configuración sobre la otra (ver Figura 4) da una idea más cabal de la complejidad detrás de un simple cambio de configuración en la red.

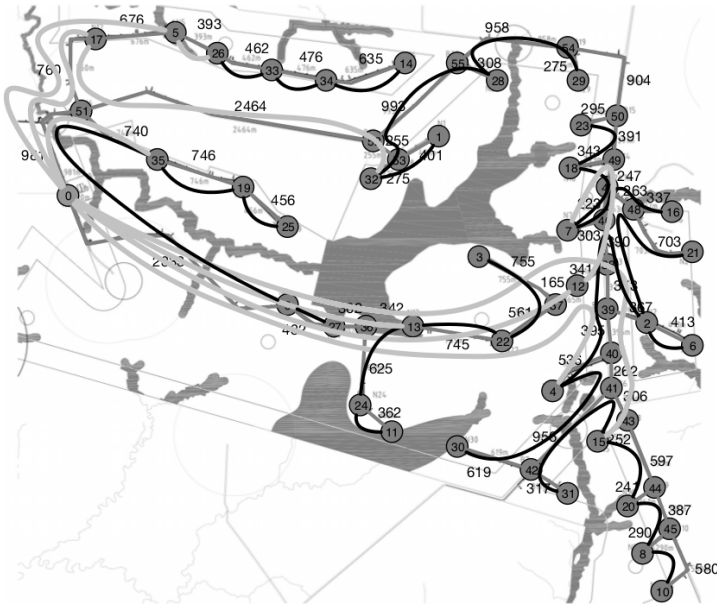


Figura 4: Resultado de superponer el nivel lógico sobre el físico. Los conductores de distinto tipo se diferencian con colores

Tabla 2: Tipo y costo de zanja en función del número de circuitos

Tipo de zanja (nc)	1	2	3	4	5	6	7	8	9	10
Número de cables	1	2	3	4	5	6	7	8	9	10
Costo x metro (c(nc))	731	1,304	1,922	2,540	3,158	3,776	4,394	5,012	5,630	6,249

En la Tabla 2 se presentan los datos de costo asociados a las zanjas. Hay un tipo de zanja para cada número de circuitos en un tramo físico dado. Instalar una zanja implica remover tierra para luego enterrar una estructura de canalizaciones dentro de la cual se despliegan los circuitos. El tamaño y costo de las canalizaciones resultó ser proporcional al número de circuitos que contiene, y como el costo de enterrar la zanja es proporcional a su tamaño, el total resulta aproximadamente lineal. Para terminar de definir la instancia del problema, el costo por circuito de entrada al CP, ccp es 770,878.

A efectos comparativos incluimos en la Tabla 3, los datos de performance de la solución original del problema por cada segmento de costo. El problema general antes descrito es el de una red en *overlay*, por lo que aun siendo conocidas las conexiones eléctricas (lógicas) podría haber más de un camino físico para implementar cada una. A ese nivel de generalidad y para las dimensiones de la aplicación en cuestión, el problema se vuelve intratable en forma exacta. Como referencia a las dificultades del uso de métodos exactos para resolver problemas de redes en *overlay* sugerimos ver [5].

Tabla 3: Performance de la solución baseline por segmento de costo

Costo Circuitos de Entrada	5,396,146	5.6 %
Costo Total de Cables	49,667,984	51.8 %
Costo Total de Zanjas	40,884,112	42.6 %
Costo Total Construcción	95,948,242	100 %

En esta instancia particular, se puede verificar que aun cuando existe un camino físico complementario (uno solo), normalmente éste es de peor calidad que el de menor distancia. Conocer de antemano el camino potencial de cada arista simplifica sensiblemente la formulación exacta del problema, que pierde ahora su condición de red en *overlay* (se vuelve de una sola capa).

3. Modelo Combinatorio

Antes de pasar a la formulación combinatoria, explicamos cómo se ha simplificado el costo de las zanjas para incluirlo en el modelo. Para el costo de una zanja se usa el resultado de una regresión lineal sobre los datos de Tabla 2, que determina que el costo por metro para enterrar nc circuitos es aproximadamente $c(nc) = 85.8 + 615.6 \cdot nc$. Llamaremos ct al costo variable de la zanja (615.6). A priori, de usarse uniformemente todos los tipos de zanja en la construcción, el error relativo esperado sería de 0.6 %.

También despreciaremos los costos fijos de instalar una zanja (85.8), ya que a lo sumo es una sola la arista física no usada en el despliegue (si fueran dos o más la red física no sería conexa), y el largo medio de las aristas respecto al total, es aproximadamente 1.8 %.

3.1. Modelo combinatorio para una capa

El modelo combinatorio presentado en esta sección está basado en el supuesto que se conoce de antemano el camino físico de cada circuito a imple-

mentar, algo que ya hemos visto se ajusta a la instancia base.

En lo que respecta a la red, son dos los grafos a considerar. En primer lugar está el conjunto de circuitos potenciales a implementar $G_L = (V, E_L)$, para el cual buscamos un árbol $T \subset G_L$, que corresponde a la red eléctrica. Recordamos que $V = \{CP\} \cup WT$, y que el índice de CP es 0. En el otro nivel está el grafo físico $G_P = (V \cup JP, E_P)$, que agrega los *junction-points* o *puntos de bifurcación* como nodos. Dado un circuito entre los nodos i y j , se supone conocido su camino (su recorrido físico). Denominaremos d_{ij} a la distancia física entre $i, j \in V$, y $tr_{ij}^f \in \{0, 1\}$ al dato correspondiente a que la arista física $f \in E_P$ es usada por el circuito entre $i, j \in V$. Dado el grafo no dirigido $G_L = (V, E_L)$ construimos el grafo dirigido $G = (V, E)$, que surge de duplicar todas las aristas de E_L (usando ambas direcciones), excepto aquellas que tienen a CP por uno de sus extremos, en cuyo caso solamente se usan las aristas incidentes a CP .

$$\left\{ \begin{array}{ll} 3 \leq \sum_{j0 \in E} x_{j0} \leq 10, & (i) \\ \sum_{ij \in E} x_{ij} = 1 & i \in V \setminus \{0\}, \quad (ii) \\ \sum_{ij \in E} x_{ij} \leq 2 & j \in V \setminus \{0\}, \quad (iii) \\ x_{ij} \in \{0, 1\} & ij \in E. \end{array} \right. \quad (1)$$

Las ecuaciones de (1) contienen parte de las reglas para construcción de la red eléctrica. La variable booleana x_{ij} indica que el arco ij es usado en la construcción del LWTG, con el dato además del sentido en el que circula la potencia hacia CP. Las restricciones (i), fuerzan a que el número de circuitos de entrada al CP esté entre 3 y 10 (requerimiento del pliego). Las restricciones en (ii) imponen que toda wind-turbine entregue su potencia a otra WT o al CP, mientras que (iii) se traduce en que de ninguna WT pueden *colgar* más de dos WTs. Observar que como el CP solamente tiene aristas incidentes, cumplir (ii) en todos los WTs implica que $\sum_{ij \in E} x_{ij} = |V| - 1$, esto es, que el resultado tiene $|V| - 1$ aristas (35 en este ejemplo). Agregar restricciones que garanticen conectividad a la solución, asegurarían que la construcción es un árbol, ya que todo subgrafo de G conexo con $|V| - 1$ aristas, debe serlo. Tanto para garantizar la conectividad del grafo resultante, como para cuantificar la carga eléctrica en cada circuito, complementaremos las restricciones de (1) con las detalladas a continuación.

$$\left\{ \begin{array}{l} - \sum_{ij \in E} y_{ij} + \sum_{jk \in E} y_{jk} = 1 \quad j \in V \setminus \{0\} \quad (i) \\ (|V| - 1)x_{ij} - y_{ij} \geq 0 \quad ij \in E \quad (ii) \\ y_{ij} \geq 0 \quad ij \in E \quad (iii) \end{array} \right. \quad (2)$$

Las variables y_{ij} representan el “flujo de potencia” en el arco ij , esto es, la cantidad de wind-turbines que entregan su potencia por ese arco en la configuración que surge de aquellos x_{ij} que valen 1. Las ecuaciones (i) establecen que la diferencia entre el flujo saliente de j y el entrante al mismo es 1, siempre que $j \neq CP$, lo que transforma al último en el único sumidero posible. Debido a (iii), las variables y_{ij} no pueden tomar valores negativos, y las restricciones (ii) aseguran que de ser positivo el valor de y_{ij} , el x_{ij} asociado debe valer 1. El problema de fondo comparte muchas características con el DCMST (Degree Constrained Minimum Spanning Tree Problem), y el modelo exacto resultante de unir (1) y (2) es similar a otros modelos conocidos (ver [1, 4]). Pero hay una diferencia destacable entre ambas versiones. En la formulación aquí presentada, la conectividad necesaria para que el resultado sea un árbol, se ha introducido mediante un sub-problema de flujo, lo que reduce la cantidad de restricciones a un número no exponencial. Finalmente, el bloque de ecuaciones (3) presiona los valores de y_{ij} hacia el mínimo posible.

$$\left\{ \begin{array}{l} \text{mín } ccp \sum_{j0 \in E} x_{j0} + \sum_{ij \in E} \sum_{k=1}^K d_{ij} c_k \cdot \eta_{ij}^k + ct \sum_{f \in E_F} \sum_{ij \in E} d_f tr_{ij}^f \cdot x_{ij} \\ y_{ij} \leq \sum_{k=1}^K m_k \cdot \eta_{ij}^k, \quad ij \in E \quad (i) \\ \sum_{k=1}^K \eta_{ij}^k = x_{ij}, \quad ij \in E \quad (ii) \\ \eta_{ij}^k \in \{0, 1\} \quad ij \in E, k = 1, \dots, K. \end{array} \right. \quad (3)$$

Los términos del objetivo a optimizar en (3) corresponden respectivamente a: el costo de los circuitos de entrada al collecting-point, el costo de los cables para la estructura de circuitos elegida, y el costo de las zanjas necesarias para enterrar esos circuitos (d_f es el largo de cada tramo físico). Las variables binarias η_{ij}^k indican que se ha usado el cable de tipo k para la implementación del circuito ij . Podemos verificar que las restricciones (ii) en (3) garantizan que en caso de implementar un circuito, éste haga uso de un sólo tipo de cable, mientras que (i) asegura que la capacidad del cable en cuestión es suficiente para soportar el flujo eléctrico en ij , esto es, que el calibre del tipo de cable para implementar

el circuito sea el mínimo correcto. La caída de tensión complica sensiblemente el modelo, y se relajó para hallar soluciones, validando a posteriori que la solución la cumplía.

3.2. Soluciones exactas al modelo combinatorio

La Figura 5 esquematiza la solución exacta para la instancia baseline presentada en la Sección 2.1. La misma fue encontrada usando CPLEX-12.6.3, en un servidor HP ProLiant DL380 G7 con 32GB de RAM, y 16 cores distribuidos en 4 procesadores Intel(R) Xeon(R) E5640@2.67GHz.

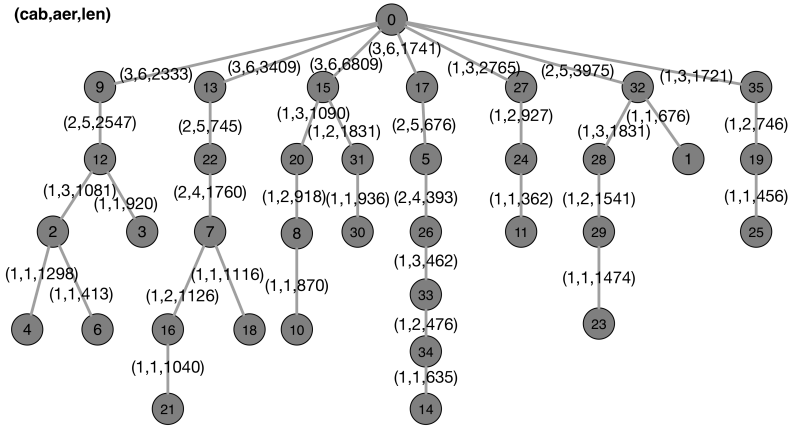


Figura 5: Layout óptimo hallado para la instancia baseline

Complementariamente detallamos en Tabla 4 los resultados de costos asociados, segmentados con el mismo criterio que en Tabla 3.

Tabla 4: Performance de la solución óptima para la instancia base

Costo Circuitos de Entrada	5,396,146	6.5%
Costo Total de Cables	43,612,776	52.2%
Costo Total de Zanjias	34,499,935	41.3%
Costo Total Construcción	83,508,857	100%

Del comparativo entre ambas tablas se destaca que el *spread* entre la óptima y la manualmente construida es de 14.9%. Además el peso relativo de cada segmento respecto al total es bastante regular. Más sutil, aunque igualmente importante son los siguientes diferencias: el layout representado en Figura 2 es eléctricamente balanceado (todos los circuitos al CP tienen 5 WTs asociados), y la solución solamente usa dos tipos de cables, lo que contribuye a la eficiencia dado que los cables se importan en bobinas de muchos metros, y la diversificación de tipos promueve el desperdicio (restos) de conductores. De hecho,

la solución original era de excelente calidad, y recién fue derrotada luego de experimentar con diversas combinaciones de heurísticas.

En cualquier caso y como veremos a continuación, las mejoras más significativas a la solución surgieron de explorar instancias con distintos tipos de conductores, combinadas con cambios menores en la red física, que enriquecían su topología y forzaban por tanto a abandonar el modelado exacto, por la ahora ineludible naturaleza *overlay* del problema.

4. Uso de Heurísticas

En esta sección se describen dos heurísticas complementarias, que permitieron resolver el problema en instancias con cambios al mapa físico que desnaturalizaban las premisas del modelo combinatorio anterior.

4.1. Heurística greedy

Esta heurística se inspira en el *Algoritmo de Prim*, aunque integra la naturaleza multicapa del problema, sus múltiples restricciones (filtra configuraciones no factibles) y la compleja estructura de costos de la solución. Las instancias quedan definidas por los mismos datos que en el modelo anterior, pero en este caso el algoritmo construye el árbol conjuntamente con los caminos físicos para los circuitos.

A partir del grafo $G_L = (V, E_L)$, se separan los nodos en dos conjuntos: $A_1 = \{CP\}$ y $A_1^c = V \setminus A_1$, correspondiendo el último al conjunto de generadores para la configuración inicial. En cada paso n se busca la arista uv (circuito) de $(A_n, A_n^c) \cap E_L$ cuya anexión minimize el incremento del costo de la solución. Supongamos que $u \in A_n$ y $v \in A_n^c$, en cuyo caso los conjuntos se actualizan según $A_{n+1} = A_n \cup \{v\}$ y $A_{n+1}^c = A_n^c \setminus \{v\}$.

Se observa que al igual que en Prim, A_n es un árbol con raíz CP para todo n . Calcular los incrementos de costo para determinar el siguiente circuito a anexar implica: computar un shortest-path-algorithm (Dijkstra) en la red física usando los costos marginales de expandir cada zanja, evaluar los costos derivados de reconfigurar los conductores en el árbol A_n aguas arriba (agregar un WT puede elevar la sección requerida de los conductores y/o llevar la caída de tensión más allá de lo permitido), o agregar el costo ccp en caso que el circuito sea directamente hacia el CP. La elección del nuevo circuito viene acompañada de un camino físico para el mismo, que pasa a ser un atributo permanente. Si durante el cómputo de los incrementos se detecta una arista cuya anexión viola el límite de circuitos de entrada a un WT, o hace que la

caída de tensión no pueda factibilizarse con cambios de conductores, esa arista se descarta como candidata. Luego de $|V| - 1$ pasos se llega a un conjunto $A_{|V|}$ que contiene todos los generadores.

No se esperaba optimalidad de esta heurística porque a diferencia de Prim los costos de las aristas cambian entre iteraciones, y al final difieren substanciales de aquellos con los que se tomó la decisión en primer lugar. No obstante, la performance mostrada por este algoritmo fue más discreta de lo esperado. De hecho, el costo de la solución construida para la instancia base fue de 107,700,472, bastante lejos del modelo combinatorio (83,508,857), pero incluso lejos del diseño manual original (95,948,242).

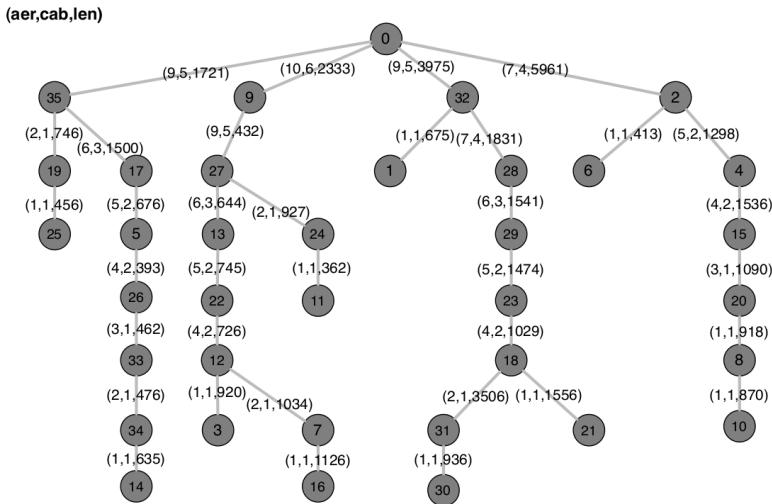


Figura 6: Layout hallado para la heurística greedy

De un análisis más detallado del algoritmo y sus construcciones, surge que el mismo está negativamente condicionado por las decisiones tomadas tempranamente, que usan información local en una construcción parcial. Se comprueba que de fijar manualmente los nodos conectados al CP, usando por ejemplo los que eligió el combinatorio (Figura 5), la heurística greedy construye el resto del árbol satisfactoriamente.

La característica anterior nos llevó a diseñar una heurística con visión más global, más precisa al momento de determinar las conexiones estratégicas, las de conductores de mayor sección, las dirigidas desde un WT hacia el CP.

4.2. Flujo de costo óptimo en una capa

Otra de las heurísticas utilizadas consistió en relajar el problema buscando colapsar los niveles lógico y físico en uno solo, a efectos de hacer tratable al problema. Las relajaciones alcanzan a: la cantidad máxima de WT's en

subtending, los límites en las caídas de tensión, y en suponer que en cualquier punto del mapa físico se podría cambiar de conductores (lo que agrega los puntos de bifurcación a las WTs).

Bajo esas premisas y para un número n conocido de WTs que entregan su potencia en un tramo físico, determinar la combinación óptima de conductores, esto es, el número x_k de conductores de cada tipo ($k = 1 \dots K$) a usar para minimizar los costos totales de zanjas y cables, es un problema combinatorio sencillo que puede tratarse recursivamente, o simplemente resolviendo (4):

$$\left\{ \begin{array}{l} \text{mín } \sum_{k=1}^K c_k x_k + 615.6 \sum_{k=1}^K x_k \\ n \leq \sum_{k=1}^K m_k x_k \\ 10 \geq \sum_{k=1}^K x_k \\ x_k \geq 0, x_k \in \mathbb{Z} \end{array} \right. \quad (4)$$

Por ejemplo, para $n = 12$ el costo por metro de la combinación óptima es \$3,735, y la combinación en sí consiste en usar dos conductores de tipo 3 ($x_3=2$) y por tanto una zanja para dos conductores. Cuando $n = 13$ el costo por metro asciende a \$4,488.6 y surge de usar tres circuitos (y la zanja correspondiente): uno de ellos con cables tipo 1 y los otros dos con tipo 2 (i.e. $x_1=1, x_2=2$).

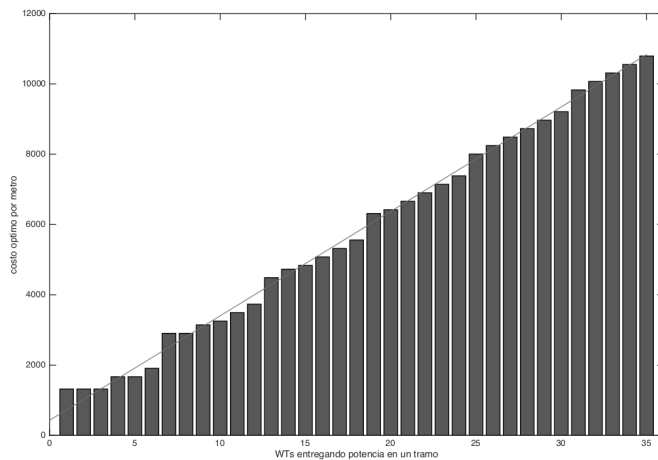


Figura 7: Costos por metro para la combinación más económica de cables y zanjas en cada tramo, para distintos números de aerogeneradores entregando su potencia a través del mismo

Se observa en Figura 7 que el costo por metro de la configuración óptima en un tramo, para un número n de aerogeneradores entregando su potencia a

través del mismo, es aproximadamente lineal, especialmente cuando más alto es n . La recta en rojo es la imagen de $an+b$, con $a=297.2$ y $b=430.2$. Por tanto, con la relajación antes mencionada y la aproximación lineal para el costo en los tramos, se obtiene un problema de flujo máximo de costo mínimo, o más precisamente un problema de flujo fijo de costo mínimo, en el que cada WT inyecta una unidad de flujo mientras que en los JPs el balance entrada-salida debe coincidir. Las ecuaciones de flujo son similares a las de (2), aunque sobre el grafo G_P y agregando $-\sum_{ij \in E} y_{ij} + \sum_{jk \in E} y_{jk} = 0$, cuando $j \in JP$.

Por el bajo error incurrido al linealizar, y por ser el nuevo problema una relajación del original, su solución nos da al mismo tiempo: información valiosa para construir una solución concreta, y una buena estimación de la cota inferior esperable. Luego de resolver el problema de flujo de costo mínimo, obtenemos el flujo de potencia ideal de cada arista física en el problema relajado, y también la combinación ideal de conductores en cada tramo. Las configuraciones son en general impracticables, ya que suelen aparecer cambios de conductores en puntos de bifurcación (algo sólo realizable en los transformadores situados en la base de WTs), o cambios entre más de tres conductores en un aerogenerador, algo tampoco soportado ya que en los WTs se puede intercambiar potencia a lo sumo entre dos circuitos de entrada y el de salida.

No obstante, al poner foco en los conductores de mayor calibre (los más caros), se consigue una buena idea de cuáles son los generadores a conectar al CP. Explorando alternativas para esas conexiones y dejando en manos de la heurística greedy el completar la construcción de una red eléctrica factible, se consiguieron soluciones de muy buena calidad. Usando casos tratables para el modelo exacto de la Sección 3, y por disponer entonces del óptimo se puede evaluar el *spread* óptimo-heurístico en algunas instancias. El *spread* resultó siempre menor a un 5 %, lo que muestra la calidad de la aproximación.

5. Resultados y Conclusiones

Si bien el diseño optimizado detallado en Sección 3.2 ya representaba un ahorro significativo (14.9 %), se trabajó intensamente junto a los técnicos de Teyma para llegar aún más lejos. Los ahorros más importantes surgieron en realidad de explorar cambios en el juego de conductores a usar, y de modificaciones en el plano físico. Concretamente se optimizó en ocho escenarios, que surgieron de combinar cuatro juegos de cables con dos planos físicos. El primer juego de conductores es el detallado en Tabla 1. En lo que al mapa físico refiere, el primero es el representado en Figura 3, mientras que el otro

es una variante del mismo, que busca incrementar la conectividad mediante cambios mínimos. Concretamente, el segundo plano es el de Figura 8.

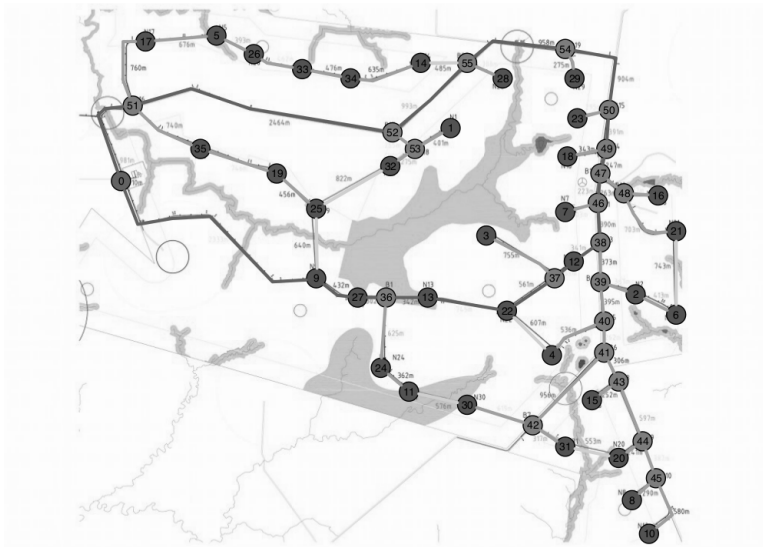


Figura 8: Estructura física ampliada

Las instancias con el mapa 1 fueron resueltas usando el modelo combinatorio de Sección 3, salvo la última, en la que se violaba la caída de tensión y hubo que refactibilizar con la heurística greedy. Para todas las instancias asociadas al mapa 2, se usó la combinación de heurísticas descrita en Sección 4.

Tabla 5: Instancias exploradas para elegir el layout a implementar

	cables 1	cables 2	cables 3	cables 4
m	5,396,146	3,083,512	3,854,390	3,854,390
a	43,612,776	37,780,944	41,011,092	37,908,840
p	34,499,935	37,467,316	33,255,310	33,255,310
1	83,508,857	78,331,772	78,120,792	75,018,540
m	4,625,268	3,083,512	3,083,512	3,083,512
a	44,939,352	27,234,446	41,376,440	30,265,446
p	29,551,638	25,928,002	25,928,002	25,734,326
2	79,116,258	56,245,960	70,387,954	59,083,284

Las soluciones remarcadas en gris, se descartaron. Las del segundo juego de cables debido a los descartes de conductores. Los cables se compran en bobinas de muchos metros, por lo que en algunas configuraciones los descartes de conductores resultaban inconvenientes. La solución correspondiente al $(map1, cables4)$, aunque luego de refactibilizada estaba dentro de la caída de tensión admisible, se entendió que se estaba peligrosamente cerca del límite, y aumentar aún más la sección de los conductores la hacía más cara que

(*map1,cables2*). El layout finalmente implementado fue (*map2,cables4*).

Vale la pena destacar que la construcción manual original era un 62.4% más cara que la implementada realmente o, lo que es equivalente, la última resultó un 38.4% más barata. Siendo estrictos, el ahorro es algo más bajo porque implicó modificaciones en caminería, pero las mismas era de menor entidad, y los ahorros finales resultaron millonarios.

La experiencia en este proyecto es un valioso ejemplo de cómo el uso de modelos formales y la asistencia numérica permiten ahorros significativos en algunos problemas de diseño, especialmente en aquellos casos donde la intuición suele fallar. Tal es el caso de los problemas de diseño de redes en *overlay*. No obstante, también se destaca que los ahorros más significativos se consiguieron explorando instancias, iterando a partir de las construcciones previas y el análisis de las debilidades en las soluciones realizado por los expertos de Teyma. Lo previo reafirma el concepto que los mejores resultados surgen de la interacción entre la informática y el conocimiento de los expertos.

Como trabajo futuro en esta línea nos planteamos implementar una metaheurística, para hibridizar con las heurísticas anteriores en aras de lograr una exploración automática y más exhaustiva del espacio de soluciones.

Referencias

- [1] L. H. Bicalho, A. S. da Cunha, y A. Lucena. Branch-and-cut-and-price algorithms for the degree constrained minimum spanning tree problem. *Computational Optimization and Applications*, 63(3):755–792, Apr 2016.
- [2] F. M. Gonzalez-Longatt, P. Wall, P. Regulski, y V. Terzija. Optimal electric network design for a large offshore wind farm based on a modified genetic algorithm approach. *IEEE Systems Journal*, 6(1):164–172, March 2012.
- [3] S. Lundberg. Configuration study of large wind parks. Technical report, Department of Electric Power Engineering, Chalmers University of Technology, Göteborg, Sweden, 2003.
- [4] S. C. Narula y C. A. Ho. Degree-constrained minimum spanning tree. *Computers & Operations Research*, 7(4):239 – 249, 1980.
- [5] C. Risso y F. Robledo. Using GRASP for designing a layered network: a real IP/MPLS over DWDM application case. *International Journal of Metaheuristics*, 2(4):392–414, 2013.

UN ENFOQUE ANALÍTICO PARA LA CONFECCIÓN DEL RANKING FIFA Y LA REALIZACIÓN DEL SORTEO DEL MUNDIAL DE FÚTBOL

SEBASTIÁN CEA^{*}
GUILLERMO DURÁN^{* **}
MARIO GUAJARDO^{***}
DENIS SAURÉ^{*}
GONZALO ZAMORANO^{*}

Resumen

Este trabajo analiza los procedimientos de ranking y sorteo de grupos de la Copa del Mundo utilizados por la FIFA. Calibrando un modelo predictivo, establecemos un ranking ideal de referencia sobre el cual juzgamos el desempeño de simples propuestas de cambio al procedimiento actual de ranking, las que a su vez están guiadas por un análisis cualitativo y estadístico del procedimiento actual. Luego analizamos el uso del ranking de equipos para el sorteo de los grupos en la fase final de la Copa del Mundo. Proponemos un modelo de programación lineal entera para balancear la *dificultad* de los grupos en dicha fase final. Ilustramos este enfoque utilizando datos de la Copa del Mundo 2014.

Palabras Clave: Investigación de Operaciones en deportes, Ranking, Copa del Mundo FIFA, Fútbol

^{*}Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile

^{**}Departamento de Matemática e Instituto de Cálculo, FCEN, Universidad de Buenos Aires, Buenos Aires, Argentina

^{***}Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway

1. Introducción

La Copa del Mundo FIFA es el evento futbolístico más popular del mundo. Organizado por la FIFA en ciclos de cuatro años, el evento congrega a las selecciones de más de 200 naciones asociadas a la FIFA. La Copa del Mundo se juega en dos fases: la fase clasificatoria y la fase final (comúnmente conocida como *Mundial*). En la fase clasificatoria, los equipos compiten dentro de sus confederaciones continentales por una cierta cantidad de cupos para la fase final. Actualmente, existen seis confederaciones: Confederación Sudamericana de Fútbol (CONMEBOL), Unión de Asociaciones Europeas de Fútbol (UEFA), Confederación Asiática de Fútbol (AFC), Confederación Africana de Fútbol (CAF), Confederación de Fútbol de Norte, Centroamérica y el Caribe (CONCACAF), y Confederación de Fútbol de Oceanía (OFC). En su edición más reciente, la fase final se disputó en Brasil en 2014 y congregó a 32 equipos, 3,4 millones de espectadores en los estadios y 3,2 billones de televidentes alrededor del mundo [8, 9].

En el formato actual de torneo, los 32 equipos clasificados al Mundial son agrupados en 8 grupos de 4 equipos cada uno. La formación de estos grupos ocurre de acuerdo a un sorteo. Una vez que los grupos han sido definidos, los equipos de un mismo grupo juegan una fase de tres partidos, todos contra todos. Los dos de mejor desempeño en cada grupo clasifican a una fase de *playoffs* que define al campeón.

En la metodología actualmente utilizada por la FIFA, el sorteo de los grupos del Mundial ocurre en base a criterios de desempeño y a criterios geográficos. El desempeño de los equipos se mide al momento del sorteo (seis meses antes del Mundial) por su posición en el *ranking FIFA* [10]. Este ranking se construye de acuerdo a una metodología que otorga puntos a cada equipo por los partidos disputados en los últimos cuatro años. Los siete equipos mejor rankeados, además del país anfitrión, son designados como sembrados (o cabezas de serie) de los grupos. Los restantes equipos son ubicados en distintos agrupamientos (llamados “pots”), según la confederación continental a la que pertenecen, y luego son sorteados a alguno de los grupos que encabezan los sembrados. Este procedimiento asegura que los sembrados no jueguen entre sí durante la fase de grupos y, además, promueve la diversidad geográfica entre los equipos que componen un mismo grupo. Si bien estos criterios parecen razonables, tanto el ranking como el sorteo han sido objetos de crítica. Por ejemplo, para el Mundial 2014, circularon varias notas de prensa deportiva y

opiniones de aficionados criticando que Colombia, Bélgica y Suiza consiguieran una posición tan alta en el ranking FIFA, a punto tal que les permitió ser cabezas de serie. Otras críticas apuntaron a la aparentemente amplia diferencia en el grado de dificultad de los distintos grupos sorteados para el Mundial. Más allá de esta evidencia anecdótica, la comunidad académica se ha gradualmente involucrado en el debate mundialístico mediante trabajos científicos. La mayor parte de los trabajos se ha enfocado en la predicción de resultados (e.g. [14], [4], [17], [7]). Algunos se han enfocado específicamente en el ranking FIFA (e.g. [15], [19], [13]) con el fin principal de estudiar o comparar su poder predictivo más que proponer modificaciones a su metodología de cálculo. Recientemente, en [6] reportan técnicas en investigación de operaciones aplicadas a la calendarización de las clasificatorias mundialistas sudamericanas. Por otro lado, la literatura sobre la formación de grupos del Mundial es muy escasa. A nuestro conocimiento, el único esfuerzo académico en esta línea es [12], que discute deficiencias del sistema de formación de grupos actualmente utilizado por la FIFA y elabora algunas propuestas de mejora.

En este artículo analizamos de forma crítica los procedimientos de construcción del ranking y de sorteo de los grupos del Mundial de la FIFA. En primer lugar analizamos la construcción del ranking FIFA. Para esto, tras demostrar una serie de falencias del procedimiento actual, establecimos un ranking *ideal* de referencia, que nos permite calificar el desempeño de posibles modificaciones sencillas a dicho procedimiento. El ranking aludido representa el resultado esperado de un torneo doble round-robin entre los equipos integrantes de la FIFA, el cual es obtenido mediante simulación de Monte Carlo. La componente principal de este procedimiento es la calibración de una variante del modelo predictivo propuesto por [14], y que es la base de la mayor parte de los modelos predictivos propuestos en la literatura. Para esto, usamos datos de partidos entre selecciones jugados entre 2009 a 2013.

Respecto a las propuestas de modificación al procedimiento de ranking, en un primer paso construimos modelos de regresión logística multinomial que nos permiten identificar variables que explican los resultados de partidos entre selecciones de fútbol. En base a nuestro análisis y los resultados de estos modelos, proponemos modificaciones simples a la metodología del ranking FIFA actual, las que son evaluadas utilizando el ranking ideal de referencia.

En segundo lugar, el trabajo estudia el uso de un modelo de programación lineal entera mixta para la formación de grupos del Mundial. Dicho modelo toma en cuenta criterios geográficos y de balance de dificultad de los grupos. Nuestros resultados muestran que es posible mejorar considerablemente la metodología actual de la FIFA para el ranking de equipos y la formación de grupos del Mundial.

El resto del artículo está organizado en cinco secciones. La Sección detalla la construcción del ranking ideal de referencia en base al modelo predictivo de [14]. En la Sección 2.1 analizamos el ranking FIFA actual e identificamos variables importantes para la predicción de resultados. En base al análisis y al resultado de esos modelos, la Sección 3.2 elabora nuevas metodologías de ranking y testea su desempeño mediante simulaciones. La Sección 4.2 formula el modelo de programación lineal entera mixta para la formación de grupos del Mundial y presenta sus resultados. El artículo concluye con una discusión final en la Sección 5.3.

2. Ranking ideal de referencia

Una parte significativa de las críticas contra el procedimiento actual se debe a la inclusión en el top 10 de equipos que tradicionalmente no han obtenido buen desempeño en competencias importantes (ver [15]). Dichas críticas son en esencia subjetivas, por lo que se requiere un procedimiento objetivo para comparar dos rankings. Esto es particularmente valioso en nuestro caso: nuestras propuestas buscan modificar de manera simple las reglas de puntuación en el procedimiento del ranking, de forma que estas tengan potencial de ser implementadas; cada modificación a las reglas impacta en el ranking, y es necesario entonces evaluar la bondad de dicha modificación.

Cualquier criterio de comparación de rankings lleva a la necesidad de un ranking *ideal* o *de referencia*, por lo que podemos pensar en el problema de construir el mismo. En este sentido, un ranking de referencia debiese reflejar la calidad de las selecciones, de forma que un equipo con mejor ranking debiese tener mayor probabilidad de derrotar a un equipo con menor ranking en terreno neutral. Este ranking debiese ser obtenido a partir de datos históricos, sin embargo encuentros entre selecciones normalmente ocurren en el marco de competencias internacionales de forma esporádica, e influenciados por la estructura de los torneos. El problema de inferir datos históricos a partir de comparaciones pareadas se remonta a un antiguo trabajo de [2]. Por otra parte, [13] presentan un estudio de sistemas de ranking para fútbol. En dicho trabajo, la calidad de un ranking está asociada a su capacidad predictiva, la cual se infiere a partir de un modelo de regresión logística que utiliza como variable explicativa el ranking de los equipos y la condición de localía. Tomando en cuenta esta idea, considerando que el resultado de un torneo varía considerablemente dependiendo del formato del mismo, y que el formato *round-robin* (o *todos contra todos*) es el que maximiza la correlación entre el mejor equipo y resultado final ([18]), en este trabajo proponemos utilizar como ranking de

referencia el resultado esperado de un torneo doble round-robin que considera el factor de localía. Para obtener un buen poder predictivo, calibramos una variación del modelo de [14] para simular el resultado de partidos entre dos equipos. Con esta herramienta a nuestra disposición procedemos a calcular el ranking esperado de la competencia round-robin a través de simulación de Monte Carlo. A continuación presentamos nuestra adaptación del modelo de [14].

2.1. Modelo Predictivo

Sean $X_{A,B}$ e $Y_{B,A}$ el número de goles convertidos por el equipo A y el equipo B , respectivamente, en un partido donde A es local. Siguiendo a [14] suponemos que $X_{A,B}$ e $Y_{B,A}$ son variables aleatorias independientes con distribución de Poisson con tasas $\lambda_{A,B}$ y $\gamma_{A,B}$, respectivamente. Además, suponemos que en un partido dado la tasa con la cual un equipo anota depende de su capacidad ofensiva, la capacidad defensiva del equipo rival y la localía. En particular, asumimos que:

$$\ln(\lambda_{A,B}) = a_A - d_B + \rho_h, \quad (1a)$$

$$\ln(\gamma_{A,B}) = a_B - d_A + \rho_a. \quad (1b)$$

En (1a) y (1b), los parámetros a_A y d_A representan las capacidades ofensiva y defensiva del equipo A , respectivamente. Las variables ρ_h y ρ_a representan correcciones a las tasas que toman en cuenta la localía (en un partido en terreno neutral ambas variables ρ_h y ρ_a son omitidas). Con esto, la probabilidad de obtener el resultado $\{X_{A,B} = m, Y_{B,A} = n\}$, con m los goles de A y n los goles de B , cuando A y B se enfrentan, está dada por:

$$P(X_{A,B} = m, Y_{B,A} = n) = \frac{e^{-\lambda_{A,B}} (\lambda_{A,B})^m}{m!} \cdot \frac{e^{-\gamma_{A,B}} (\gamma_{A,B})^n}{n!}. \quad (2)$$

El modelo descrito arriba cuenta con dos parámetros por cada equipo, más aquellos asociados a los factores de localía y visita. Los parámetros del modelo son calibrados usando *maximum likelihood estimation* (MLE), utilizando datos de todos los partidos oficiales FIFA entre los años 2009 y 2013.

Para obtener nuestro ranking de referencia simulamos 10^5 torneos doble *round-robin* en terreno neutral considerando sólo a los equipos participantes en la Copa Mundial 2014, y promediamos el número de puntos obtenidos por cada equipo cuando una victoria, empate y derrota otorgan 3, 1 y 0 puntos respectivamente. La Tabla 1 muestra el ranking de referencia y una estimación de la probabilidad de ser campeón de cada país obtenida a través de las simulaciones.

Tabla 1: Simulación de torneos round-robin entre equipos participantes en Copa Mundial 2014.

Orden	País	Prob. Campeón (%)	Orden	País	Prob. Campeón (%)
1	Brasil	24,45	17	Ghana	0,17
2	España	20,49	18	Costa de Marfil	0,15
3	Argentina	15,33	19	Corea del Sur	0,076
4	Holanda	12,49	20	Suiza	0,072
5	Alemania	5,3	21	Japón	0,068
6	Inglaterra	5,26	22	Bélgica	0,042
7	Francia	3,27	23	Australia	0,03
8	Uruguay	3,0	24	Nigeria	0,031
9	Ecuador	2,94	25	USA	0,027
10	Colombia	2,0	26	Portugal	0,026
11	Croacia	1,18	27	Honduras	0,012
12	Rusia	1,07	28	Camerún	0,003
13	Italia	1,04	29	Costa Rica	0,003
14	México	0,95	30	Argelia	0,002
15	Chile	0,19	31	Grecia	0,001
16	Bosnia	0,18	32	Irán	0

3. Análisis del ranking FIFA

Con un ranking referencial disponible, pasamos a estudiar el procedimiento actual del ranking FIFA. Introducida en 2006, la metodología actual del ranking FIFA se basa en el puntaje (conocido como “rating”) de las 211 selecciones adultas asociadas a la FIFA. El rating de un equipo se basa en el puntaje asignado a cada uno de los partidos disputados por dicho equipo, de acuerdo a la metodología que describimos a continuación.

El puntaje P asignado a un equipo por disputar un partido cualquiera está dado por la multiplicación de 4 factores:

$$P = M \cdot I \cdot T \cdot C. \quad (3)$$

El factor M denota los puntos obtenidos en el encuentro: 3 si el equipo gana el partido; 1 en caso de empate; 0 en caso de derrota. El factor I denota la importancia del encuentro: 4 si es Copa del Mundo; 3 si es fase final de torneo de Confederación (Copa América, Copa de Oro, Eurocopa, etc.) o Copa Confederaciones; 2.5 si es eliminatoria mundialista o en el ámbito de confederación; y 1 si es un partido amistoso o torneo menor de confederación. El factor T

denota la fortaleza del rival la que se calcula como 201 menos el ranking FIFA del oponente (e.g. de tal manera que jugar contra el top 1 del ranking tiene asociado un factor F igual a 200); en caso que el oponente tenga un ranking inferior al 150, la fuerza del contrincante se fija en un mínimo de 50. Finalmente, el factor C denota el promedio de la fuerza de las confederaciones a la cuales pertenecen los dos equipos que participan del partido.¹ Por ejemplo, la Tabla 2 muestra los puntos obtenidos por Honduras y Chile en el partido que disputaron en el Mundial Sudáfrica 2010, cuyo resultado fue un triunfo de Chile por 1-0.

Tabla 2: Ejemplo Puntaje FIFA.

	Honduras	v/s	Chile
Posición en ranking	38		22
M : Resultado del partido	0		3
I : Importancia del partido	4		4
T : Fuerza del competidor	178		162
Confederación	CONCACAF		CONMEBOL
Fuerza de la confederación	0.94		0.94
C : Fuerza de confederaciones		$(0.88+1)/2$	
$P = M \cdot I \cdot T \cdot C$	0		1227

Para un equipo i , definimos $G_i := \{(j_k, t_k) : k = 1 \dots\}$ el conjunto de partidos disputados por el equipo i , donde j_k y t_k denotan el rival y la fecha (medida en años) del encuentro k . Adicionalmente, para $s = 1, \dots, 4$, definimos $G_{i,s}(t)$ el conjunto de índices correspondientes a los encuentros disputados por el equipo i entre los años $t - s$ y $t - s + 1$. Sea $R_i(t)$ el rating del equipo i en la fecha t . De acuerdo al procedimiento actual, este rating se calcula como una suma ponderada de los promedios de puntos obtenidos durante los últimos cuatro años, según la siguiente fórmula:

$$R_i(t) := \sum_{s \in \mathbb{N}} \alpha_s \left(\frac{1}{\max\{|G_{i,s}(t)|, 5\}} \sum_{h \in G_{i,s}(t)} P_{i,h} \right), \tag{4}$$

donde $P_{i,h}$ denota el puntaje obtenido por el equipo i en el encuentro h . El procedimiento actual deprecia los puntajes obtenidos en el pasado de acuerdo al siguiente esquema: $\alpha_1 = 1; \alpha_2 = 0,5; \alpha_3 = 0,3; \alpha_4 = 0,2$; y $\alpha_s = 0$ para $s > 4$.

¹Este factor de fuerza se calcula de acuerdo al número de victorias que han obtenido los equipos de cada Confederación en las últimas tres ediciones de la World Cup. Los valores de fuerza antes del mundial Brasil 2014 fueron 1 para CONMEBOL y UEFA, 0.88 para CONCACAF, 0.86 para AFC y CAF, y 0.85 para OFC. Ver [10] para más detalles.

El ranking FIFA ordena mensualmente a los 211 países asociados, otorgando el ranking número 1 al equipo que posea el mayor rating, y así sucesivamente a los posteriores. El principal uso del ranking es determinar a los sembrados para el sorteo de la fase de grupos con la cual comienza la fase final de la Copa del Mundo: los 7 mejores rankeados más el país anfitrión son sembrados en los 8 grupos con los que comienza la copa. Este hecho les otorga cierta ventaja, dado que las chances de pasar la primera fase o fase de grupos se ven incrementadas dado que evitan partidos difíciles en la ronda inicial.

3.1. Deficiencias del ranking actual

Si bien la metodología actual permite rankear a los equipos de acuerdo a un cálculo relativamente sencillo, esta posee serias deficiencias que discutimos a continuación.

Partidos Amistosos. El ranking de un equipo, dado en (4), es una suma ponderada de los promedios de los puntos obtenidos en los últimos 4 años por un equipo. Considerando la baja importancia asociada a los partidos amistosos ($I = 1$), es fácil notar que a los países con buen rating no les conviene disputar este tipo de partidos. Ilustramos este punto con el caso de Chile, que en Agosto de 2014 tenía un promedio en el último año de 673,61 puntos (ver Tabla 3). Jugar un partido amistoso contra el equipo rankeado top 1 y vencerlo otorga

Tabla 3: Rating de Chile en Agosto de 2014.

Año	Promedio anual	α_t	Contribución
2014	673,61	1	673,61
2013	391,47	0,5	195,73
2012	469,7	0,3	140,91
2011	447,07	0,2	89,41
$R_{aug,2014}$			1099,66

el máximo puntaje posible para estos encuentros, equivalente a 600 puntos ($3 \cdot 1 \cdot 200 \cdot 1 = 600$), que es inferior al promedio de Chile a dicha fecha. Por lo tanto, incluso un triunfo en un partido amistoso contra el mejor equipo en ese momento disminuiría su rating y, a su vez, lo haría probablemente retroceder en el ranking mundial. Este punto quedó en evidencia en la selección de sembrados para el Mundial 2014: La Tabla 4 muestra los diez equipos que encabezaban el ranking FIFA en Octubre 2013, cuando se seleccionaron los sembrados.

Tres de los sembrados (Colombia, Bélgica y Suiza) lograron su ranking, en parte, por haber jugado menos amistosos durante el año 2013 en comparación a los otros equipos que estaban disputando esas plazas. A Octubre de 2013,

Tabla 4: Ranking FIFA Octubre 2013.

Posición	País	Puntos	Posición	País	Puntos
1	España	1513	6	Uruguay	1164
2	Alemania	1311	7	Suiza	1138
3	Argentina	1266	8	Holanda	1136
4	Colombia	1178	9	Italia	1136
5	Bélgica	1175	10	Inglaterra	1080

Colombia y Suiza habían jugado sólo 3 partidos amistosos en el último año y Bélgica lo había hecho en 4 ocasiones, mientras que Holanda e Inglaterra habían disputado 5 partidos de esta categoría e Italia lo había hecho en 6 oportunidades. Un análisis via simulación revela que equipos con menor rating hubiesen logrado ser sembrados en la copa del mundo de no haber jugado partidos amistosos el año previo al sorteo (por ejemplo, Chile, que ocupaba el puesto 12 al momento del sorteo, hubiese sido segundo en el ranking).

Depreciación de puntaje y calendarización. Un segundo problema con la metodología actual es que la depreciación anual de puntajes favorece a algunas confederaciones, debido a la calendarización de sus torneos de confederación. De lo explicado anteriormente, es claro que los equipos reciben mayor puntaje desde las copas de confederación (Copa América, Copa de Oro, Eurocopa, Copa de Asia, Copa Africana, Copa de las Naciones de la OFC), sin embargo, estas no se realizan en las mismas fechas, por lo tanto se deprecian de manera distinta: los países pertenecientes a confederaciones que juegan su torneo más cerca del mundial siguiente se ven beneficiados. Por ejemplo, equipos europeos son beneficiados respecto a países sudamericanos dado que su torneo continental se juega sólo 2 años antes del mundial siguiente, mientras que el de los sudamericanos se realiza 3 años antes (en la antesala al Mundial 2014, la Copa América se jugó en 2011 y la Eurocopa en 2012).

Puntos por resultado. En primer lugar, si bien el cálculo de M es acorde al número habitual de puntos obtenidos por partido en un torneo de fútbol (3 en caso de triunfo, 1 en caso de empate, y 0 en caso de derrota), postulamos que esto impone una diferencia desmedida en puntaje obtenido en un partido si se empata con los mejores del mundo en comparación a si se obtiene un triunfo ante equipos de medianía de tabla, en igualdad de formato de partido (ambos amistosos, o eliminatoria, o mundial). Por ejemplo empatar con el top 1 del ranking en un partido amistoso entrega como máximo 200 puntos, mientras que ganarle a un equipo ubicado en la posición 100 reporta como mínimo 255 puntos (estos puntajes dependen de las confederaciones que disputen el partido). Más allá de lo discutible de la diferencia entre los méritos de ambos

casos, lo que parece contradictorio entonces es que sea eventualmente más conveniente para un equipo disputar un partido contra un equipo mediocre en vez del honor de enfrentar al mejor del ranking.

Localía y Confederación. Además de las deficiencias sobre los factores actualmente considerados por el ranking FIFA, nos parece importante analizar la relevancia de otros factores que no son actualmente considerados. Un factor relevante en este sentido es el asociado a la condición de localía. De acuerdo a la fórmula utilizada actualmente, el puntaje asignado a un equipo en un partido es independiente de si el equipo jugó de local o visita. Sin embargo, existe consenso respecto a que los equipos mejoran su desempeño cuando juegan de local. Este *efecto de localía* ha sido reconocido en la literatura como un factor clave para explicar resultados deportivos. Particularmente en el fútbol, [5] relacionan esta ventaja de localía a la presión ejercida sobre los árbitros y el aliento de los aficionados, mientras [16] atribuye esta ventaja al mayor conocimiento del campo de juego por parte del equipo local. Otro factor a considerar es la influencia de la confederación en el cálculo de puntos para el ranking. En nuestra opinión, su relevancia no es clara a priori. Adicionalmente nos parece importante extender el concepto de localía mas allá del país de origen de un equipo y llevarlo al ámbito regional. Esto es motivado por los resultados en los torneos más importantes a nivel de selecciones adultas: en 13 de los 15 mundiales disputados en territorio americano o europeo, el campeón fue un equipo del mismo continente que el país anfitrión.

3.2. La condición de localía

Para estimar la relevancia de este factor relativo al ranking FIFA, y para sacar prescripciones que puedan transformarse en recomendaciones de mejoras concretas, en esta sección analizamos el poder predictivo de la condición de localía. Considerando los factores que se toman en cuenta en el procedimiento de ranking, procedemos a estimar directamente la probabilidad de un resultado en función de la diferencia en el ranking y el factor de localía. A diferencia de la sección anterior, en esta sección utilizamos una regresión logística multinomial (ver [7] para un enfoque alternativo), remitimos al lector a [1] para más detalles e interpretación de la regresión logística.

Sea $Y_{i,k} \in \{\text{ganar, empatar, perder}\}$ el resultado del encuentro k para el equipo i . Nuestro modelo base, al cual nos referimos como MRLn, plantea que

$$P(Y_{i,k} = x) = \begin{cases} \frac{e^{V_{i,k}}}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{victoria} \\ \frac{1}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{empate} \\ \frac{e^{U_{i,k}}}{1+e^{V_{i,k}}+e^{U_{i,k}}} & x = \text{derrota,} \end{cases} \quad (5)$$

donde definimos

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} \quad (6)$$

$$U_{i,k} = \beta_0 + \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{j_k \text{ es local}\}} + \beta_{visita} \cdot 1_{\{j_k \text{ es visita}\}} \quad (7)$$

Adicionalmente, consideramos las siguientes especificaciones para las variables $V_{i,k}$ (la modificación respectiva a $U_{i,k}$ sigue la misma lógica).

- Modelo sin factor de localía (MR).

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)).$$

- Modelo con localía de confederación (MRLnLc):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} + \beta_c \cdot C_{i,k},$$

donde la variable binaria $C_{i,k}$ toma el valor 1 si y solo si el partido k del equipo i se juega en el continente al que pertenece i e i no es local.

- Modelo con factor de confederación (MRLnFc):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{local} \cdot 1_{\{i \text{ es local}\}} + \beta_{visita} \cdot 1_{\{i \text{ es visita}\}} + \beta_{AFC} * AFC_i + \beta_{OFC} * OFC_i + \beta_{CAF} * CAF_i + \beta_{CNMBL} * CNMBL_i + \beta_{UEFA} * UEFA_i.$$

Aquí, las variables binarias AFC_i , CAF_i , OFC_i , $CNMBL_i$, $UEFA_i$ indican si el equipo pertenece o no a la respectiva confederación (se omite la confederación CONCACAF, que tomamos como confederación de referencia).

El primero de estos modelos se considera para tener un punto de referencia respecto al poder predictivo del ranking FIFA por sí solo. El segundo de los modelos se considera para establecer la influencia de la localía a nivel de confederación, mientras que el tercero busca medir la influencia de considerar variables asociadas a la confederación de los equipos (recordar que el procedimiento actual de ranking incluye dicho tipo de variable).

Para estimar los modelos, recolectamos datos sobre la totalidad de partidos jugados entre los años 2005 y Octubre del 2013, entre selecciones adultas de fútbol asociadas a alguna de las seis confederaciones continentales que son miembros de FIFA. La base de datos resultante incluye 8049 partidos, cuya caracterización se presenta en la Tabla 5.

Los modelos de regresión logística fueron implementados en el software *R Studio*. La estimación de los parámetros fue realizada de acuerdo al método de máxima verosimilitud, que intenta encontrar los $k + 1$ parámetros para los cuales la probabilidad de observar los datos sea la más alta posible, donde k

Tabla 5: Partidos de Fútbol de Selecciones nacionales adultas 2005 - Oct 2013.

	Amistosos	Clasificatorias	Fase Final de Confederación	Torneo Copa del Mundo	Total
Con localía	2716	3079	189	10	5994
Cancha neutral	1170	282	485	118	2055
Total	3886	3361	674	128	8049

Tabla 6: Resultados modelos de regresión.

Variable	MR		MRLn		MRLnLc		MRLnFc	
	Coficiente	Error	Coficiente	Error	Coficiente	Error	Coficiente	Error
β_0	0,3071	0,0274	0,3019	0,0519	0,1478	0,0895	0,1478	0,0895
β_r	0,0131	0,0003	0,014	0,0003	0,014	0,0003	0,014	0,0003
β_{local}			0,3463	0,0614	0,5005	0,0954	0,3601	0,0954
β_{visita}			-0,5058	0,0645	-0,3517	0,0974	-0,467	0,0974
β_c					0,2049	0,0975		
β_{AFC}							0,0611	0,06
β_{OFC}							0,477	0,2012
β_{CNMBL}							0,3136	0,1021
β_{UEFA}							-0,1454	0,0631
β_{CAF}							0,1705	0,0633

es la cantidad de características de un equipo al momento del partido. Los resultados se muestran en la Tabla 6.

En los resultados del modelo MR, el signo positivo obtenido para el parámetro β_r indica que a un equipo con mejor ranking se le asocia una mayor probabilidad de ganar que el equipo adversario. Por otra parte, el signo y valor del intercepto (β_0) indica que ante partidos con equipos de similar ranking, es más probable que el resultado sea un triunfo para alguno de los equipos a que el resultado sea un empate. Adicionalmente, al comparar ambos estimadores se puede apreciar que las probabilidades de empatar o ganar de un equipo se igualan ($V_{i,k} = 0$) cuando el rival está $\frac{\beta_0}{\beta_r} \approx 23$ unidades arriba en el ranking.

Los resultados para modelo MRLn permiten inferir que las tres variables de este modelo son estadísticamente significativas (debido a que el intervalo de confianza de todos los parámetros a un 95 % de confianza no incluyen el valor 0). En cuanto a la variable β_{local} , se puede ver que su signo es positivo, en línea con los resultados anteriores en la literatura que han mostrado la importancia de este factor. De la misma forma, vemos que la variable β_{visita} toma un valor negativo, e incluso en módulo mayor que la variable β_{local} . Esto no es significativo si consideramos que en cada partido que no se juega en terreno neutral, hay un equipo local y uno visita. En este sentido, lo importante acá es

la magnitud de la diferencia entre estas variables. Al respecto, podemos usar dicha diferencia para estimar a cuánta diferencia de ranking equivale el factor localía. Esta relación se puede obtener resolviendo la siguiente ecuación:

$$\beta_{local} - \beta_{visita} = 2 * x * \beta_r \Rightarrow x = 30,4. \quad (8)$$

que se deduce igualando la ecuación de (6) para ambos rivales.

Este resultado indica que los partidos más parejos dentro del ámbito de selecciones adultas se producen cuando el equipo local tiene una fuerza de 30 unidades más baja en comparación al equipo visitante.

Comparamos los distintos modelos descritos arriba utilizando el *Criterio de Información Bayesiano* (BIC).² Nuestros resultados se presentan en la Tabla 7, donde se puede apreciar que el modelo MRLn obtiene un mejor desempeño (menor BIC) que el modelo MR.

Tabla 7: Resumen Resultados Modelos propuestos.

Modelo	Log Verosimilitud	N° Param.	AIC	BIC
MR	-7.704	2	15.412	15.426
MRLn	-7.409	4	14.828	14.856
MRLnLc	-7.407	5	14.825	14.860
MRLnFc	-7.391	9	14.800	14.862

Los resultados de la estimación del modelo MRLnLc, muestran que la variable β_c es una variable significativa y posee signo positivo, por lo cual un equipo que juega un partido en su continente tiene mayor probabilidad de ganar que su equipo contrincante, si este contrincante es de otro continente. Sin embargo, notamos que el efecto de localía continental es considerablemente menor que el efecto de localía nacional. Además, si bien la variable β_c es significativa, la Tabla 7 muestra que el aumento de log verosimilitud del modelo MRLnLc no compensa la agregación de un parámetro extra por sobre el modelo MRLn, por lo que el modelo MRLn continúa siendo el que mejor ajusta a los datos. Los resultados de la estimación del modelo MRLnFc muestran que la β_{AFC} no es significativa, es decir, estadísticamente no es distinta de 0 por lo tanto es igual a la variable omitida, que en este caso es $\beta_{CONCACAF}$, por lo cual países de ambas confederaciones se comportan de manera similar. Por otro lado países de la CONMEBOL, OFC y CAF presentan un mejor desempeño

²Este índice está dado por $BIC = -2 \cdot \ln(L) + k \cdot \ln(n)$, donde L es la verosimilitud asociada con el MLE, k el número de parámetros estimados por el modelo, y n el número de observaciones. El valor del BIC aumenta ya sea por la cantidad de variables o por la varianza no explicada de la variable dependiente, para más detalles ver, por ejemplo, [11].

que CONCACAF, mientras que países de la UEFA tienen peores resultados ³. El BIC de MRLnFc es mayor que el de los dos modelos anteriores.

En conclusión, el mejor modelo parece ser el MRLn, en cuanto a que es el que mejor explica los datos con el mínimo número de parámetros. Al ser la importancia del factor de localía una variable tan determinante en el resultado de un partido, esperaríamos entonces que fuera un factor tomado en cuenta al calcular el puntaje que un partido contribuye al rating de un equipo.

4. Propuestas de mejora y resultados

El ranking ideal construido en la Sección 2 lo utilizamos en este trabajo sólo como referencia. Consideramos que el ranking a ser implementado en la práctica por FIFA debe ser construido a través de una fórmula sencilla, tal cual se hace actualmente, de modo que pueda ser fácilmente interpretado por los aficionados al fútbol.

4.1. Propuestas de Mejora al Procedimiento de Ranking

A partir de las deficiencias develadas y los resultados obtenidos por los modelos en la sección anterior, procedemos a generar distintas propuestas de mejoras.

P1: Incluir localía y omitir confederación. Los resultados obtenidos en la sección anterior dan cuenta de la alta importancia de la variable localía nacional y de la relativamente baja importancia de la variable confederación. Intentando capturar esto proponemos eliminar la variable C de la fórmula de puntaje actual (3) y ajustar el factor de fuerza del rival de acuerdo a la prescripción dada por (8). Esto es, $P_{i,k} = M_{i,k} \cdot I_{i,k} \cdot T_{i,k}$, donde

$$T_{i,k} = \begin{cases} \max\{(186 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega de local} \\ \max\{(216 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega de visita} \\ \max\{(201 - R_{j_k,t_k}, 50)\} & \text{si } i \text{ juega en estadio neutral} \end{cases} \quad (9)$$

³Si bien algunos países pertenecientes a la UEFA han tenido grandes logros en torneos de mayor envergadura (como Alemania e Italia), la confederación tiene en total 55 países, entre los cuales se encuentran países que consistentemente han ocupado los últimos escalafones del ranking FIFA (tales como Malta, San Marino y Liechtenstein) lo cual ayuda a explicar los resultados

Esta fórmula impone una penalización respecto a la fórmula actual equivalente a 30 puestos del ranking, cuando un equipo juega de local.

P2: Omitir partidos amistosos. Como alternativa para mitigar lo perjudicial de los partidos amistosos, proponemos no considerar los partidos amistosos para efecto de cálculo de ratings. Esto elimina los incentivos a no jugar partidos amistosos (los que son claramente importantes, dado que permiten probar nuevos jugadores y experimentar nuevas estrategias). En términos concretos proponemos redefinir $G_{i,s}(t)$ en (4) como el conjunto de partidos *no amistosos* jugados por un equipo entre las fechas $t - s$ y $t - s + 1$, para todo s .

P3: Modificación en los puntos por victoria. Para mitigar las grandes diferencias en puntaje obtenido en un partido si se empata con los mejores del mundo en comparación a si se obtiene un triunfo ante equipos de mediano o bajo ranking, proponemos otorgar sólo dos puntos por victoria en vez de tres en la fórmula (9). De esta manera, un empate con equipos del top ten del ranking otorga una cantidad de puntos similar o incluso superior a derrotar a equipos de ranking mediocre. En términos concretos, proponemos utilizar

$$M_{i,k} = \begin{cases} 2 & \text{si } Y_{i,k} = \text{victoria} \\ 1 & \text{si } Y_{i,k} = \text{empate} \\ 0 & \sim \end{cases} \quad (10)$$

Notamos que esta propuesta se encuentra respaldada de forma parcial por los resultados en [3] los que muestran que bajo dicha asignación de puntajes, el ranking esperado al final del torneo será el mismo que el obtenido aplicando una variación al modelo probabilístico de [2], el cual deriva rankings mediante comparaciones de a pares (muy similar a regresiones logísticas)

P4: No depreciar puntos. Para corregir las deficiencias causadas por el desajuste de fechas en la calendarización de los torneos importantes de las distintas confederaciones, proponemos omitir la depreciación de puntaje que actualmente utiliza el ranking FIFA. En términos concretos proponemos utilizar

$$\alpha_s = \begin{cases} 1 & s \in \{1, 2, 3, 4\} \\ 0 & \sim \end{cases} \quad (11)$$

Al no depreciar los puntajes por año, el puntaje de un equipo en cualquier momento reflejará su desempeño en los últimos cuatro años, abarcando

todos los torneos de confederaciones posibles en los que un país puede participar.

4.2. Resultados

En nuestros experimentos, testeamos las propuestas de la sección anterior en forma incremental. Denotamos a la propuesta **P1-X** aquella que incluye las propuestas **P1** hasta la **PX**, inclusive. La Tabla 8 muestra los 16 primeros puestos de los rankings propuestos y el ranking FIFA, todos ellos en Octubre del año 2013 (fecha utilizada por la FIFA para sembrar los cabezas de serie del mundial más reciente).

Tabla 8: Rankings obtenidos mediante las cuatro propuestas analizadas más el ranking FIFA en Octubre de 2013.

	P1-1	P1-2	P1-3	P1-4	FIFA
1	España	Brasil	Brasil	Brasil	España
2	Argentina	España	España	España	Alemania
3	Alemania	Argentina	Argentina	Argentina	Argentina
4	Uruguay	Alemania	Alemania	Alemania	Colombia
5	Colombia	Estados Unidos	Chile	Uruguay	Bélgica
6	Costa de Marfil	Chile	Italia	Holanda	Uruguay
7	Italia	Holanda	Uruguay	Chile	Suiza
8	Estados Unidos	Uruguay	Estados Unidos	Costa de Marfil	Holanda
9	Brasil	Italia	Holanda	Italia	Italia
10	Bélgica	Costa de Marfil	Colombia	Japón	Inglaterra
11	Holanda	Colombia	Costa de Marfil	Estados Unidos	Brasil
12	Ghana	Bélgica	Bélgica	Inglaterra	Chile
13	Chile	Ghana	Inglaterra	Colombia	Estados Unidos
14	México	Japón	Ghana	Corea del Sur	Portugal
15	Suiza	México	Francia	Grecia	Grecia
16	Inglaterra	Francia	Japón	México	Bosnia

Es obviamente imposible aunar opiniones para dilucidar cual entre estas cuatro propuestas y la que actualmente usa la FIFA provee un *mejor* ranking. Una medida objetiva de comparación es la distancia de cada uno de estos rankings respecto al ranking ideal de referencia desarrollado en la Sección . Como medida de distancia utilizamos el error cuadrático medio (MSE) respecto al ranking de referencia:

$$MSE_j = \frac{1}{N} \sum_{i \in N} \left(R_i^j(oct, 2013) - R_i^*(oct, 2013) \right)^2, \quad (12)$$

donde MSE_j representa el error cuadrático medio asociado a la propuesta **P1-j** y $R_i^j(\cdot)$ denota el ranking del equipo i bajo la propuesta **P1-j**, (el superíndice

* denota el ranking de referencia), y N es la cantidad de países incluidos en el modelo. Mientras menor sea el MSE del ranking, mejor la propuesta.

La Tabla 9 muestra la comparación de las propuestas en términos de sus MSEs, los cuales revelan que la cuarta propuesta es la de mejor desempeño. Le sigue el ranking FIFA, con un error cuadrático medio 15.6% mayor que nuestra mejor propuesta. Las restantes propuestas muestran peor desempeño, especialmente la número 1, que se aleja en 41.9% del ajuste de la número 4. Concluimos entonces que, si bien el ranking FIFA presenta un aceptable desempeño, es posible mejorar dicho ranking utilizando el conjunto de nuestras propuestas.

Tabla 9: Desempeño del ranking FIFA y de las propuestas.

Propuesta	MSE
Ranking FIFA (Octubre 2013)	66,25
P1 (localía incluida y confederación omitida)	81,31
P1-2 (+ partidos amistosos omitidos)	72,87
P1-3 (+ modificación en los puntos por victoria)	69,87
P1-4 (+ sin depreciación de puntaje)	57,31

5. Formación de grupos en sorteo de la Copa del Mundo

El principal uso del ranking FIFA es determinar a los sembrados para el sorteo de la fase de grupos con la cual comienza la fase final de la Copa del Mundo. Es importante recordar que en esta etapa, los equipos sólo se enfrentan a equipos del mismo grupo. Analizamos a continuación el proceso de formación de estos grupos, y proponemos un esquema más balanceado mediante una formulación de programación entera.

5.1. Deficiencias del sistema actual de sorteo

Bajo el sistema actual, los grupos del Mundial a los cuales los países son sorteados se forman mediante una elección aleatoria. Los 32 equipos participantes en la Copa Mundial son divididos en 4 *pots*. El pot 1 contiene a los equipos que serán sembrados (o cabezas de serie) de los grupos: el país organizador y los siete de mejor ranking entre los restantes equipos. Los otros tres

pots contienen equipos de acuerdo a zonas geográficas. Para el Mundial Brasil 2014, el pot 2 contuvo a los equipos sudamericanos no sembrados y a los equipos africanos (7 en total), el pot 3 a los equipos norteamericanos y asiáticos (8 en total), y el pot 4 a los equipos europeos no sembrados (9 en total). Para que todos los pots contuvieran ocho equipos, uno de los equipos del pot 4 fue asignado mediante sorteo al pot 2. La Tabla 10 muestra la configuración de los pots resultantes.

Tabla 10: Pots de equipos y ranking FIFA en la antesala al Mundial 2014 (entre paréntesis el ranking relativo, asignándole a Brasil el ranking relativo 1 por ser el país anfitrión).

Pot 1	Pot 2	Pot 3	Pot 4
11 Brasil (1)	9 Italia (10)	13 USA (13)	8 Holanda (9)
1 España (2)	12 Chile (12)	24 México (23)	10 Inglaterra (11)
2 Alemania (3)	17 Costa de Marfil (17)	31 Costa Rica (24)	14 Portugal (14)
3 Argentina (4)	22 Ecuador (21)	34 Honduras (27)	15 Grecia (15)
4 Colombia (5)	23 Ghana (22)	44 Japón (28)	16 Bosnia (16)
5 Bélgica (6)	32 Argelia (25)	49 Irán (29)	18 Croacia (18)
6 Uruguay (7)	33 Nigeria (26)	56 Corea (30)	19 Rusia (19)
7 Suiza (8)	59 Camerún (32)	57 Australia (31)	21 Francia (20)

Luego, los grupos son formados sorteando la asignación de los equipos, de tal manera que cada uno de los ocho grupos contiene exactamente un equipo de cada pot. Existe una restricción adicional que impide que 3 equipos europeos y 2 equipos sudamericanos formen parte del mismo grupo. Si bien este sistema promueve la diversidad geográfica dentro de cada grupo, perjudica a algunos equipos, específicamente a aquellos con mejor ranking dentro de su propio pot que con alta probabilidad serán situados en grupos más difíciles. Por ejemplo, en 2014 Estados Unidos se vio altamente afectado con este sistema ya que compartió el pot 3 con equipos mucho más débiles, con los cuales estaba impedido de enfrentarse en la fase de grupos. Además Estados Unidos estaba obligado a enfrentar a un cabeza de serie y a otro equipo europeo del pot 4 (cuyos ranking varían entre 8 y 21). Otro país perjudicado fue Italia, el sorteado para pasar del pot 4 al 2, que en consecuencia quedó compartiendo pot con países de ranking muy inferior, como lo son Argelia, Nigeria y Camerún. La configuración final de grupos del Mundial 2014 se muestra en la Tabla 11. El número que antecede al país es su Ranking FIFA respectivo en el mes de Octubre del 2013, mes en el cual se determinaron los cabezas de serie. Entre paréntesis aparece el ranking relativo entre las 32 selecciones mundialistas. Brasil por ser el país anfitrión obtiene el primer lugar.

Como medida de dificultad de un grupo, utilizamos la suma de los rankings

Tabla 11: Cuadro Final Mundial Brasil 2014.

Grupo	Equipos			
A	11 Brasil (1)	18 Croacia (18)	24 México (23)	59 Camerún (32)
B	1 España (2)	8 Holanda (9)	12 Chile (12)	57 Australia (31)
C	4 Colombia (5)	15 Grecia (15)	17 Costa de Marfil (17)	44 Japón (28)
D	6 Uruguay (7)	31 Costa Rica (24)	10 Inglaterra (11)	9 Italia (10)
E	7 Suiza (8)	22 Ecuador (21)	21 Francia (20)	34 Honduras (27)
F	3 Argentina (4)	16 Bosnia (16)	49 Irán (29)	33 Nigeria (26)
G	2 Alemania (3)	14 Portugal (14)	23 Ghana (22)	13 USA (13)
H	5 Bélgica (6)	32 Argelia (25)	19 Rusia (19)	56 Corea (30)

de los equipos que lo conforman. Cuando menor es la suma, mayor es el nivel de dificultad. La Tabla 12 muestra esta medida para los grupos del Mundial 2014, revelando un gran desbalance entre los diferentes grupos. De acuerdo a la suma de rankings, los grupos B, D y G son los de mayor dificultad, mientras que los grupo H, A, E y F son los más débiles. Notar que Italia y Estados Unidos quedaron en los grupos más fuertes (Italia finalmente quedó eliminado en primera ronda, mientras que Estados Unidos clasificó segundo en su grupo).

Tabla 12: Suma de Ranking relativo y Ranking FIFA para los grupos del Mundial 2014.

Grupo	A	B	C	D	E	F	G	H
Suma Rank. Rel.	74	54	65	52	76	75	52	80
Suma Rank. FIFA	112	78	80	56	84	101	52	112

5.2. Modelo de Programación Entera para balancear los grupos

Nuestra propuesta es intentar corregir el desbalance en la dificultad de los distintos grupos, mediante un modelo de programación lineal entera. La formulación del modelo se presenta a continuación. En dicho modelo, la variable de decisión es en qué grupo se asigna cada país participante. Mas allá de las restricciones lógicas (todos los grupos tienen 4 países, los países son asignados a un solo grupo, etc), el modelo restringe el número de países de una misma confederación asignados a un grupo, y considera como función objetivo el minimizar la diferencia entre la máxima y mínima suma de rankings de los integrantes de un grupo.

5.2.1. Formulación matemática

Conjuntos

G : grupos.

C : confederaciones.

I : equipos.

S : equipos sembrados ($S \subset I$).

J_c : equipos de la confederación c ($J_c \subset I, c \in C$).

Parámetros

R_i : ranking del equipo i ($i \in I$).

M_c : máximo número de equipos de la confederación c que pueden pertenecer a un mismo grupo ($c \in C$).

Variables de Decisión

$$x_{ig} = \begin{cases} 1 & \text{si el equipo } i \text{ es asignado al grupo } g \\ 0 & \text{si no} \end{cases}$$

w_{min} : menor de las sumas de los rankings de los equipos dentro de un mismo grupo.

w_{max} : mayor de las sumas de los rankings de los equipos dentro de un mismo grupo.

Restricciones

$$\sum_{g \in G} x_{ig} = 1 \quad \forall i \in I \quad (13)$$

$$\sum_{i \in I} x_{ig} = 4 \quad \forall g \in G \quad (14)$$

$$\sum_{i \in S} x_{ig} = 1 \quad \forall g \in G \quad (15)$$

$$\sum_{i \in J_c} x_{ig} \leq M_c \quad \forall g \in G, c \in C \quad (16)$$

$$w_{min} \leq \sum_{i \in I} R_i x_{ig} \quad (17)$$

$$w_{max} \geq \sum_{i \in I} R_i x_{ig} \quad (18)$$

$$x_{ig} \in \{0,1\}, w_{min} \geq 0, w_{max} \geq 0 \quad \forall i \in I, g \in G \tag{19}$$

Función Objetivo

$$\text{mín } f = w_{max} - w_{min} \tag{20}$$

Las restricciones (13) aseguran que todo equipo sea asignado a exactamente un grupo. Las restricciones (14) establecen que cada grupo contiene a 4 equipos, mientras que las restricciones (15) aseguran que uno de estos equipos sea sembrado. Las restricciones (16) imponen la cota superior en el número de equipos de una misma confederación que pueden ser asignados a un mismo grupo (el Mundial 2014, se permitió a lo más un equipo de cada confederación, excepto la europea para la cual se permitieron dos). Las restricciones (17)-(18) ayudan a calcular las sumas mínima y máxima de los rankings de los equipos dentro de un mismo grupo, mientras que la función objetivo (20) busca minimizar la diferencia entre estas dos sumas.

5.3. Resultados

El modelo fue implementado en AMPL/CPLEX y su resolución tomó unos pocos segundos para todas las instancias consideradas. Las tablas 13 y 14 muestran grupos obtenidos utilizando el ranking FIFA y el ranking relativo, respectivamente.

Tabla 13: Grupos obtenidos por el modelo utilizando el ranking FIFA.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
Alemania	España	Suiza	Uruguay	Bélgica	Brasil	Colombia	Argentina
Chile	Irán	Ecuador	Honduras	Australia	México	Corea	Japón
Camerún	Costa de Marfil	Argelia	Ghana	USA	Nigeria	Italia	Costa Rica
Inglaterra	Bosnia	Francia	Rusia	Holanda	Grecia	Portugal	Croacia
Fuerza	83	83	82	83	83	83	83

Tabla 14: Grupos obtenidos por el modelo utilizando el ranking relativo.

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6	Grupo 7	Grupo 8
Brasil	Uruguay	Argentina	Suiza	España	Colombia	Bélgica	Alemania
Camerún	Corea	Australia	Chile	Irán	Costa Rica	Ecuador	Rusia
Italia	Francia	Bosnia	Japón	USA	Nigeria	Argelia	Honduras
México	Holanda	Grecia	Croacia	Ghana	Inglaterra	Portugal	Costa de Marfil
Fuerza	66	66	66	66	66	66	66

En ambos casos, el nivel de dificultad de los grupos obtenidos es considerablemente más parejo que los grupos obtenidos por el sorteo FIFA. Esto queda claramente reflejado en las figuras 1 y 2, que grafican la suma del ranking FIFA y relativo de los equipos en cada grupo, respectivamente. Usando el ranking FIFA, el modelo obtiene un valor objetivo óptimo $f^* = 1$, mientras que en los grupos sorteados en la realidad la diferencia entre los grupos de mayor y menor suma de rankings fue de 66. Usando el ranking relativo, el modelo obtiene un valor objetivo óptimo $f^* = 0$, mientras que en los grupos sorteados en la realidad la diferencia entre los grupos de mayor y menor suma de rankings fue de 28.

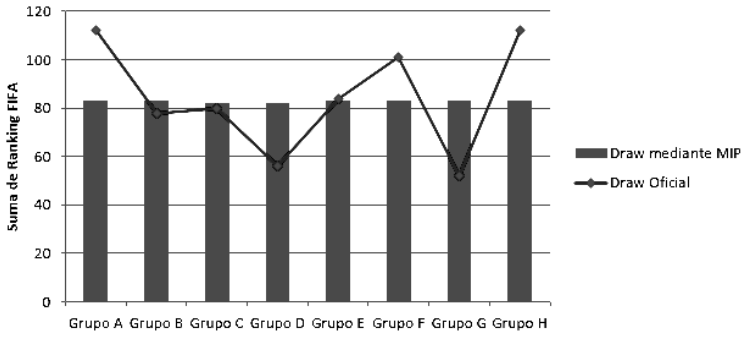


Figura 1: Suma del ranking de los equipos en cada grupo obtenidos por el modelo usando el ranking FIFA.

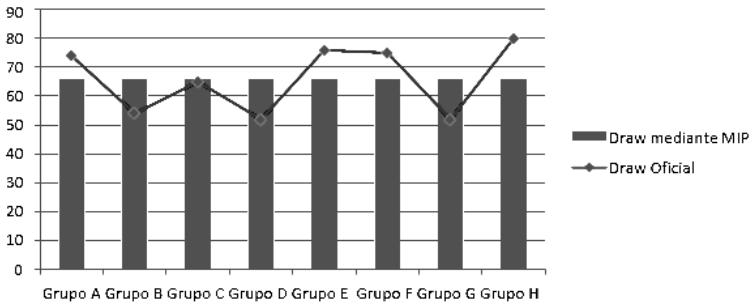


Figura 2: Suma del ranking de los equipos en cada grupo obtenidos por el modelo usando el ranking relativo.

6. Conclusiones

Este artículo ha revelado falencias importantes en la metodología actual del ranking FIFA, y genera a su vez, un conjunto de propuestas de mejoras basadas en los resultados de modelos matemáticos desarrollados para este estudio y aplicados a datos de partidos internacionales. Primero, nuestro estudio empírico sugiere que el puntaje entregado en un partido no refleja de buena manera cuán meritorio fue para cada equipo obtenerlo. Si bien la metodología de FIFA entrega mayor puntaje por jugar contra equipos con mejor ranking, no hace referencia alguna a otras variables tan relevantes, como lo es por ejemplo el factor localía. Por otro lado, el ranking actual permite a países sacar ventaja evitando jugar partidos amistosos en el año previo a la designación de los cabezas de series del Mundial. Esto debido principalmente a la escasa entrega de puntos de este tipo de partidos en comparación con cualquier otro tipo de formato, combinado con la fórmula de promediar los puntajes de los juegos disputados. Nuestro trabajo empírico muestra como este factor favoreció a países como Bélgica, Suiza y Colombia para ser sembrados en el Mundial 2014. Otro factor que estudiamos es el de la depreciación del promedio de puntos de los últimos 4 años. Este sistema genera disparidad entre países de distintos continentes debido a que no todos juegan su torneo continental en el mismo año, por ende, ante igualdad de resultados en los demás torneos, es probable que los equipos de confederaciones que jueguen sus torneos más cerca del próximo mundial se vean beneficiados.

Mediante modelos de regresión logística multinomial, y el uso intensivo de datos, analizamos la importancia de distintas variables en el resultado de partidos de fútbol entre selecciones. Los resultados muestran lo fundamental que es la variable localía, ya que entrega una gran ventaja al equipo que posea dicho atributo. Este factor acompañado con la diferencia de Ranking entre equipos fueron los de mejor ajuste a la data histórica. De nuestros resultados, el efecto de jugar de local es comparable con tener una diferencia de ranking de 30 puestos a favor. Otros factores, como la pertenencia a una determinada confederación y la localía continental tienen mucho menor poder predictivo.

Con el fin de mejorar el ranking FIFA actual, generamos propuestas que modifican su metodología. Estas modificaciones consisten en incluir la variable localía, eliminar la variable confederación, omitir partidos amistosos, reestructurar los puntos por partido ganado y cambiar el sistema de depreciación de puntaje por año. Para testear el desempeño de estas modificaciones y com-

pararlas con la metodología actual, utilizamos el criterio de error cuadrático medio y un ranking referencial creado en base a simulaciones. Los resultados muestran que la incorporación de estas modificaciones en conjunto mejoran considerablemente el desempeño del ranking FIFA, cuyo error cuadrático medio fue 15,6% mayor que el de nuestra mejor propuesta.

Finalmente, mediante un modelo de programación lineal entera mixta, logramos generar grupos que mantienen los criterios geográficos de la FIFA y a la vez son más parejos en nivel de dificultad. Aplicando el modelo a los datos del Mundial 2014 y utilizando como medida de dificultad de un grupo la suma de los rankings de los equipos que lo componen, logramos formar grupos mucho más balanceados que los grupos sorteados por la FIFA.

Referencias

- [1] S. Anderson, A. De Palma, y J. Thisse. *Discrete choice theory of product differentiation*. MIT press, 1992.
- [2] R. Bradley y M. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.
- [3] R. Davidson. On extending the bradley-terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328, 1970.
- [4] M. Dixon y M. Robinson. A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(3):523–538, 1998.
- [5] P. Downward y M. Jones. Effects of crowd size on referee decisions: Analysis of the fa cup. *Journal of sports sciences*, 25(14):1541–1545, 2007.
- [6] G. Durán, M. Guajardo, y D. Sauré. Scheduling the south american qualifiers to the 2018 fifa world cup by integer programming. *European Journal of Operational Research*, 262(3):1109–1115, 2017.
- [7] D. Dyte, S. Clarke, et al. A ratings based poisson model for world cup soccer simulation. *Journal of the Operational Research society*, 51(8):993–998, 2000.
- [8] FIFA. 2014 fifa world cup brazil technical report and statistics. Technical report, 2014.

- [9] FIFA. 2014 fifa world cup brazil television audience report. Technical report, 2014.
- [10] FIFA. Fifa ranking, marzo 2017.
- [11] J. Friedman, T. Hastie, y R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.
- [12] J. Guyon. Rethinking the fifa world cup final draw. *Journal of Quantitative Analysis in Sports*, 11(3):169–182, 2015.
- [13] J. Lasek, Z. Szlávik, y S. Bhulai. The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, 1(1):27–46, 2013.
- [14] M. Maher. Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118, 1982.
- [15] I. McHale, S. Davies, et al. Statistical analysis of the effectiveness of the fifa world rankings. *Statistical thinking in sports*, páginas 77–90, 2007.
- [16] R. Pollard. Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, 4(3):237–248, 1986.
- [17] H. Rue y O. Salvesen. Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3):399–418, 2000.
- [18] P. Scarf y M. Yusof. A numerical study of tournament structure and seeding policy for the soccer world cup finals. *Statistica Neerlandica*, 65(1):43–57, 2011.
- [19] A. Suzuki, L. Salasar, J. Leite, y F. Louzada-Neto. A bayesian approach for predicting match outcomes: the 2006 (association) football world cup. *Journal of the Operational Research Society*, 61(10):1530–1539, 2010.

POLÍTICAS DE DESINCENTIVO EN EL CONSUMO DE ALCOHOL Y EFECTO EN LOS ACCIDENTES DE TRÁNSITO

VALERIA SCAPINI *
CINTHYA VERGARA ** ***
DIEGO ÁLVAREZ *

Resumen

Los accidentes de tránsito generan altos costos económicos y sociales dentro de un país. El alcohol se identifica como uno de los factores de riesgo clave dentro de este tipo de accidentes. En Chile, para disminuir el riesgo de accidentes provocados por conductores bajo los efectos del alcohol, se aprobaron la Ley N° 20.580, llamada “Tolerancia Cero” y la Ley N° 20.770, llamada “Emilia”. El presente estudio tiene la finalidad de estudiar el impacto que ha tenido sobre la cantidad de accidentes la variación exógena generada por la aprobación de estas leyes, a través de la estimación de 3 modelos lineales. Se recopilieron datos estadísticos del Ministerio de Salud, entre los años 2008 y 2015, que entregan la cantidad de atenciones de urgencias de accidentes de tránsito, junto con la zona geográfica, la cantidad de automóviles en circulación, el ingreso medio de los hogares, nombre de los consultorio de salud, año y semanas estadísticas. Como resultado se obtiene que, en general, ambas leyes se relacionan con una disminución en la cantidad de accidentes de tránsito, sin embargo, no todas las zonas presentan esta correlación. Tanto el ingreso de los hogares como el número de automóviles contribuyen al número de accidentes en magnitudes que varían cada recinto de atención.

Palabras Clave: Consumo de Alcohol, Accidentes de Tránsito, Legislación, Modelos Lineales, Modelos Cuantitativos.

* Escuela de Ingeniería Comercial, Universidad de Valparaíso, Santiago, Chile

** Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

*** Planificación y Control de Gestión, CRS Hospital Provincia Cordillera, Chile

1. Introducción

Los accidentes de tránsito son un serio problema que aqueja a la sociedad día tras día trayendo consigo sufrimiento, muertes y daños colaterales con altos costos sociales y económicos. En Chile, entre los años 1997 y 2012, los accidentes de tránsito corresponden a la doceava causa de muerte, superando a causas como la insuficiencia cardíaca y enfermedades vinculadas al corazón, varios tipos de tumores, homicidios o Virus de la Inmunodeficiencia Humana (VIH). Además, producen una serie de externalidades negativas a través de daños a la propiedad, demoras en las vías, uso de recursos como carabineros, servicios de salud y bomberos.

Con ciudades cada vez más complejas y parques automotores cada vez más grandes, distintos países se han visto en la necesidad de tomar medidas de distintas índole para mejorar tanto las condiciones de tránsito como la seguridad de calles y caminos, los vehículos y los propios usuarios de la vía pública. La Organización Mundial de la Salud (OMS) identifica como factores de riesgo clave la velocidad excesiva, conducción bajo los efectos del alcohol, no uso de casco, cinturón de seguridad y dispositivo de retención para niños y distracciones debido a que generan mayor probabilidad de causar un accidente al conducir un vehículo. De acuerdo con cifras de la OMS cada año mueren 1.25 millones de personas en todo el mundo debido a accidentes de tránsito donde, la mitad de ellas, son “usuarios vulnerables de la vía pública” como peatones, ciclistas y motociclistas siendo además la principal causa de muerte entre grupo etario de 15 a 29 años de edad [15, 16].

En particular, dentro de los factores que aumentan el riesgo de accidentes vehiculares, la influencia del alcohol, en conductores o peatones, provoca alrededor del 20% de las muertes en accidentes de tránsito [7, 19]. El alcohol es conocido por ser una sustancia que afecta al individuo que lo consume en diferentes estadios o niveles de conciencia. En primer lugar afecta el nivel psicomotor, disminuyendo la capacidad de juzgar distancias, velocidad o precisión para seguir una trayectoria; en segundo lugar el nivel psicológico, aumentando el grado de euforia y agresividad, la concentración se ve disminuida, y además se produce una sensación de aversión al riesgo; y en tercer lugar el nivel físico, dificultando la capacidad de seguir objetos con la vista, o mantener su vista fija en un punto en específico [1].

Dado a estas consideraciones, en Chile se han aprobado dos modificaciones a la Ley del Tránsito, que tienen el objetivo de desincentivar el consumo de

alcohol mediante sanciones más severas, combatir la accidentabilidad y mortalidad en siniestros de tránsito. Estas modificaciones reciben el nombre de Ley Tolerancia Cero o Ley N° 20.580 y Ley Emilia o Ley N° 20.770; ambas corresponden variaciones exógenas debido al cambio en la tolerancia de la ciudadanía que buscan evitar accidentes con muertes producto de la ingesta de alcohol.

El presente estudio tiene la finalidad de estudiar cuál ha sido el impacto que ha tenido sobre la cantidad de accidentes la implementación de las leyes “Tolerancia Cero” y “Emilia” en la Región Metropolitana, tomando como fecha de estudio el periodo comprendido entre los años 2008 y 2015. Con el fin de cumplir éste objetivo se recopilaron datos estadísticos del Ministerio de Salud, que entregan la cantidad de atenciones de urgencias de accidentes de tránsito, que en conjunto con la zona geográfica, la cantidad de automóviles en circulación y el ingreso medio de los hogares, se estiman cuatro modelos de regresión lineal que explican de manera íntegra el comportamiento de estas variables.

2. Trabajos relacionados

Actualmente se han realizado estudios acerca de los factores que tienen un efecto en la ocurrencia de accidentes de tránsito. Fridstrøm [4, 5, 6] realizó un estudio para determinar la importancia que tiene la econometría sobre el estudio de los accidentes de tránsito, dentro del cual describe tipos de regresiones que son apropiados para medir la accidentabilidad. En [5] utiliza el modelo econométrico llamado TRULS para explicar la accidentabilidad mensual dentro en 10 provincias de Noruega durante una extensión de tiempo de 22 años. Este modelo tiene la característica de no solo explicar el número de accidentes, víctimas por categoría y nivel de gravedad, sino que además cuenta con ecuaciones que entregan datos como tasas de posesión de vehículos privados, la cantidad de tránsito explicados en términos de vehículos por kilómetro, tasas de uso de cinturón de seguridad, entre otros. También explica que los accidentes de tránsito son imprevisibles, pero evitables, y se debe trabajar considerando la causalidad y la aleatoriedad de este tipo de problemas.

Posteriormente Graham y Glaister [8] miden la incidencia que tienen en los accidentes de tránsito la escala urbana, la densidad de población, y el tipo de lugar en consideración. Para realizarlo utilizaron censos ingleses como unidad de estudio y desarrollaron modelos binomiales negativos para llevar a cabo el análisis. Se concluye que la incidencia de víctimas de peatones y heridos de diversa gravedad es más alta en zonas residenciales que en áreas de negocio,

mientras que la relación entre urbana y víctimas de peatones se ve disminuida al examinar varios muy poblados.

Griebe [9] describe los hallazgos principales acerca de los estudios realizados por el mismo autor sobre modelos de predicción de accidentes en juntas urbanas y enlaces urbanos de carreteras. El autor busca proponer modelos de accidentes simples para predecir la cantidad esperada de accidentes de tránsito en cruces urbanos y enlaces viales, e identificar factores que afectan a la seguridad vial. En su desarrollo dieron cuenta que los modelos para enlaces de carretera resultan ser menos complicados que las uniones, dado a que poseen un patrón de accidentabilidad más uniforme dado un flujo de tráfico más simple. Las variables explicativas resultaron ser más significativas para los modelos de carreteras, el diseño de las carreteras y la geometría del camino, mientras que menos significativas al momento de analizar uniones o empalmes. La variable más importante en ambos resultados fue el flujo de tráfico vehicular.

Noland y Quddus [13] realizaron una investigación en donde utilizan como datos los distritos de enumeración de Londres con la finalidad de investigar como la congestión vehicular puede afectar la seguridad vial. En primer lugar, se plantean una hipótesis que propone que las condiciones de tráfico congestionado pueden aumentar el número de choques e interacciones entre vehículos, pero su gravedad normalmente es inferior en comparación a rutas de flujo libre no congestionadas. Su análisis utilizó modelos de conteo negativo binomial para determinar los factores que afectan a las víctimas. En sus diferentes estadios de gravedad, difieren entre periodos de tiempo congestionados en comparación con periodos de tiempo sin existencia de congestión vehicular. Los resultados sugieren que los efectos de la infraestructura pueden interactuar con niveles de congestión, pero aun así resultan ser no concluyentes a la hora de considerar a la congestión como mayor causa atenuante.

Utilizando datos de lesiones y accidentes fatales en Pensilvania entre los años 1969 y 2000, Agüero-Valverde y Jovanis [2] realizaron un modelo bayesiano que incluye efectos espaciales y temporales e interacciones espacio tiempo, que se compara con la estimación binomial negativa tradicional de la frecuencia anual de accidentes a nivel de condado. Las variables ocupadas incluyen situación socio-demográfica, condiciones meteorológicas, infraestructura de transporte y cantidad de viajes. Los resultados muestran que los condados con mayor nivel de pobreza, aquellos con mayor kilometraje de carretera y densidad de carretera han visto aumentado significativamente el riesgo de accidentes; la precipitación total es significativa y positiva en modelos binomiales negativos, pero no es significativa mediante estimación bayesiana. La correlación espacial, la tendencia temporal y las interacciones espacio tiempo son significativas en los modelos de accidentes con lesiones mediante estima-

ción bayesiana. Se evidencia la existencia de correlación espacial en los datos de choque y entregan un mecanismo para cuantificar y reducir el efecto de esta correlación. Existe una alta probabilidad de que la correlación espacial sea aún más importante en los segmentos de carretera donde hay intersecciones.

Wedagama y Dissanayake [21] realizaron un estudio que investiga la influencia de factores relacionados con accidentes que tienen víctimas mortales en carretera utilizando la técnica de relación logística. Los modelos de regresión logística se desarrollaron por separado, considerando motocicletas y todos los vehículos, situándose en Bali de Indonesia como caso de estudio. En base a los datos de accidentes de la Policía Estatal de Bali, se emplearon siete variables para explicar el modelo. El estudio evidenció que las posibilidades de un accidente mortal debido a los motociclistas masculinos y los automovilistas fueron de 0.3 y 0.4 respectivamente inferior a las mujeres. Por lo tanto, las posibilidades de las motociclistas y automovilistas eran aproximadamente del 79 % y 72 % respectivamente, por lo que contribuyen más a los accidentes mortales que los hombres. Otro factor considerado para el análisis es la edad, que fue significativa para influir en todas las muertes por accidentes de tránsito, obteniendo un impacto del 50 % sobre éstas.

Haque, Chin y Lim [10] evaluaron cómo los factores de comportamiento del conductor influyen en el riesgo de accidente e identificaron el grupo más vulnerable de conductores. Para explorar el comportamiento del piloto, desarrollaron un cuestionario con 61 ítems de búsqueda de sensaciones impulsivas, agresivas y comportamiento que podría contribuir algún riesgo. Agruparon el riesgo de colisión mediante el algoritmo de k-medias y desarrollaron un modelo logarítmico lineal que relaciona el comportamiento del conductor con el riesgo de colisión. Los resultados muestran que los motociclistas involucrados en el accidente tienen una puntuación más alta en rasgos conductuales que evidencien peligros. Los conductores agresivos son más propensos a caer bajo el grupo de alto riesgo, mientras que la búsqueda de sensaciones impulsivas no es significativa. Estos datos contribuyen a realizar las campañas de seguridad vial.

Con el objetivo de destacar los factores que influyen en la ocurrencia de accidentes en lugares peligrosos de caminos urbanos, Obaidat y Ramadan [14] realizaron un estudio en donde correlacionaron las características de accidentes con diferentes factores, tales como, elementos geométricos del camino, velocidad de tráfico, tipo de pavimento, condiciones de iluminación, tipo de colisiones y condiciones del tráfico. El estudio considero 28 zonas peligrosas en las carreteras de Amman, Jordania. Casa uno de los lugares peligrosos escogidos tenía al menos 20 accidentes y dos registros de fatalidad por año. Se desarrollaron diferentes modelos para correlacionar las características del accidente

con las variables estudiadas. Se encontró que los modelos logarítmico y lineal fueron los modelos más significativos y parsimoniosos que se pueden utilizar para predecir la relación entre las características de los accidentes como variable dependiente y las otras variables independientes. Las variables que más contribuyeron a los accidentes de tránsito en lugares peligrosos fueron velocidad media, velocidad de desplazamiento, grado máximo y medio de curvas horizontales, número de curvas verticales, anchura media, tipo de superficie de carretera, iluminación, número de vehículos por hora, número de instalaciones de cruce de peatones y porcentaje de camiones en la ruta. La finalidad de este estudio es superar los problemas asociados con accidentes de tránsito en lugares peligrosos y mejorar la seguridad.

El estudio realizado por Seva, Flores y Gotohio [20] en el año 2013 tiene busca determinar las variables personales y ambientales significativas en la predicción de accidentes de motocicletas en Filipinas, comparar los resultados con hallazgos en otros países y proponer intervenciones gubernamentales para mejoras. Los datos fueron recolectados de 177 participantes, quienes a través de una encuesta que fue contestada en un centro de licencia en la ciudad de Metro Manila. Se utilizó regresión logística para predecir la probabilidad de un accidente con las variables consideradas para el modelo. Se obtuvo que tres variables eran predictores significativos en los accidentes en motocicleta: edad, conducción y tipo de unión. Los conductores jóvenes resultaron ser más propensos a estar involucrados en accidentes, el hecho de cometer una violación a las leyes del tránsito aumenta la probabilidad de ocurrencia de un accidente, mientras que la conducción en las uniones también aumentan significativamente el número de accidentes. Estas variables son de ayuda sobre campañas gubernamentales.

Por último, Anowar y Raihan [3] analizaron los patrones de accidentes en intersecciones seleccionadas en una arteria urbana en la ciudad de Daca, Bangladés. Los accidentes de intersección representan cerca del 40% del total de accidentes que ocurren en la ciudad metropolitana de Daca. El estudio intentó identificar las principales causas, factores y tipos de accidentes en las intersecciones problemáticas, y sugerir medidas adecuadas para reducir los accidentes. El estudio investigó 2 tipos de intersecciones: Intersecciones urbanas y suburbanas con 4 tipos de patrones de accidentes (colisión fatal, grave, simple y justa).

Los estudios revisados muestran distintos factores asociados a accidentes permitiendo comprender su comportamiento y contribuyendo al diseño y cuantificación de distintas medidas o políticas públicas ligadas a mejores condiciones viales.

3. Ley Tolerancia Cero y Ley Emilia

En distintos países las tasas de accidentes de tránsito asociados al alcohol han sido reducidas disminuyendo el límite legal aceptable. La Ley ° 20.580 conocida como “Ley Tolerancia Cero” entró en vigencia en el año 2012 como modificación de la Ley de Tránsito bajando los grados de alcohol permitidos en la sangre para conducir de acuerdo al objetivo de sancionar con mayor severidad la conducción de vehículos por personas que se encuentren en estado de ebriedad cuando dicho delito provocare lesiones graves, gravísimas o la muerte de la víctima [11].

Partiendo de este objetivo se reformuló los gramos de alcohol en la sangre definiendo como “*bajo la influencia del alcohol*” a aquellos conductores con 0,3 a 0, 8 gramos por litro de sangre y en “*estado de ebriedad*” a los conductores con más de 0,8 gramos de alcohol por litro de sangre.

La normativa legal llamada “Ley Emilia” que hace referencia a la Ley N° 20.770 [12] toma ese nombre debido a una petición ciudadana que frente a muerte de la menor fallecida Emilia Silva Figueroa solicitaba penas más fuertes para personas que condujesen bajo la influencia de alcohol. Entró en vigencia el año 2014 y en ella se sanciona con cárcel efectiva de al menos un año a los conductores en “*estado de ebriedad*” que generen lesiones graves, gravísimas o la muerte con cárcel efectiva de al menos 1 año. Además, establece como delito fugarse del lugar del accidente y negarse a realizar el alcoholtest o la alcoholemia. En caso de que un conductor sea encontrado con alcohol en la sangre, pero sin que haya causado lesiones gravísimas o la muerte, no se aplica la “Ley Emilia”, sino la “Ley Tolerancia Cero”

4. Construcción de modelos

La finalidad de este estudio es medir la efectividad de leyes “Tolerancia Cero” y “Emilia” sobre la reducción de la cantidad de accidentes de tránsito, es decir, el efecto causal que existe entre la aplicación de las leyes que buscan desincentivar el consumo de alcohol y el número de accidentes de tránsito, entendiéndose como variación exógena a la aprobación de ambas leyes. Si bien varios estudios han realizado análisis para medir la ocurrencia de accidentes considerando sus factores, no existe estudio a la fecha que logre plasmar empíricamente la efectividad de estas leyes en Chile.

Para ello, se utilizará el método de evaluación pre-post que está basado en la evaluación de las diferencias que muestra un sólo un grupo, el que fue afectado por un cambio exógeno o un shock externo (cambio en la legislación vigente), midiendo los efectos de este cambio en función a los datos obtenidos antes y después del suceso. En palabras de Pomeranz [17] *“una evaluación pre-post mide el cambio en el tiempo tomando en cuenta el estado inicial del grupo. En este caso, se mide el impacto como la diferencia entre la situación anterior y la situación posterior a una intervención. El análisis pre-post es una manera muy común de evaluar programas. Muchas veces este tipo de análisis retrospectivo parece conveniente si los datos de la situación anterior al programa existen”*.

El método de diferencias pre-post toma como supuestos que el grupo de comparación está compuesto por los mismos participantes y que esta intervención es el único factor que influye en cambio del resultado, por lo que sin esta intervención el resultado o la condición de la población se habría mantenido sin cambio aparente. Las ventajas de este procedimiento están dadas por la existencia de datos que se pueden analizar de manera retrospectiva, sin necesidad de requerir datos de personas que no son partícipes del cambio, mientras que las desventajas están dadas por la falta de consideración de otros factores que, si pueden inferir en el cambio de situación en la población, esto puede afectar al supuesto clave de que la intervención es el único factor que influye en el cambio. Este método de evaluación supone que no existe una tendencia en los datos, sin embargo el análisis del número de accidentes de tránsito muestra que éstos han aumentado en el tiempo. Es importante señalar que esta tendencia no se contradice con el objetivo del estudio planteado que busca determinar una disminución en el número de accidentes de tránsito. A continuación se muestra un gráfico con la tendencia de éstos en el tiempo.

La estimación de los efectos tomando datos antes y después de la aprobación de las leyes se lleva a cabo mediante Regresiones Lineales con Mínimos Cuadrados Ordinarios (MCO) y variables de control que permiten mejorar la estimación del efecto de las variables bajo estudio.

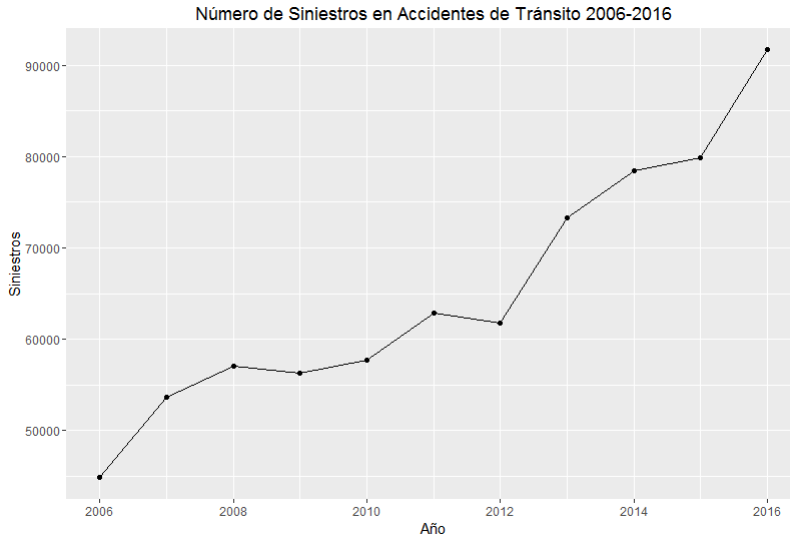


Figura 1: Número de siniestros de accidentes de tránsito en Chile 2006-2016

4.1. Datos recolectados

Para estudiar el efecto que tienen la “Ley Tolerancia Cero” y “Ley Emilia” en los accidentes de tránsito se utilizó una serie de datos desde el año 2008 al 2015 extraída del Departamento de Estadística e Información de Salud (DEIS) respecto a la cantidad de atenciones de urgencias semanales por causas de accidentes de tránsito tomando como supuesto que tiene una alta correlación con el número de accidentes de tránsito. Específicamente se utilizaron estadísticas de los hospitales y SAPU pertenecientes a la Región Metropolitana, que contiene información semanal a nivel de zona y comuna, y registra información completa para el período estudiado.

Para obtener los datos de vehículos en circulación se utilizaron los anuarios parque de vehículos en circulación obtenidos del Instituto Nacional de Estadísticas (INE). Específicamente se utilizaron los datos entre los años 2008 y 2015 del total de vehículos de la Región Metropolitana, que fueron ingresados para cada comuna, bajo el supuesto que el número de vehículos se mantiene fijo durante el año. Esta base de datos contiene los permisos de circulación otorgados por todas las municipalidades según tipo, según región, provincia y comuna, y registra información completa para el periodo estudiado.

Para la variable ingreso se utilizó el indicador de “Ingreso medio” de los hogares obtenido desde Observatorio Urbano del Ministerio de Vivienda y Urbanismo. Esta base de datos contiene el ingreso medio de cada comuna del país, la cual para ser utilizada fue corregida en base al deflactor obtenido del Índice de Precios al Consumidor (IPC) entregado por el Banco Central, expresándose

los valores en cifras del año 2009. Las variables comuna y zona son obtenidas desde la base de datos del Departamento de Estadísticas e Información de Salud (DEIS).

La base de datos consolidada contiene 43.190 datos y las variables de accidentes de tránsito, ingreso y vehículos, debido a la gran dispersión que presentan sus valores, se transformaron aplicando logaritmo natural. Para estimar los modelos se utilizó el lenguaje y ambiente para análisis estadístico R-cran y RStudio [18].

4.2. Evaluación de las Leyes mediante modelos de regresión

Los modelos de regresión lineal consideran como variable dependiente o explicada la cantidad de atenciones de urgencia por accidentes de tránsito, mientras que las variables independientes consideradas fueron: $ley_{tolerancia0}$, ley_{Emilia} (variable binaria que toma valor 1 con la aplicación de cada una de las leyes y valor 0 de lo contrario).

Los resultados obtenidos en esta estimación se presentan en la Tabla 1

Tabla 1: Estimación Modelo 1

Variable	Estimación
Ley Tolerancia Cero	-0.0563 (*)
Ley Emilia	-0.1249 (***)
R-Cuadrado	0.716
Efecto Fijo Consultorio	Sí
Efecto Fijo Año/Semana	Sí
Número de observaciones	43.190
Signif. codes: 0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 () 1	

Posteriormente se agregaron las variables: logaritmo natural del número de vehículos y logaritmo natural del ingreso medio de los hogares por comuna.

Obteniendo como resultados la estimación presentada en la Tabla 2.

Tabla 2: Estimación Modelo 2

Variable	Estimación
Ley Tolerancia Cero	0.569520
Ley Emilia	0.049903 (*)
ln(vehículos)	0.000184 (***)
ln(ingreso)	0.013714 (*)
R-Cuadrado	0.718
Efecto Fijo Consultorio	Sí
Efecto Fijo Año/Semana	Sí
Número de observaciones	43190
Signif. codes: 0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 () 1	

Por último, el tercer modelo corresponde al anterior que fue aplicado para cada zonas geográfica en particular.

Los resultados obtenidos de la estimación se presentan en la Tabla 4.2.

Tabla 3: Estimación Modelo por Zona Geográfica

Variable	Centro	Norte	Occidente	Oriente	Sur	Sur Oriente
Ley Tolerancia Cero	0.0630	-0.1179	-0.08227	0.3007 (**)	0.0861	-0.3619(**)
Ley Emilia	-0.4583 (***)	0.004416	-0.0721	0.0218	0.02728	-0.2283(*)
ln(vehículos)	-0.3855 (.)	1.21 (***)	1.9117 (***)	0.2604	1.376 (***)	-0.948 (***)
ln(ingreso)	-15.4022	30.70	-96.3383 (***)	5.0095	-14.23	10.74
R-cuadrado	0.9229	0.6131	0.6109	0.7834	0.7375	0.6253
E.F. Control	Sí	Sí	Sí	Sí	Sí	Sí
E.F. Años/Semana	Sí	Sí	Sí	Sí	Sí	Sí
Número de observaciones	1158	5492	2489	2305	2530	2805
Signif. codes: 0 (***) 0.001 (**) 0.01 (*) 0.05 (.) 0.1 () 1						

4.3. Discusión de resultados

El primer modelo busca explicar de manera general si las leyes “Tolerancia Cero” y “Emilia” han sido efectivas en la Región Metropolitana. Al respecto se puede decir que ambas leyes se relacionan con una disminución en el número de accidentes. Sin embargo, la disminución del número de accidentes es mayor con la aplicación de la “Ley Emilia”.

El segundo modelo mide la efectividad de ambas leyes considerando la cantidad de vehículos en circulación por comuna y el ingreso medio del hogar por comuna de la Región Metropolitana. Los resultados muestran que sólo tres de las cuatro variables consideradas resultan ser significativas, a saber: la “Ley Emilia”, el nivel de ingreso y los vehículos, y que se relacionan con un aumento en el promedio de accidentes. Específicamente, en el caso de la cantidad de vehículos en circulación, se puede decir que a mayor cantidad de vehículos motorizados particulares existen se tiene mayor congestión vehicular y aumenta la probabilidad de ocurrencia de accidentes de tránsito. Del mismo modo, el nivel de ingreso tiene una correlación positiva con el número de accidentes, ya que el mayor ingreso percibido por los hogares se relaciona con un mayor consumo, adquisición y uso de vehículos motorizados particulares, lo cual aumenta la probabilidad de ocurrencia de accidentes de tránsito. Dado que el parque automotriz ha aumentado en el tiempo, esta situación podría verse empeorada, por lo que deben implementarse mejoras o incentivos para la utilización de transportes de carácter colectivo que logren disminuir la cantidad de flujo vehicular. Llama la atención el hecho de que la “Ley Emilia” se relacione con un aumento en el promedio de accidentes, sin embargo, esto se puede deber a la tendencia creciente que existe en el número de accidentes en el tiempo.

El tercer modelo es una recopilación de todas las variables anteriores considerando la aplicación de las leyes, el número de vehículos en circulación, el ingreso medio del hogar para cada zona en la Región Metropolitana. Estas zonas corresponden a una división de la región que permite agrupar a los diferentes centros de salud dentro localizados en un área geográfica, con el fin de determinar las zonas en las cuales la ley ha sido más efectiva. Los resultados varían considerablemente dependiendo de la zona geográfica. En el caso de la “Ley Tolerancia Cero” su efecto es estadísticamente significativo en las zonas *Oriente* y *Sur Oriente* con el efecto esperado en cuanto a una disminución en la cantidad de accidentes sólo para esta última. Para el caso de la “Ley Emilia” su efecto es estadísticamente significativo en las zonas *Centro* y *Sur Oriente*, con el efecto esperado en cuanto a una disminución en la cantidad de accidentes. Las variables de ingreso y cantidad de vehículos muestran resul-

tados que difieren según sea la zona en estudio, encontrando zonas en donde los accidentes aumentan y otras en donde disminuyen. La única zona que ha presentado una disminución en el promedio de accidentes ante la aplicación de ambas leyes corresponde a la zona *Sur Oriente*.

Las causas de que en las zonas existan diferencias pueden estar dadas por la cantidad de vehículos existentes en la zona, el estado de las calles, la frecuencia o presencia de fiscalización, la existencia de autopistas, tamaño e infraestructura de la zona geográfica, entre otras. Estas diferencias podrían causar que en algunas zonas exista mayor riesgo al conducir por factores inherentes de la zona que no están relacionados con el consumo de alcohol.

Es necesario señalar que, de manera general, ambas leyes se relacionan con una disminución en el valor esperado de accidentes, sin embargo, al controlar por nivel de ingreso y cantidad de vehículos, estos resultados se distorsionan, sólo se mantienen en la zona *Sur Oriente* de la Región Metropolitana.

5. Conclusiones

Este trabajo tuvo como finalidad identificar cómo han impactado las leyes “Tolerancia Cero” y “Ley Emilia” a la cantidad de accidentes ocurridos en la Región Metropolitana de acuerdo con 3 modelos lineales identificando los periodos de entrada en vigencia de las leyes. Con ello, los resultados entregan información acerca de la efectividad de ambas leyes, separados por zona y además incluyendo las variables de vehículos en circulación e ingreso medio del hogar por comuna.

Al analizar los modelos es posible ver que de manera general que ambas leyes se relacionan con una disminución de los accidentes de tránsito. Sin embargo, al agregar al análisis las variables de ingreso y vehículos se obtiene que ambas leyes no se relacionan con una disminución. Tanto el ingreso de los hogares como el número de automóviles muestran una correlación positiva con el número de accidentes. Sin embargo, al desagregar el modelo por zona geográfica, se obtiene que ambas leyes son efectivas sólo en la zona *Sur Oriente* de la región. Esto puede ser debido a que los accidentes no necesariamente ocurren en las zonas donde las personas viven, así como, que la investigación no incorpora las atenciones de urgencia por la misma causa realizadas en establecimientos de salud privados.

En resumen, los resultados muestran concordancia con los patrones esperables de disminución de accidentes. Por su parte, los modelos permiten de manera simple identificar si existe impacto en la cantidad de accidentes

de tránsito gracias a las leyes y cuantificar en cierto grado su magnitud. Sin embargo, es importante considerar que los datos corresponden a atenciones primaria de urgencias de establecimientos públicos por accidentes de tránsito, sin detallar las causas específicas que los causan y, además, no todos los accidentes en su totalidad derivan a una atención de urgencia. De esta manera, a futuro se podría ampliar el estudio incluyendo la severidad del accidente (leve, moderado o grave), el número de muertes ocasionadas por dicho accidente, la ingesta de alcohol. Es importante conocer cuántas víctimas de accidentes han quedado con discapacidad permanente y saber si han sido indemnizadas, así como si ha sido indemnizada la familia en caso de la ocurrencia de víctimas fatales. Del mismo modo, se podría hacer un seguimiento a cada accidente con el fin de determinar a cuántos de ellos se les ha aplicado de manera efectiva las sanciones que consideran ambas leyes.

En términos generales la investigación realizada logra cumplir con el objetivo planteado entregando una evaluación objetiva sobre el impacto de las leyes aún frente a una gran limitación sobre los datos disponibles. Este mismo análisis con un mayor horizonte de tiempo y más detalles sobre las causas de los accidentes permitiría mejorar la cuantificación del impacto de estas leyes. Los modelos realizados permiten evaluar la implementación de políticas públicas sin necesidad de complejos cálculos, con resultados fáciles de analizar y utilizando herramientas de uso libre, lo cual constituye una metodología de análisis asequible para los organismos públicos y analistas del sector. Finalmente esperamos que este estudio contribuya a comprender de mejor manera los efectos de las leyes una vez aprobadas y, de este modo, mejorar los procesos legislativos y su ejecución.

Referencias

- [1] AChS. Efectos del alcohol al conducir y ley tolerancia cero, 2016.
- [2] J. Aguero-Valverde y P. Jovanis. Spatial analysis of fatal and injury crashes in pennsylvania. *Accident Analysis & Prevention*, 38(3):618–625, 2006.
- [3] S. Anowar, M. Alam, y M. Raihan. Analysis of accident patterns at selected intersections of an urban arterial. In *Proceedings of the 21st ICTCT Workshop*, 2008.
- [4] L. Fridstrøm. Econometric models of road use, accidents, and road investment decisions. *Volume II*, 1999.

- [5] L. Fridstrøm. The truls-1 model for norway. In *Structural road accident models: The international DRAG family*, páginas 97–126. Emerald Group Publishing Limited, 2000.
- [6] L. Fridstrøm. Un enfoque econométrico para los accidentes de tránsito, 2008.
- [7] L. González-Wilhelm, J. Johnson, R. Carnevali, y G. Ruiz. Ley chilena de tolerancia cero al alcohol: fortalezas, falencias y carencias que no deben ser obviadas. *Revista médica de Chile*, 140(7):939–945, 2012.
- [8] D. Graham y S. Glaister. Spatial variation in road pedestrian casualties: the role of urban scale, density and land-use mix. *Urban Studies*, 40(8):1591–1607, 2003.
- [9] P. Greibe. Accident prediction models for urban roads. *Accident Analysis & Prevention*, 35(2):273–285, 2003.
- [10] M. Haque, H. Chin, y B. Lim. Effects of impulsive sensation seeking, aggression and risk-taking behaviors on the vulnerability of motorcyclists. *Asian Transport Studies*, 1(2):165–180, 2010.
- [11] MTT. Ley 20.580. modifica Ley N° 18.290, aumentando las sanciones por manejo en estado de ebriedad, bajo la influencia de sustancias estupefacientes o sicotrópicas, y bajo la influencia del alcohol. Ministerio De Transportes Y Telecomunicaciones; Subsecretaría De Transportes, 2012.
- [12] MTT. Ley 20.770. modifica la ley del tránsito, en lo que se refiere al delito de manejo en estado de ebriedad, causando lesiones graves, gravísimas o, con resultado de muerte, 2014.
- [13] R. Noland y M. Quddus. Congestion and safety: A spatial analysis of london. *Transportation Research Part A: Policy and Practice*, 39(7):737–754, 2005.
- [14] M. Obaidat y T. Ramadan. Traffic accidents at hazardous locations of urban roads. *Jordan Journal of Civil Engineering*, 6(4):436–447, 2012.
- [15] OMS. Lesiones causadas por el tránsito. *Centro de prensa Organización Mundial de la Salud - Nota descriptiva*, 2016.
- [16] OMS y OPS. Informe 2016: Seguridad vial en las américas. *Organización Panamericana de la Salud 2016*, 2016.
- [17] D. Pomeranz. Métodos de evaluación. *Harvard Business School*, 2011.

- [18] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [19] SEMDA. El consumo de alcohol en Chile: situación epidemiológica. *Documento trabajado por SENDA-MINSAL*, 2016.
- [20] R. Seva, G. Flores, M. Gotohio, y N. Paras. Logit model of motorcycle accidents in the Philippines considering personal and environmental factors. *International Journal for Traffic and Transport Engineering*, 3(2):173–184, 2013.
- [21] D. Wedagama y D. Dissanayake. The influence of accident related factors on road fatalities considering Bali province in Indonesia as a case study. *Journal of the Eastern Asia Society for Transportation Studies*, 8:1905–1917, 2010.

PROGRAMACIÓN ENTERA PARA EL DISEÑO DE JORNADAS LABORALES DE REPONEDORES EN LA INDUSTRIA DEL RETAIL

JAIME MIRANDA *
FERNANDA PALACIOS **

Resumen

De forma periódica, las compañías proveedoras de la industria del Retail deben realizar la programación de las visitas para la reposición de sus productos en sus puntos de ventas. Esta programación de actividades incluye tanto, el diseño de las jornadas laborales semanales de los reponedores, así como, el secuenciamiento de visita de los puntos que cada reponedor debe realizar diariamente. Tanto el diseño de las jornadas laborales como el secuenciamiento de visitas deben considerar una serie de requerimientos operativos y propios del negocio, así como se debe adaptar a la demanda de los consumidores finales, transformando estos dos problemas en tareas complejas a realizar manualmente. Ser eficiente en la reposición de sus productos es una tarea fundamental para evitar los quiebres de stock en los puntos de venta, mejorando la disponibilidad de los productos y, por ende, las ventas de estas compañías.

En este artículo se presenta un enfoque de solución que permite resolver el problema asociado a la gestión del personal de reposición de productos para una compañía proveedora de la industria del Retail basado en modelos de programación lineal entera mixta. Este enfoque fue utilizado para diseñar las jornadas laborales diarias y semanales para la compañía SC Johnson. Los resultados obtenidos muestran que las soluciones entregadas por nuestro enfoque permiten cubrir un mayor número de puntos de ventas, aumentando su cobertura desde un 68 % a más de un 88 % de los puntos de venta, utilizando la misma dotación de personal respecto del enfoque manual.

Palabras Clave: Gestión de personal, programación de operaciones, ruteo.

*Departamento de Control de Gestión y Sistemas de Información, Facultad de Economía y Negocios, Universidad de Chile, Santiago, Chile.

**S. C. Johnson & Son

1. Introducción

Hoy en día muchas compañías utilizan las diferentes cadenas de supermercado como puntos de venta para que sus productos puedan llegar a sus clientes finales. Un aspecto fundamental para hacer rentable un punto de venta es que siempre los clientes encuentren disponibles en las góndolas los productos que andan buscando, tanto en tamaño como en forma. Cuando un punto de venta tiene problemas de gestión de suministro se provocan los quiebres de stock, generando que un consumidor no encuentre un producto en su lugar habitual, variedad y forma deseada. En consecuencia, un buen indicador de efectividad que mide la gestión de un punto de ventas es el porcentaje de quiebres de stock o Faltantes de Mercadería en Góndolas (FMG) que posee en el tiempo.

Existen muchas razones por las cuales existen FMG en un punto de venta siendo posible mencionar por ejemplo: 1) problemas de gestión del inventario del producto, 2) que los supermercados no realicen a tiempo sus pedidos de mercadería a sus proveedores, 3) malos pronósticos de demanda, 4) problemas de gestión de envío desde los centros de distribución o 5) problemas asociados a la reposición y entrega de los mismos proveedores, entre otros. Es claro que cuando un consumidor no compra un producto debido a que no se encuentra disponible en un punto de venta existe una pérdida para la compañía, pues su nivel de satisfacción es necesariamente menor. Esta baja en la satisfacción puede ocasionar diferentes comportamientos en los consumidores como cambios de marca y formatos, cambios de supermercado, postergaciones de la compra o simplemente no comprarán el producto. Por tanto, cualquiera sea la reacción de un consumidor existirán pérdidas económicas y de imagen, tanto para el supermercado, como para las compañías que tienen sus productos en las góndolas.

Según algunos estudios (GS1, 2012) en los países desarrollados el FMG alcanza un 8 %, mientras que en Latinoamérica bordea el 15 % existiendo una gran dispersión. Por ejemplo en el año 2012, en República Dominicana el FMG es de un 28.07 %, mientras que en Uruguay y Costa Rica es de un 6.36 % y un 7.3 % respectivamente. Particularmente, en Chile el FMG bordea el 15 % siendo la mala gestión en la reposición de los productos el causante del 41 % de estos quiebres de stock.

En general, la reposición de los productos de una góndolas en un punto de venta la realiza la misma cadena de Retail con personal interno. Si consideramos que para las cadenas de Retail el personal interno de reposición es un

recurso escaso y que estas compañías direccionan sus esfuerzos en la reposición sobre los productos que le entregan un mayor beneficio, es natural pensar que en la práctica no siempre repongan los productos de los diferentes proveedores en las góndolas. Esto último genera que los productos de los proveedores menos rentables no sean atendidos con la misma regularidad. Por este motivo, algunas compañías proveedoras de la industria del Retail han decidido contar con sus propios reponedores para apoyar la gestión de reposición interna de los distintos puntos de venta. Este apoyo busca tener siempre en góndola todos los productos de la empresa proveedora independiente de los esfuerzos que realice un punto de venta particular.

En términos prácticos, la frecuencia y regularidad con que se reponen los productos en las góndolas de un punto de venta está relacionada principalmente con los patrones de consumo que tienen los consumidores del producto en el tiempo y con el volumen de ventas o tamaño que posee un punto de venta particular. En la práctica, estas empresas proveedoras atienden uno o más puntos de venta en un día, siendo estos puntos desde pequeños supermercados de barrio hasta mega supermercados, observándose además que las ventas de productos tienen marcados factores estacionales, claras tendencias y *peaks* en el tiempo. Por tanto, la reposición de productos y, por ende, las jornadas laborales de los reponedores, deben ajustarse al consumo de los productos y al volumen total de ventas.

Respecto de las jornadas laborales de los reponedores, éstas deben respetar una serie de restricciones por ley, como por ejemplo respetar un número máximo de horas trabajadas por día y semana, así como se deben respetar otras condiciones impuestas por las mismas cadenas del Retail, como mantener el mismo reponedor que visita un punto de venta particular o tener un número mínimo de horas de reponedor en cada uno de sus puntos. En consecuencia, la confección de la programación de jornadas laborales diarias y semanales de los reponedores se transforma en un problema difícil de resolver en la práctica al ser un problema altamente combinatorial.

En este artículo se presenta un enfoque de solución basado en 2 modelos de optimización secuenciales que resuelven el problema del diseño de las jornadas laborales diarias y semanales para una empresa proveedora de la industria del Retail. Con el primer modelo se busca ajustar la demanda semanal por reposición de productos de los distintos puntos de venta dentro de patrones de visita diarios predeterminados con el fin de hacer calzar la demanda con las jornadas laborales de los reponedores. Mientras que el segundo modelo tiene como objetivo determinar las jornadas laborales de los reponedores a nivel diario y semanal, es decir, se determina el secuenciamiento de visita diario de los distintos puntos de venta. Este enfoque fue aplicado a la Compañía SC

Johnson (SCJ) en su filial de Chile.

La estructura de este artículo es como sigue. La sección 2 presenta la revisión de la literatura existente, la sección 3 presenta en detalle le problema estudiado, la sección 4 presenta el enfoque de solución propuesto, la sección 5 presenta los resultados experimentales. Finalmente, la sección 6 presenta las conclusiones de este trabajo.

2. Revisión de la literatura

La programación del personal corresponde al proceso por el cual se construyen programaciones horarias de las tareas que realiza el personal de una organización con tal de satisfacer una demanda por un producto o servicio. Se ha mostrado en la literatura que es extremadamente difícil encontrar buenas soluciones de forma manual, dado que el problema tiene un alto número de condiciones y requerimientos, lo que hace difícil encontrar soluciones óptimas que minimicen costos, que asegure cargas de trabajo parejas y que satisfagan condiciones asociadas a los lugares de trabajo [5].

Existe una amplia literatura en Investigación de Operaciones asociada a la gestión del personal en diferentes industrias, como por ejemplo: transporte de aerolíneas [4], servicios de correo [2], enfermería [3] y mantenimiento de aviones [1], entre otros. En general, este tipo de problemas han sido resueltos mediante enfoques heurísticos [8] o mediante modelos de optimización [9]. Una interesante revisión de la literatura existente la presentan Van den Bergh *et al* [10].

Para mejorar la gestión de la cadena de suministro y, por ende, dar un mejor servicio al cliente, es fundamental gestionar de manera eficiente el recurso humano con que se cuenta [6]. Por tanto, la programación de actividades de este recurso es un punto crucial para la eficiencia de las operaciones de la cadena de suministro. Sin embargo, si consideramos que la distribución de los periodos de mayor carga trabajo durante el día o semana es desigual, genera que en muchas ocasiones no sea posible hacer calzar turnos estándares de 8 horas. Este último elemento aumenta la complejidad para los administradores, pues aumenta la combinatoriedad del problema [3]. Al existir un desbalance de la carga de trabajo genera que en ciertas ocasiones se tengan que asignar turnos más cortos o más largos, que se obligue a algunos empleados a tomar un día libre o contratar trabajadores a tiempo parcial [7].

Cabe destacar que a diferencia de los trabajos presentados anteriormente, en nuestro caso además de tener que programar las visitas de los reponedores,

incorporamos dentro de nuestro enfoque dos elementos adicionales asociados a la definición del secuenciamiento de visita de los puntos de venta, es decir optimizar el ruteo, además de diseñar una solución para poder hacer una distribución eficiente de la carga de trabajo durante los días de la semana para los reponedores.

3. Descripción del caso de estudio

S.C. Johnson (SCJ) es una compañía estadounidense con sede en Chile, fabricante de diferentes productos farmacéuticos, productos de cuidado personal, perfumes y productos para bebés. Actualmente, esta compañía comercializa sus productos en más de 1.000 puntos de ventas distribuidos en todo Chile.

Para cada punto de venta es posible determinar el número total vendido por semana para cada producto, siendo el número total de horas de reposición que demanda cada punto de venta directamente proporcional al número total de cajas vendidas en dicho punto. Si llamamos m al número total de cajas vendidas en una semana en un punto de venta y n al número de cajas por hora que un reponedor puede restituir en las góndolas, el número total de horas de reposición de dicho punto será igual a m/n . Cabe destacar que en este problema se deben definir tanto las jornadas diarias como semanales de manera conjunta para los reponedores, por tanto se hace necesario determinar la distribución de las horas semanales de cada punto de venta en los diferentes días de la semana. Al resolver este último problema será posible determinar qué días de la semana se debe realizar una visita y cuál debe ser su duración.

Actualmente, la empresa cuenta con 68 reponedores, los cuales cubren cerca del 68% de los puntos de venta de la compañía visitándolos en algún momento de la semana. Los reponedores visitan entre 1 y 4 puntos de ventas diariamente con un máximo de 9 puntos de venta a la semana. Las jornadas laborales de los reponedores deben respetar una serie de condiciones, como por ejemplo: un máximo de 8 horas diarias y 45 horas semanales, contemplar al menos dos días libres cada semana y que al menos un día domingo esté libre cada dos semanas. Otro aspecto que agrega mayor complejidad a la resolución del problema recae en que existen ciertas cadenas de supermercados que exigen que un punto de venta siempre sea visitado por el mismo conjunto de reponedores. Esta condición busca tener un mayor control del ingreso de personal externo al supermercado.

En general, la cobertura busca visitar siempre a los puntos de venta que consumen la mayor cantidad de productos en los días en que éstos poseen la

mayor venta. Sin embargo, al hacer un análisis más fino de la cobertura, estas visitas no siempre cumplen con la cantidad de horas que el punto de ventas necesita. La Figura 1 muestra el porcentaje de cobertura de los puntos de venta que son cubiertos respecto del porcentaje de horas asignadas al punto de venta respecto del total de la solución propuesta por el enfoque manual. De este gráfico, se deduce que solo al 57% de los puntos de venta se les asigna al 100% de las horas requeridas para la reposición de los productos y, por tanto, más del 40% de los puntos de venta posee un número menor de horas requeridas lo que se traduce en un pobre abastecimiento de las góndolas de productos. Cabe destacar que SCJ utiliza un enfoque manual que resuelve el problema mediante la prueba y el error.

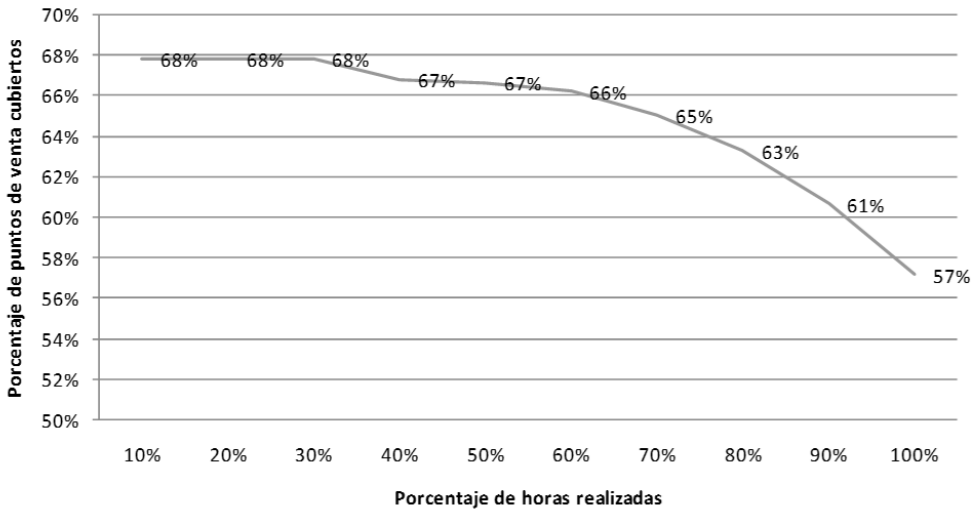


Figura 1: Cobertura de los puntos de venta respecto del porcentaje de horas asignadas.

De acuerdo con lo anterior, el desafío principal de este trabajo es desarrollar un enfoque de solución analítico que permita optimizar las jornadas laborales diarias y semanales de los reponedores. En este ámbito se busca por un lado maximizar cobertura total y asegurar coberturas mínimas a ciertas cadenas con los recursos que actualmente posee SCJ.

4. Descripción del enfoque de solución

El enfoque de solución propuesto tiene como objetivo determinar las jornadas diarias y semanales de visitas a los puntos de venta de los reponedores. Este

enfoque contempla 3 fases secuenciales las cuales se describen a continuación.

4.1. Fase I: Distribución de horas de reposición semanal

La primera etapa de nuestro enfoque busca determinar la distribución diaria de las horas de reposición de cada punto de venta en una semana. La forma en que se realizó esta tarea fue por medio de la programación de actividades sobre la base de dos tipos de patrones: 1) Patrones de días (por ejemplo lunes-miércoles-viernes) y 2) patrones de horas diarias (por ejemplo 2 horas diarias o 3.5 horas diarias). Estos patrones de visita definen los días y número de horas que son permitidos para que un reponedor visite un punto de venta y reponga las góndolas de productos. A continuación se describe el modelo de programación lineal binario utilizado en esta fase.

Conjuntos de índices:

- i : Puntos de venta.
- d : Días de la semana.
- p : Patrones de visita

Parámetros:

- V_{id} = Número total de cajas vendidas en el punto de venta i el día d .
- H_i = Número total de horas de reposición requeridas por el punto de venta i .
- Dia_i = Número de días que debe ser visitado el punto de venta i en una semana. Este parámetro define el patrón de días.
- $Horas_p$ = Número de horas que contiene el patrón p .
- Pat_visita_{ip} = 1 si el punto de venta i puede utilizar el patrón de horas p ; 0 en caso contrario.
- Pat_Dia_{dp} = 1 si el día d está contenido en el patrón de horas p ; 0 en caso contrario.
- Max_{Dif} = Máxima diferencia en el número de horas que se pueden asignar entre días consecutivos de una semana para un punto de venta.

VARIABLES DE DECISIÓN:

- $X_{idp} = \begin{cases} 1, & \text{si el punto de venta } i \text{ utiliza un patrón de horas } p \text{ el día } d \\ 0, & \text{en otro caso} \end{cases}$
- $Y_{ip} = \begin{cases} 1, & \text{si el punto de venta } i \text{ utiliza el patrón de horas } p. \\ 0, & \text{en otro caso} \end{cases}$
- $W_{id} = \begin{cases} 1, & \text{si el punto de venta } i \text{ es visitado el día } d \\ 0, & \text{en otro caso} \end{cases}$

Función Objetivo:

Esta función busca maximizar que el número total de horas asignadas para la reposición de las góndolas sea un día que el punto de venta tenga el mayor volumen de ventas. Dado que los mayores días de venta hay más quiebres de góndolas, se busca que esos días se asignen un mayor número de horas de reposición.

$$Maxz = \sum_i \sum_d \sum_p X_{idp} \cdot Horas_p \cdot V_{id}$$

Restricciones:

1. Satisfacer para cada punto de venta el número total de horas semanales requerido:

$$\sum_d \sum_p X_{idp} \cdot Horas_p = H_i \quad \forall i.$$

2. Asignar el número de días en que se debe visitar cada punto de venta:

$$\sum_d \sum_p X_{idp} = Dia_i \quad \forall i.$$

3. No asignar más de un patrón de horas por día a cada punto de venta:

$$\sum_p X_{idp} \leq 1 \quad \forall i, d.$$

4. Relación de variables y activación de la variable W :

$$\sum_p X_{idp} = W_{id} \quad \forall i, d.$$

5. Eliminar patrones de visitas que no son factibles para un punto de venta:

$$Y_{id} \leq Pat_{visitaip} \cdot Pat_{Dia_{dp}} \quad \forall i, p.$$

6. No utilizar más de un patrón de visita por punto de venta:

$$\sum_p Y_{ip} \leq 1 \quad \forall i.$$

7. No asignar más de un máximo de horas de diferencia (Max_{Dif}) entre los días de visita d y d' a un mismo punto de venta:

$$\sum_p X_{idp} \cdot Horas_p - \sum_p X_{id'p} \cdot Horas_p \leq Max_{Dif} + M \cdot (2 - W_{id} - W_{id'})$$

$$\forall i, d, d'. M \gg 1$$

$$\sum_p X_{idp} \cdot Horas_p - \sum_p X_{id'p} \cdot Horas_p \leq -Max_{Dif} - M \cdot (2 - W_{id} - W_{id'})$$

$$\forall i, d, d'. M \gg 1$$

8. Naturaleza de las variables:

$$X_{idp}, Y_{ip}, W_{id} \in \{0, 1\} \quad \forall i, p, d.$$

4.2. Fase 2: Generación del conjunto total de rutas factibles

La segunda etapa del enfoque de solución tiene como objetivo determinar el conjunto de rutas de visita factibles que puede realizar un reponedor. Cabe destacar que el resultado entregado en la Fase 1 permite obtener para cada punto de venta el número total de horas que necesita para la reposición de las góndolas cada día. Por tanto, la Fase 1 determina la demanda de horas de reposición y su distribución semanal. Con esta información, se diseñó un algoritmo de enumeración explícito que entrega el conjunto total de rutas factibles que pueden realizarse cada día. Una ruta determina el secuenciamiento de visita de los puntos de venta en un día y, se considera que es factible cuando la suma total de los tiempos viaje más los tiempos que se usarán para la reposición de productos no supere las 8 horas. Además, para cada ruta se define un costo calculado como la diferencia entre el máximo tiempo de una ruta (8 horas) y el largo total de la ruta. Las rutas que tienen una mayor diferencia de horas se consideran rutas más costosas al tener un mayor tiempo ocioso del reponedor.

4.3. Fase 3: Determinación de jornadas laborales y coberturas

La tercera etapa del enfoque de solución determina las jornadas laborales, diarias y semanales, del conjunto de reponedores de manera simultánea, mediante un modelo de programación lineal binaria. Cabe destacar que el diseño de las jornadas laborales considera todas las leyes laborales existentes, así como condiciones que tienen que ver con coberturas mínimas a ciertas cadenas de puntos de venta. El horizonte de planificación es de 2 semanas pues hay condiciones que son impuestas cada 14 días.

Conjunto de índices:

- i : Puntos de venta.
- d : Días de la semana.
- r : Rutas factibles.
- m : Reponedores.
- c : Cadenas de puntos de venta.
- $E_d =$ Conjunto de los 5 días seguidos al día d . Por ejemplo, si el día d es lunes, $E_{lunes} =$ lunes, martes, miércoles, jueves, viernes, sábado.

Parámetros:

- $b_{ri} = 1$ si la ruta r contiene el punto de venta i ; 0 en caso contrario.
- $a_{ri} =$ Número de horas de reposición asignadas al punto de venta i que está contenido en la ruta r ; Este parámetro toma un valor igual a 0 si no está contemplada una visita a dicho punto de venta en la ruta r .
- $N_i =$ Número total de días de visitas asignado al punto de venta i .
- $CoberturaMin_c =$ Número mínimo de horas de visita que deben tener todos los puntos de venta de la cadena c .
- $MaxMercad_i =$ Número máximo de reponedores distintos que pueden visitar el punto de venta i .
- $rutadia_{rd} = 1$ si la ruta r debe ser realizada el día d ; 0 en caso contrario.
- $rutad_r = 1$ si la ruta r debe ser realizada el día domingo; 0 en caso contrario.

- $f_{ic} = 1$ si el punto de venta i pertenece a la cadena de supermercados c ; 0 en caso contrario.
- C_{rm} = Costo de la ruta r si es asignada al reponedor m .
- CC_m = Costo de contratación del reponedor m .

Variables de decisión:

- $x_{rm} = \begin{cases} 1, & \text{si el reponedor } m \text{ realiza la ruta } r \\ 0, & \text{en otro caso.} \end{cases}$
- $z_{mi} = \begin{cases} 1, & \text{si el reponedor } m \text{ visita el punto de venta } i \\ 0, & \text{en otro caso.} \end{cases}$
- $g_{im} = \begin{cases} 1, & \text{si } m \text{ es el reponedor "principal" del punto de venta } i \\ 0, & \text{en otro caso.} \end{cases}$
- $y_m = \begin{cases} 1, & \text{si el reponedor } m \text{ trabaja algún día dentro del horizonte} \\ 0, & \text{en otro caso.} \end{cases}$

Función objetivo:

Esta función busca minimizar los costos totales de la operación, es decir, minimizar el costo total de la asignación de rutas más el costo total asociado a la contratación de reponedores.

$$\text{Min } z = \sum_r \sum_m C_{rm} x_{rm} + \sum_m CC_m y_m$$

Restricciones:

1. Todos los puntos de venta deben ser visitados por alguna ruta y reponedor:

$$\sum_m \sum_r b_{ri} x_{rm} = 1 \quad \forall i.$$

2. Cada ruta se puede realizar a lo más una vez por el conjunto de reponedores:

$$x_{rm} \leq 1 \quad \forall r.$$

3. Cada reponedor hace a lo más una ruta por día:

$$\sum_r rutadia_{rd}x_{rm} \leq 1 \quad \forall m, d.$$

4. Un reponedor puede trabajar a lo más 6 días seguidos:

$$\sum_{j \in E_d} \sum_r rutadia_{rd}x_{rm} \leq 6 \quad \forall m, d.$$

5. Cada reponedor debe tener un domingo libre cada dos semanas:

$$\sum_r rutad_r x_{rm} \leq 1 \quad \forall m.$$

6. Relación de variables y activación de la variable z_{mi} :

$$\sum_r b_{ri}x_{rm} \leq N_i z_{mi} \quad \forall m, i.$$

7. Número máximo de reponedores distintos por punto de venta:

$$\sum_m z_{im} \leq MaxMercad_i \quad \forall i.$$

8. Activación variable que identifica el reponedor “principal” de punto de venta:

$$\sum_r b_{ri}x_{rm} \leq P_i g_{mi} \quad \forall m, i.$$

9. Cada punto de venta tiene un reponedor “principal”:

$$\sum_m g_{im} \geq 1 \quad \forall i.$$

10. Activación variable que verifica si el reponedor m tiene asignada alguna ruta:

$$\sum_r x_{rm} \leq 12y_m \quad \forall m.$$

11. Naturaleza de las variables:

$$x_{rm}, z_{ip}, g_{im}, y_m \in \{0, 1\} \quad \forall r, m, i, p.$$

5. Resultados experimentales

En esta sección se presentan los resultados obtenidos por la aplicación del enfoque de solución propuesto respecto del procedimiento manual que utilizaba SCJ al momento de la aplicación. La instancia utilizada contempla 290 puntos de venta pertenecientes a 9 cadenas de supermercados diferentes los cuales se encuentran distribuidos geográficamente en todo Santiago (ver Figura 2). Cabe destacar que los reponedores se movilizan utilizando el transporte público de Santiago. Los tiempos de viaje para el transporte público fueron obtenidos utilizando el servicio entregado por *Google Maps* <https://developers.google.com/maps/?hl=es>.

Para la resolución de los modelos de optimización se utilizó el software comercial GAMS 23.5.2 y el solver Gurobi 3.0.1. El tiempo de resolución del modelo utilizado en la Fase 1 fue de 1.5 minutos, mientras que el tiempo de resolución del modelo utilizado en Fase 3 es de 20 horas. En ambos casos se obtuvo la solución óptima del problema con GAP igual a 0. Respecto del tiempo computacional del algoritmo de enumeración explícita fue de 4.5 minutos, entregando 20.144 rutas factibles.

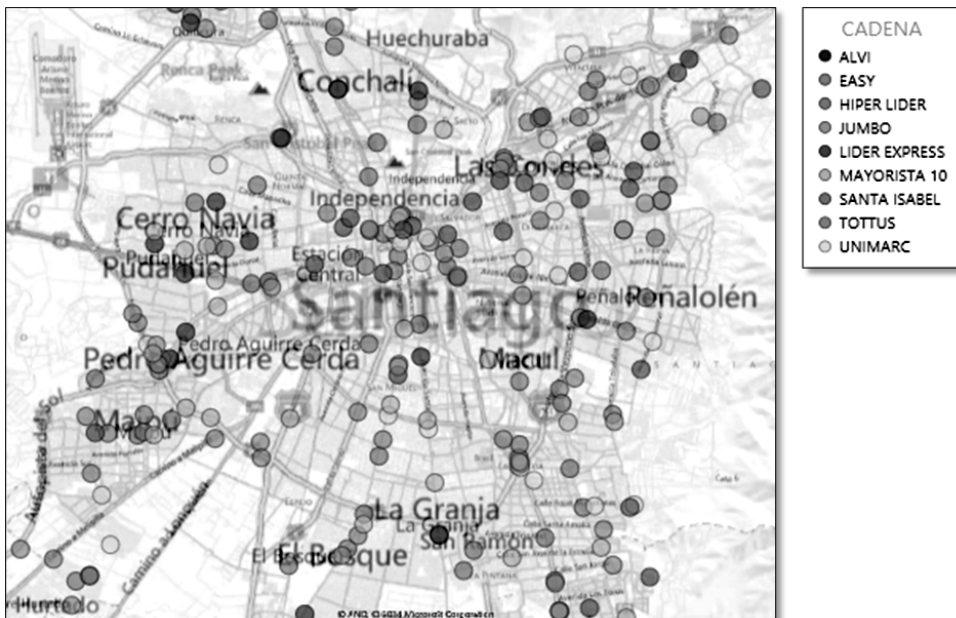


Figura 2: Distribución de los puntos de venta y cadenas de Santiago.

La Figura 3 muestra el porcentaje de cobertura de los puntos de venta respecto del porcentaje de horas realmente asignadas respecto del total, para un igual número de reponedores. La solución propuesta visita prácticamente todos los puntos de venta al menos una vez y, de estos puntos de venta, el 88 % son visitados todas las horas que requieren, a diferencia de la solución manual que sólo alcanza un 68 % del total. Por tanto, en promedio la solución propuesta por nuestro enfoque supera en más de 20 puntos porcentuales a la solución manual.

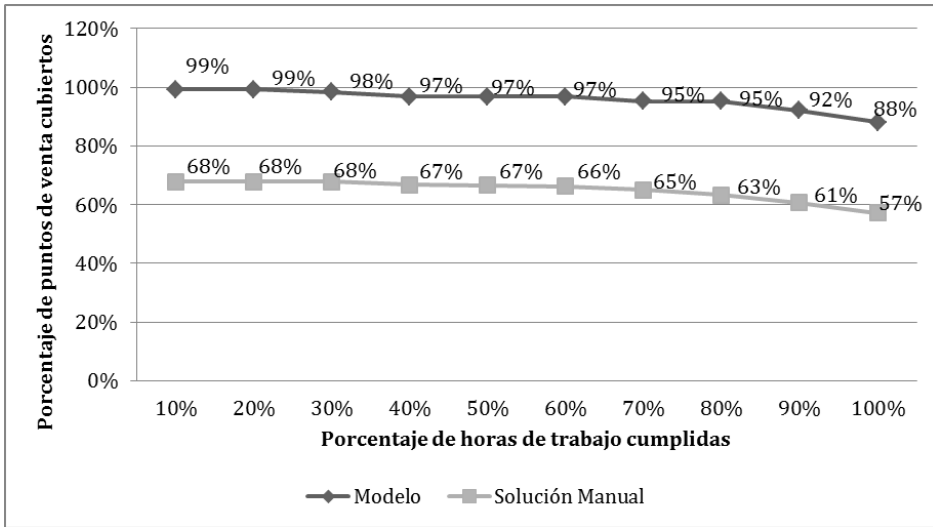


Figura 3: Cobertura de los puntos de venta de Santiago respecto del número de horas asignadas.

6. Conclusiones

En este trabajo se presenta un enfoque de solución que permite resolver el problema de reposición de productos de la compañía SC Johnson para su red de puntos de venta.

El enfoque propuesto genera resultados satisfactorios para la compañía, pues las soluciones entregadas permiten cubrir un mayor número de puntos de venta con la misma dotación de personal superando en más de 20 puntos porcentuales a la solución manual. Basado en los resultados, se concluye que con la incorporación de nuestro enfoque de solución la compañía se verá beneficiada la gestión de las visitas de los reponedores, encontrando soluciones más eficientes y a mínimo costo.

Referencias

- [1] H. Alfares. Aircraft maintenance workforce scheduling a case study. *Journal of Quality in Maintenance Engineering*, 5(2):78–89, 1999.
- [2] J. Bard, C. Binici, et al. Staff scheduling at the united states postal service. *Computers & Operations Research*, 30(5):745–771, 2003.
- [3] E. Burke y T. Curtois. New approaches to nurse rostering benchmark instances. *European Journal of Operational Research*, 237(1):71–81, 2014.
- [4] E. Butchers, P. Day, A. Goldie, S. Miller, J. Meyer, D. Ryan, A. Scott, y C. Wallace. Optimized crew scheduling at air new zealand. *Interfaces*, 31(1):30–56, 2001.
- [5] A. Ernst, H. Jiang, M. Krishnamoorthy, y D. Sier. Staff scheduling and rostering: A review of applications, methods and models. *European journal of operational research*, 153(1):3–27, 2004.
- [6] D. Graham. Warehouse of the future. *Frontline Solutions*, 4(4):20–20, 2003.
- [7] A. Ladier, G. Alpan, y B. Penz. Joint employee weekly timetabling and daily rostering: A decision-support tool for a logistics platform. *European Journal of Operational Research*, 234(1):278–291, 2014.
- [8] Q. Lequy, G. Desaulniers, y M. Solomon. A two-stage heuristic for multi-activity and task assignment to work shifts. *Computers & Industrial Engineering*, 63(4):831–841, 2012.
- [9] E. Naudin, P. Chan, M. Hiroux, T. Zemmouri, y G. Weil. Analysis of three mathematical models of the staff rostering problem. *Journal of Scheduling*, 15(1):23–38, 2012.
- [10] J. Van den Bergh, J. Beliën, P. De Bruecker, E. Demeulemeester, y L. De Boeck. Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3):367–385, 2013.

ALGORITMOS DE ZONIFICACIÓN PARA EL PROBLEMA DE LA RECOLECCIÓN DE RESIDUOS URBANOS: EL CASO DE ESTUDIO DE UNA CIUDAD ARGENTINA

MARCELO BIANCHETTI *
GUILLERMO DURÁN **
IVO KOCH ***
JAVIER MARENCO ****

Resumen

La programación de los aspectos logísticos de la recolección de residuos urbanos plantea una serie de problemas de optimización combinatorial de difícil resolución en la práctica. Uno de estos problemas consiste en la división del área a recorrer en zonas, cada una de ellas correspondiente a un camión recolector. Para ello es preciso respetar restricciones de volumen de recolección por cuadra y sobre la flota de vehículos, así como también restricciones cualitativas de similitud de las zonas entre sí. En este artículo analizamos el problema aplicado a la recolección de residuos en la ciudad de San Miguel de Tucumán en Argentina. Estudiamos para ello distintos algoritmos heurísticos de zonificación que nos permiten proponer una partición en zonas alternativa a la actual, de forma tal de minimizar la cantidad de camiones utilizados.

Palabras Clave: Recolección de residuos, Zonificación, Heurísticas.

* Instituto de Industria, Universidad Nacional de General Sarmiento, Buenos Aires, Argentina

** Departamento de Matemática e Inst. de Cálculo FCEN-UBA y CONICET, Argentina, Departamento de Ingeniería Industrial, FCFM, Universidad de Chile, Santiago, Chile

*** Instituto de Industria, Universidad Nacional de General Sarmiento, Buenos Aires, Argentina

**** Instituto de Ciencias, Universidad Nacional de General Sarmiento y Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

1. Introducción

La recolección de residuos ha sido objeto de numerosos trabajos de investigación debido a la amplitud de problemas que surgen de la misma, tanto para el área de grafos como de la programación lineal entera; pero el mayor esfuerzo se ha hecho sobre la optimización del ruteo de vehículos y la asignación de horarios (ver por ejemplo [1, 5, 9, 13, 15, 19, 20, 21, 23, 27]). El presente trabajo aborda un problema distinto: el de la organización en zonas de una región de manera tal de minimizar la flota de camiones a utilizar. Este problema de *diseño territorial*, esto es, de agrupamiento de áreas geográficas básicas en zonas más grandes que sean aceptables de acuerdo a determinados criterios de planeamiento fue estudiado como parte de numerosas aplicaciones. Aparece en diseño de distritos electorales [16], de áreas de ventas [6], de distritos escolares [10], de territorios de servicio para prestaciones sociales como hospitales o unidades administrativas [22], y de distritos policiales [4], entre otros. Puede encontrarse una muy buena reseña en [17]. El problema de diseño territorial aplicado a la recolección de residuos y a operaciones de remoción de nieve y esparcimiento de sal en las calles fue analizado en [14] y [24], respectivamente. Los autores abordaron el problema de la recolección de residuos utilizando otras técnicas a las estudiadas aquí en [2] y [3].

El caso de estudio de del presente artículo será el Municipio de San Miguel de Tucumán, ciudad capital de la Provincia de Tucumán, situada en el noroeste de la Argentina.

El servicio de recolección municipal de residuos en esta ciudad se lleva a cabo en una franja horaria diurna y una nocturna. Esto se debe en parte a la falta de camiones suficientes para cubrir todas las zonas en paralelo, y en parte a cuestiones de tránsito y seguridad.

De martes a sábados, la cantidad de camiones con la que se cuenta alcanza para cubrir íntegramente las zonas; pero al no haber recolección los domingos, los lunes se incrementa el volumen de residuos a recolectar, formando un cuello de botella en el área del microcentro, que es un área densamente poblada delimitada por cuatro grandes avenidas. Aquí sus recorridos se trazan de una manera especial, haciendo un barrido horizontal y otro vertical. Es decir que los camiones sólo la recorren en líneas rectas que entran o salen de la misma, girando sólo fuera de ésta. La manera requerida por parte del municipio para optimizar este problema es asignar mayor cantidad de camiones a esta área. Como la cantidad de camiones es fija, para ello es preciso retirarlos de otras zonas de la misma franja horaria. A tal fin, desarrollamos en este trabajo heurísticas basadas en programación lineal entera con el objetivo de rezonificar el área nocturna recuperando la mayor cantidad de camiones posible para destinarlos al área del microcentro. Las heurísticas desarrolladas son originales

y difieren de las halladas en la bibliografía. En [14] los autores plantean técnicas metaheurísticas basadas en modificar sucesivamente los “centros” de las zonas: asignar el centro de una zona a otra o intercambiar dos centros, para intentar mejorar la zonificación; en nuestro trabajo los centros son un dato de entrada para el algoritmo y no se modifican dinámicamente. Por otra parte, el trabajo [24] coincide en una de las ideas de este artículo, de hacer crecer las zonas agregando ciclos. Sin embargo, esta estrategia fue reemplazada en este estudio por la agregación de aristas individuales a las zonas, por dar mejores resultados computacionales.

El presente trabajo se estructura de la siguiente manera: en el resto de esta sección precisamos en detalle los requerimientos del problema y damos algunas definiciones preliminares. En la Sección 2 desarrollamos los algoritmos que empleamos en la resolución. A continuación, la Sección 3 muestra los resultados computacionales obtenidos. La última sección indica algunas posibles líneas de trabajo futuro.

1.1. Consideraciones particulares del problema

El servicio se cumple en 40 zonas diurnas y nocturnas disjuntas, que abarcan 12.320 cuadras aproximadamente. Actualmente el área diurna cubre un total de 6.982 cuadras, de las cuales 1.710 pertenecen a calles o avenidas mano única mientras que 5.272 a calles o avenidas doble mano. Esta región se divide en 19 zonas con un camión asignado a cada una. El servicio nocturno cubre 21 zonas ubicadas en el centro de la ciudad con un total de 5.338 cuadras, de las cuales 3.515 pertenecen a avenidas y calles mano única, y 1.823 doble mano. Dentro de estas 21 zonas céntricas, el microcentro de la ciudad delimitado por las avenidas Sarmiento al Norte, Avellaneda - Sáenz Peña al Este, Roca al Sur y Alem - Mitre al Oeste es cubierto por 9 camiones. Este microcentro está compuesto por 2.092 cuadras. Las 3.246 cuadras restantes de la región, de las cuales 1.696 pertenecen a calles o avenidas mano única, y 1.550 a avenidas o calles doble mano son cubiertos por los 12 camiones restantes.

Procuraremos que las zonas en que buscamos dividir el área nocturna del municipio sean similares; es decir, que un camión tarde aproximadamente el mismo tiempo en recolectar los residuos en cada una de ellas. Buscamos entonces la menor varianza del conjunto de los tiempos que tardan todos los camiones en realizar sus servicios respectivos.

En las áreas a zonificar de la ciudad en cuestión hay contenedores y recolección domiciliaria. Existen ambas modalidades porque no en todas las cuadras existen contenedores; en las que carecen de ellos, la recolección es por cada domicilio. Por ello, es necesario que los camiones recorran absolutamente todas las calles.

Según los datos brindados por el municipio, el tiempo que consume un camión en recolectar residuos en una cuadra de una zona nocturna es 5 veces mayor en promedio que el que tarda en atravesarla sin recolectar, y 3 veces

mayor en el caso de la zona diurna. La razón de esta diferencia es que las zonas nocturnas cuentan con edificios altos y son más densamente pobladas, por lo que el tiempo de recolección en cada cuadra se incrementa. Se considera que el volumen de recolección dentro de cada una de las zonas (excluido el microcentro) es aproximadamente lineal con respecto a las longitudes de las cuadras a recolectar.

Sea la velocidad promedio de un camión V_p y V_r en cuadras de paso y recolección respectivamente. Se estima que $V_p = 30$ km/h mientras que $V_{r_{nocturna}} = 6$ km/h y $V_{r_{diurna}} = 9$ km/h, aproximadamente. Como el personal de recolección trabaja 8 hs diarias, la zona de cobertura de un camión en la región nocturna se limita a 48 km, mientras que en las diurnas ésta puede tener una longitud de 72 km.

Un camión puede recorrer calles de zonas vecinas si de esa manera reduce el tiempo total empleado para recorrer su zona. Asumimos que la capacidad de los camiones es suficiente para recolectar los residuos de su zona sin necesidad de descargar. Del mismo modo se considera que la autonomía de los vehículos es suficiente para salir del depósito, recolectar y volver a descargar, sin necesidad de recargar combustible.

1.2. Descripción formal del problema

Para las definiciones, propiedades y algoritmos de teoría de grafos no precisadas aquí, remitimos al lector a [26]. El mapa M de la región a estudiar se modelará como un grafo mixto $G = (V_G, E_G, A_G)$, tal que el conjunto de aristas no dirigidas $E_G \subseteq (V_G \times V_G)$ representará los segmentos de las calles y avenidas doble mano, sus arcos dirigidos $A_G \subseteq (V_G \times V_G)$ serán los segmentos de calles y avenidas mano única (su sentido estará dado por el orden de sus nodos), y el conjunto de vértices V_G representará los extremos e intersecciones de segmentos de caminos vehiculares.

Un k zonificado será una partición del grafo G en k subgrafos conexos (denominados *zonas* o *áreas*) $G_i = (V_i, E_i, A_i)$ tales que $\forall i: V_i \subseteq V_G, E_i \subseteq E_G \cap (V_i \times V_i), A_i \subseteq A_G \cap (V_i \times V_i)$, cumpliendo además que $\forall i \neq j: E_i \cap E_j = \emptyset, A_i \cap A_j = \emptyset$.

Para calcular el tiempo que consumiría un camión en recolectar todos los residuos de su zona es necesario resolver el problema del *cartero chino* (CPP) [7, 8, 18, 25] sobre el subgrafo asociado. Formalmente, dado un grafo mixto $G = (V, E)$ con longitudes positivas asignadas a sus aristas $L: E \rightarrow \mathbb{R}_+$, el problema del cartero chino consiste en encontrar un circuito de longitud mínima que pase por cada arista de G al menos una vez.

En un grafo dirigido o no dirigido el CPP puede ser resuelto en tiempo polinomial. Por otro lado, si el grafo es mixto el problema de decisión asociado es NP-Completo, aún si todas las longitudes de las aristas son iguales, el grafo es planar y su grado máximo es 3 [12]. Como la zonificación involucra resolver el CPP sobre grafos mixtos, nuestro problema es a su vez NP-Completo.

Una condición necesaria y suficiente para que exista una solución del CPP, es que G sea fuertemente conexo. Un grafo con estas características se denominará *factible*.

Como el origen del grafo es un mapa, debe ser factible. Además, como la simplificación preserva factibilidad, entonces para zonificar es necesario dividir un grafo G , que se sabe factible, en k subgrafos (G_1, \dots, G_k) disjuntos que se desean factibles. Este objetivo no siempre es alcanzable. Por ejemplo, si G es un grafo consistente en un único circuito simple (sin nodos repetidos), no es posible dividirlo en k subgrafos factibles disjuntos, para cualquier valor de $k > 1$.

Para cualquier particionamiento (G_1, \dots, G_k) del grafo factible $G = (V_G, E_G, A_G)$ en k subgrafos disjuntos, existen k conjuntos de aristas (C_1, \dots, C_k) tales que $G_i \cup C_i$ es factible para todo $i = 1 \dots k$. Sea $G_i = (V_{G_i}, E_{G_i}, A_{G_i})$, basta tomar $C_i = (A_G \cup E_G) \setminus (A_{G_i} \cup E_{G_i})$. Luego, siempre es posible factibilizar las zonas agregándoles aristas del grafo G . Por eso en este trabajo se definió que una zona está compuesta por aristas *obligatorias*, que son las que representan los caminos que un camión *debe* recolectar, y *optativas*, que son las que representan los caminos que *puede* atravesar (sin recolectar).

Definimos en este trabajo el peso $L : (E_G \cup A_G) \rightarrow \mathbb{R}^+$ para las aristas optativas de G como la distancia euclideana de los nodos que componen la arista sobre la cual se evalúa. El peso de las aristas obligatorias será 5 veces este valor (aplicaremos nuestros algoritmos a las zonas nocturnas). El tiempo que consume una arista en el recorrido de un camión entonces estará dado por $T : E_G \cup A_G \rightarrow \mathbb{R}$, tal que:

$$T(e) = \begin{cases} L(e)/V_r & \text{si } e \text{ es arista obligatoria} \\ L(e)/V_p & \text{si } e \text{ es arista optativa.} \end{cases}$$

Para calcular el *costo* de la zona, es decir el tiempo consumido por el camión en recorrerla recolectando los residuos de sus calles obligatorias, es necesario resolver esta variante del cartero chino sobre el subgrafo mixto asociado factibilizado con arcos y aristas optativas. Como un arco o una arista obligatoria, una vez recolectados los residuos, se convierte en arco o arista de paso, se determinó para el cálculo del circuito que todos los arcos y las aristas fueran tratadas como de paso, diferenciando únicamente las obligatorias forzando a que aparezcan al menos una vez. Luego, al circuito resultante se le suma el tiempo que consume la recolección. Si denominamos $f = \frac{\text{costo}_{\text{recoleccion}}}{\text{costo}_{\text{paso}}}$ entonces este tiempo estará dado por $f - 1$ veces el costo de pasar por los arcos y aristas obligatorias de la zona (se decrementa en 1 por ya haber contado una vez cada arista obligatoria).

Dos zonas se dirán *similares* si un camión tarda aproximadamente el mismo tiempo en recolectar los residuos en cada una de ellas.

Una *zonificación óptima* de la región en k zonas se definirá como aquel particionamiento del grafo asociado en k subgrafos conexos similares, siendo

además el promedio de los resultados del CPP en todos ellos el menor posible y dando prioridad a un grado de similitud mayor.

El requisito informal de la forma de las áreas se traduce en buscar que los subgrafos sean conexos, y que la varianza de las distancias euclidianas de los nodos más alejados del centro de la misma en las distintas direcciones sea pequeña.

2. Algoritmos desarrollados

La idea principal del método de resolución es generar las áreas en forma iterativa e incremental, de manera tal que en cada una de estas iteraciones la similitud entre ellas sea la mayor posible.

Como los subgrafos que representan las áreas generadas hasta el momento deben representar submapas del mapa original, buscamos que sea posible recorrerlos, i.e., que sea posible resolver en ellos el problema del cartero chino. Por este motivo se pedirá que dichos subgrafos sean fuertemente conexos. La manera de garantizar ese requisito será demostrar que cada subgrafo se puede describir como una unión de circuitos.

De aquí surgen dos de los principales criterios analizados en este trabajo: agregar en cada iteración o bien un circuito, o bien una arista (o arco). En el segundo caso será necesario agregarle al área como optativos un conjunto de circuitos que la cubran completamente para garantizar factibilidad.

Para decidir a qué zona agregar la arista o arco o circuito la hipótesis fundamental del presente trabajo es que la solución del CPP está fuertemente ligada con la cantidad de aristas y arcos que posee el grafo a resolver. Por este motivo se analizaron diversos criterios de manera de intentar agregar la arista o circuito a aquel grafo cuyo resultado del CPP fuera menor. Se procura además minimizar la cantidad de ejecuciones del CPP dado su costo computacional.

El Algoritmo 1 resume estas ideas. La resolución comienza convirtiendo un mapa de la región del municipio actualmente cubierta por el turno nocturno en formato *openStreetMap* (.osm) [11] al digrafo G que lo modela (línea 2), seguido de una simplificación del mismo que da lugar a un grafo equivalente H (línea 3). Posteriormente se verifica la factibilidad de H , se generan las distintas zonas H_i de forma heurística (línea 9), se factibilizan (línea 11) y resuelven los modelos de programación lineal entera mixta (PLEM) asociados para obtener una solución del problema del cartero chino en cada una (líneas 12 y 13). Esas soluciones son por último reescritas como recorridos (línea 14). Dedicamos el resto de esta sección al análisis de cada uno de estos pasos.

2.1. De un mapa al modelo

Los datos del mapa M del municipio en formato .osm constan de un conjunto de nodos $N_M = \{n_1, \dots, n_j\}$, un conjunto de caminos $W_M = \{w_1, \dots, w_k\}$,

Algoritmo 1 Algoritmo de zonificación

```

1: function ZONINGAMAP( $M : osmMap, K : osmKnodes, H : heuristic$ )
2:    $G := osmToGraph(M)$ 
3:    $S := simplifyGraph(G)$ 
4:    $S, B, E := graphByCircuits(S)$ 
5:   if not  $checkFactibility(S)$  then
6:      $printErrors(S, G, B)$ 
7:     STOP
8:   end if
9:    $GrownAreas := growAreas(S, K, B, H)$ 
10:  for  $k \in K$  do
11:     $Area_k := zoning(GrownAreas, k)$ 
12:     $LP_k := graphToModel(Area_k)$ 
13:     $Solution_k := LPSolver.Solve(LP_k)$ 
14:     $Path_k := solutionToPath(Solution_k, S, G)$ 
15:  end for
16: end function

```

y un conjunto de restricciones R_M . Cada *nodo* $n_i \in N_M$ tiene asociado un par de coordenadas $(latitud_i, longitud_i) \in \mathbb{R}^2$ y un conjunto de atributos. Un *camino* $w_i \subseteq N_M \times \dots \times N_M$ está definido como una sucesión ordenada de un subconjunto del conjunto de nodos y un listado de atributos. Dependiendo de éstos, el camino representará una calle doble mano, una avenida, un río, una vía de tren, el perímetro de un edificio público, etc. En caso de ser una calle mano única, el orden de los nodos indicará el sentido de la misma. Las *restricciones* $r_i \subseteq W_M \times N_M \times W_M$ determinan los giros prohibidos de un camino a otro camino a través de un nodo.

La conversión de este formato al grafo G sobre el cual se trabajará vendrá dada por:

- el conjunto de aristas no dirigidas $E_G \subseteq (V_G \times V_G) \subseteq (N_M \times N_M)$ cumplirá que $\forall (x, y) \in E_G : \exists w_i \in W_M$ tal que $w_i = (n_1, \dots, x, y, \dots, n_n)$ y cuyos atributos indican avenida o calle residencial de doble sentido.
- sus arcos dirigidos $A_G \subseteq (V_G \times V_G)$ cumplirán que $\forall (x, y) \in A_G : \exists w_i \in W_M$ tal que $w_i = (n_1, \dots, x, y, \dots, n_n)$ y cuyos atributos indican avenida o calle residencial de un solo sentido.
- el conjunto de vértices $V_G \subseteq N_M$ cumplirá que $v \in V_G \leftrightarrow \exists y : (v, y) \in (E_G \cup A_G) \vee (y, v) \in (E_G \cup A_G)$.

Se utilizará la distancia terrestre entre dos nodos para la función de peso $L : (E_G \cup A_G) \rightarrow \mathbb{R}^+$ para las aristas de G .

En este trabajo realizamos una simplificación del mapa original, para obtener instancias equivalentes más reducidas. En los mapas utilizados las curvas se representan mediante varios segmentos rectos para representar más fielmente la realidad, y para evitar superposición de caminos, por lo que una cuadra puede comprender un conjunto de aristas o arcos; reemplazamos estos casos por un solo segmento recto. Además, eliminamos los nodos *redundantes*, que unen exactamente dos segmentos con el mismo sentido y categoría (calle residencial o avenida). Denominaremos $H = (V_H, E_H, A_H)$ al grafo simplificado que resulta de la transformación, que describimos en detalle debajo:

i) El conjunto de aristas no dirigidas $E_H \subseteq (V_H \times V_H) \subseteq (V_G \times V_G)$ representan los segmentos comprendidos entre los extremos, intersecciones o cambios de categoría (esto es, avenida o calle) de los caminos doble mano. Cumplen la propiedad de que para toda arista (x, y) del grafo H , con $x \neq y$, existe un camino $C = \{n_1, \dots, n_k\} \subseteq V_G$ compuesto por aristas doble mano de G tal que todos los nodos internos de C son redundantes salvo sus extremos n_1 y n_k . Además se debe cumplir una de las tres condiciones siguientes: o bien $C = (x, y)$ o bien $(x, y) = (n_1, n_k) \notin E_G$ y ninguna de las aristas que forman a C están en H , o bien (x, y) pertenece a C siendo que o C es un circuito o la arista que reemplazaría al camino C ya existe en el grafo H .

ii) Los arcos dirigidos $A_H \subseteq (V_H \times V_H) \subseteq (V_G \times V_G)$ representan los segmentos comprendidos entre los extremos, intersecciones o cambios de categoría (esto es, avenida o calle) de los caminos de única mano. Verifican que para todo $(x, y) \in A_H$, con $x \neq y$ existe un camino dirigido $C = \{n_1, \dots, n_k\} \subseteq V_G$ compuesto por arcos pertenecientes a A_G con todos sus nodos internos redundantes, sus extremos n_1 y n_k no redundantes. Además, debe cumplirse que $C = (x, y)$, o bien $(x, y) = (n_1, n_k) \notin A_G$ y ninguno de los arcos que forman a C están en H , o bien (x, y) pertenece a C siendo que o C es un circuito o el arco que reemplazaría al camino C ya existe en el grafo H .

iii) El conjunto de vértices $V_H \subseteq V_G$ representa únicamente los extremos e intersecciones de caminos vehiculares, o la unión de dos segmentos de los mismos con distinto sentido y/o categoría. Se cumplirá que un vértice v pertenecerá a V_H si y sólo si existe un vértice y tal que la arista o arco (v, y) o su inversa es un arco o arista del grafo H .

2.2. Verificación de factibilidad general

Para verificar la factibilidad del grafo asociado al mapa a zonificar, se han utilizado dos algoritmos de complejidad polinomial que recurren a la siguiente propiedad, que puede probarse fácilmente.

Proposición 2.1. *Sea $G = (V_G, E_G, A_G)$ un grafo mixto, si vale que el grafo no dirigido $F = (V_F, E_F) : V_F = V_G$ y $E_F = (A_G \cup E_G)$ es conexo, y si toda arista de G tiene un circuito que la contiene, es decir $\forall (x, y) \in (E_G \cup A_G)$,*

existe un camino C tal que $C = \{(y, n_1), \dots, (n_k, x)\} \subseteq (E_G \cup A_G)$, entonces G es fuertemente conexo.

Por lo tanto lo que se buscó fue que existiera un circuito para cada arista y arco de G , y que el grafo F asociado fuera conexo. Para hallar un circuito para cada arista del grafo mixto $H = (V_H, E_H, A_H)$ se utilizó el algoritmo de Dijkstra para camino mínimo entre dos nodos (es decir, la versión que detiene el algoritmo original al hallar la solución para los nodos en cuestión) aplicado entre cada par distinto (x, y) con $(y, x) \in A_H$. El camino C hallado que une y con x es mínimo. Un arco sin circuito implica que el problema del cartero chino no tiene solución. Un mapa de una ciudad tiene que ser factible por ser la representación de las calles utilizadas para ir de algún lugar a cualquier otro. De existir arcos sin circuito, puede ser que los datos sean incorrectos, con lo que debe corregirse el mapa; o bien puede darse que este conjunto de calles intransitables se deba a un mal recorte del mapa original. Esto es, que se le haya quitado un segmento a algún camino que sale y entra a la ciudad. En el grafo simplificado H , el conjunto de arcos sin circuito está compuesto por aquellos que representan calles mano única pertenecientes a un camino sin salida o sin entrada. A su vez dichos arcos pueden particionar el grafo original generando circuitos desconectados del resto del grafo. La manera de solucionar este inconveniente es eliminar las aristas sin circuito.

Verificamos si el grafo H es conexo mediante el algoritmo *Breadth First Search (BFS)*, aplicado al grafo no dirigido $F = (V_F, E_F) : V_F = V_H \wedge E_F = (A_H \cup E_H)$, revisando que todos los nodos sean alcanzables a partir de uno cualquiera.

2.3. Generación de los subgrafos

Sean un grafo mixto $F = (V_F, E_F, A_F)$ y un circuito mixto $C = (V_C, E_C, A_C)$, definimos la *unión* de ambos como el grafo $(C \cup F) = (V_F \cup V_C, E_F \cup E_C, A_F \cup A_C)$. Decimos que dicha unión es *conexa* si $(V_C \cap V_F) \neq \emptyset$.

Proposición 2.2. *Sea G un grafo mixto factible, y C un circuito mixto, si la unión $G \cup C$ es conexa entonces el grafo resultante es factible.*

Corolario 2.1. *Una unión finita conexa de circuitos determina un grafo factible.*

En consecuencia, partiendo de un subgrafo H_i del grafo factible H , es posible agregar circuitos de H a H_i hasta dejarlo factible si éstos contienen las aristas que en H_i no pertenecen a ningún circuito.

Para generar una partición del grafo H que se sabe factible en subgrafos similares, se evaluó un conjunto de heurísticas que coinciden estructuralmente en la serie de pasos especificada en el Algoritmo 2, pero difieren en la combinación de criterios utilizados.

Estudiamos a continuación los criterios para dividir las aristas del grafo original en conjuntos.

Algoritmo 2 growAreas

- 1: Partir de una ubicación tentativa de los k centros de las zonas a generar. Estos centros, que tienen latitud y longitud, se denominan k -nodos. Asignar a cada uno un grafo vacío.
 - 2: Dividir las aristas del grafo original en h conjuntos según algún criterio (el particionamiento por aristas o por circuitos descrito debajo).
 - 3: Calcular las distancias de dichos conjuntos a los k -nodos. Para esto es necesario definir una noción de distancia.
 - 4: Mientras sea posible, seleccionar el k -nodo con el grafo más reducido (también debe definirse este criterio), y agregarle los nodos y aristas correspondientes al conjunto de aristas más cercano al mismo, de forma que lo mantengan conexo. Si no hay, marcar dicho k -nodo como *terminado* e intentar con el siguiente. El procedimiento continúa hasta marcar todos los k -nodos como terminados.
-

- **Particionamiento por Circuitos.** Este procedimiento utiliza la división en circuitos mencionada anteriormente. Este criterio garantiza factibilidad, pero genera zonas no disjuntas. Luego de finalizada su ejecución, debemos determinar a qué zona pertenece cada una de las aristas de las intersecciones.
- **Particionamiento por Aristas.** Con este criterio en el que cada arista es un conjunto, se garantiza que las zonas generadas sean disjuntas, pero no se tiene noción sobre su factibilidad. Por este motivo será necesario garantizar la factibilidad mediante el agregado de circuitos opcionales. Para ello se agregan para cada arista sin circuito, en carácter de optativas las aristas pertenecientes al circuito mínimo que la contiene.

Definimos además dos nociones de distancia entre un punto k del espacio y un conjunto no nulo de aristas C , en ambos $dist(x, y)$ definida entre dos puntos es la distancia euclideana.

- **Lejano:** $dist(k, C) = \text{máx}\{dist(k, y) | y \in e \wedge e \in C\}$.
- **Promedio:** $dist(k, C) = \text{prom}\{dist(k, y) | y \in e \wedge e \in C\}$.

Como en cada paso es necesario determinar si un grafo es menor que otro en relación con el problema a resolver se han definido tres criterios. Para las siguientes definiciones se utiliza que C es la solución óptima del CPP en el grafo G , $obl(G)$ y $opt(G)$ son los conjuntos de aristas y arcos obligatorios y optativos respectivamente de G , y $peso(e)$ determina el peso de una arista o arco.

- **CPP (G)** = $F_{obl} \sum_{e \in obl(C)} peso(e) + F_{opt} \sum_{e \in opt(C)} peso(e)$.

Con este criterio, es necesario factibilizar el grafo cada vez que se lo quiera utilizar. Requiere definir dos factores $F_{obl}, F_{opt} \in \mathbb{R}$ que marcan la importancia que se le da a las aristas optativas u obligatorias al determinar el peso del grafo. En este trabajo se deja fijo F_{opt} y se varía solamente F_{obl} .

- **Peso Total (G)** = $F_{obl} \sum_{e \in obl(G)} peso(e) + F_{opt} \sum_{e \in opt(G)} peso(e)$.

Más simple que el criterio anterior, reemplaza resolver el CPP por sumar las aristas del grafo factibilizado.

- **Peso Parcial (G)** = $\sum_{e \in obl(G)} peso(e)$.

Este criterio toma en cuenta solamente las aristas obligatorias. Puede verse como un caso particular de Peso Total en el que $F_{obl} = 0$ y $F_{opt} = 1$, pero se ha considerado aparte porque no requiere que el grafo G sea factible, lo cual acelera su cálculo.

Por último, una heurística determina cuál de esos criterios se aplica en cada paso, es decir para seleccionar el menor subgrafo al cual agregarle el conjunto de aristas más cercano. Las desarrolladas en el presente trabajo han sido las siguientes:

- **CPP/Total/Parcial:** Usar en todos los pasos el mismo criterio según se decida al inicio del algoritmo de generación de las zonas: si las áreas son G_i para $i = 1, \dots, k$, en cada paso el grafo al cual agregar el conjunto de aristas (según el criterio elegido) estará dado por

$$\min_{i=1}^k \text{PesoParcial}(G_i) \quad \text{ó} \quad \min_{i=1}^k \text{PesoTotal}(G_i) \quad \text{ó} \quad \min_{i=1}^k \text{CPP}(G_i).$$

- **CPP-N y Total/Parcial:** Usar el criterio CPP cada N pasos, y en los $N - 1$ intermedios usar Total ó Parcial según se decida al inicio de la generación de áreas. En la iteración j el grafo elegido estará dado por

$$\begin{cases} \min_{i=1}^k \text{CPP}(G_i) & \text{si } j \text{ mód } N = 0 \\ \min_{i=1}^k \text{PesoParcial}(G_i) \quad \text{ó} \quad \min_{i=1}^k \text{PesoTotal}(G_i) & \text{sino.} \end{cases}$$

- **CPP-N, Total/Parcial y Feedback:** Usar el criterio CPP cada N pasos, y en los $N - 1$ pasos intermedios utilizar una función de realimentación basada en los resultados del último paso en el que se resolvió CPP y el resultado de Total o Parcial según se decida al inicio del algoritmo de generación. Sea j un paso de la iteración, $h = j \text{ mód } N$. La selección de la menor área $G_i \in \{G_1, \dots, G_k\}$ estará dada por

$$\min_{i=1}^k \text{CPP}_{feed}(G_i),$$

donde

$$CPP_{feed}(G_i) = \begin{cases} F_m(G_i) & \text{si } h > 0 \\ P_{eso_j}(G_i) & \text{si } h = 0, \end{cases}$$

tal que F_m es una de las 4 funciones de realimentación desarrolladas con el fin de extender el impacto del resultado del cartero chino durante las iteraciones que siguen a su cómputo. La cantidad de estas iteraciones en las cuales se modifica el resultado del criterio común con la aplicación de la función de realimentación dependerá de una estimación de la cantidad de aristas o arcos a agregar para equilibrar todos los resultados del CPP calculado. Asimismo el grado de influencia de dicha función sobre cada área dependerá de la diferencia entre el resultado del criterio común en el comienzo de la realimentación con el resultado del mismo en la iteración actual. Definimos a continuación las funciones F_m .

Sea $n = \frac{j}{N}$. Para la primera función de realimentación, el valor resultante para el paso $n \times N$ debe ser el de resolver el CPP. En los pasos intermedios $nN + h < (n + 1) \times N$ este valor va aumentando junto con la diferencia entre el resultado de la heurística en el paso actual y el resultado en el paso nN . La forma de estimar la cantidad de aristas o arcos a agregar se aclara a continuación.

En primer lugar se intentó precisar de forma experimental cuánto influye una arista o arco del grafo en el recorrido solución del CPP. Se obtuvo como resultado

$$edge_w = \frac{CPP_{result}}{|edges|} \cong 0,7586.$$

Es decir que agregar una arista en promedio aumenta en 0,7586 el valor del resultado del CPP. Luego se buscó determinar la cantidad de aristas que se deben agregar en el paso nN para igualar todos los resultados de resolver el CPP en las áreas, lo cual se representa como:

$$|edgesToAdd|_{nN} = \sum_{x \in CPP_{results_{nN}}} \frac{[\text{máx}(CPP_{results_{nN}})] - x}{edge_w}.$$

Como en cada iteración se agrega una arista al sistema, dicho valor es disminuído en 1 cada paso hasta llegar a 0. Aquí termina la realimentación y se vuelve al criterio intermedio seleccionado. Entonces en los pasos j no múltiplos de N

$$|edgesToAdd|_j = \max(0, |edgesToAdd|_{(j-1)} - 1).$$

Con esos valores, la primer función de realimentación utilizada para calcular el valor resultante en cada zona en la iteración j queda definida como

$$F_1 = \begin{cases} CPP_{nN} + F_{feed} \times (P_{eso_j} - P_{eso_{nN}}) & \text{si } h > 0 \\ P_{eso_j} & \text{si } h = 0, \end{cases}$$

donde

$$\begin{aligned} CPP_{nN} &= CPP \text{ de la iteración } nN, \\ Peso_j &= \text{PesoParcial o } PesoTotal \text{ de la iteración } j, \\ Peso_{nN} &= \text{PesoParcial o } PesoTotal \text{ de la iteración } nN, \\ F_{feed} &= |edgesToAdd|_j. \end{aligned}$$

La segunda fórmula busca generar una combinación lineal entre los resultados de las dos heurísticas utilizadas, de manera tal que a medida que se suceden las iteraciones intermedias, disminuye el peso de la realimentación hasta resolver nuevamente el CPP (los parámetros fueron determinados mediante experimentación)

$$F_2 = \begin{cases} CPP_{nN} \times (F_{feed} - h) + 9 \times F_{feed} \times (Peso_j - Peso_{nN}) & \text{si } h > 0 \\ Peso_j & \text{si } h = 0. \end{cases}$$

La tercera y la cuarta función utilizan la misma idea

$$F_3 = \begin{cases} (CPP_{nN} + Peso_j - Peso_{nN}) \times F_{feed} + h \times Peso_j & \text{si } h > 0 \\ Peso_j & \text{si } h = 0, \end{cases}$$

$$F_4 = \begin{cases} (CPP_{nN} + done_j) \times F_{feed} + h \times Peso_j \times \frac{F_{obl}}{F_{opt}} & \text{si } h > 0 \\ Peso_j & \text{si } h = 0, \end{cases}$$

donde $done_j = (Peso_j - Peso_{nN}) \times \frac{F_{obl}}{F_{opt}}$, es decir cuánto aumentó el área desde la iteración nN hasta la j .

2.4. Factibilización de los subgrafos

Luego de generar las k áreas, éstas deben factibilizarse para cada iteración en la cual se agreguen aristas sueltas. Dada un área, se recorre su conjunto de aristas obligatorias y para cada una de éstas se agregan como optativas todas las aristas de alguno de los circuitos que la contienen. Estos circuitos son precomputados al dividir el grafo simplificado en circuitos en el tercer paso del algoritmo general.

En caso de que se esté utilizando el criterio que agrega circuitos, las zonas al finalizar cada iteración son ya factibles; lo que resta es determinar qué aristas son obligatorias y cuáles optativas en cada una de ellas. En el presente trabajo, por motivos que se detallarán en la sección de experimentos y resultados, no se desarrolló dicho algoritmo diferenciador, por lo que este criterio sólo es posible aplicarlo en los pasos intermedios tomando todas las aristas como obligatorias. Una vez generadas las áreas, para calcular los recorridos se da la opción de determinar manualmente los límites de las zonas marcando optativas y obligatorias.

2.5. Resolución del CPP

Sea el grafo $G = (V, E, A)$ la representación de un área con calles obligatorias y optativas, mano única y doble mano, en la cual E será un conjunto de pares ordenados, a pesar de representar aristas no dirigidas. Sean también las funciones *required* y *optional*, que evaluadas sobre una arista indican obligatoriedad o no de incluirla en el recorrido,

$w : (E \cup A) \rightarrow \mathbb{R}^+$ la función de peso de las aristas de G .

$$OBL_A = \{a \in A : \text{required}(a)\},$$

$$OBL_E = \{e \in E : \text{required}(e)\},$$

$$OBL_{Einv} = \{e_{ij} | e_{ji} \in E \wedge \text{required}(e_{ji})\},$$

$$OPT = \{e \in (A \cup E) : \text{optional}(e)\}.$$

El modelo tendrá una variable entera $obl_{uij} \geq 1$ para cada arista $e_{ij} \in OBL_A$ y una variable entera $opt_{ij} \geq 0$ para cada $e_{ij} \in OPT$, que denotarán la cantidad de apariciones de las aristas correspondientes en el circuito solución. Se contará además con dos variables enteras $obl_{dij}, obl_{dji} \geq 0$ para cada $e_{ij} \in OBL_E$ que representarán la cantidad de apariciones de cada arista obligatoria.

La función objetivo buscará minimizar la longitud de dicho circuito:

$$\sum_{e_{ij} \in OBL_A} w(e_{ij}) \times obl_{uij} + \sum_{e_{ij} \in OBL_E} w(e_{ij}) \times (obl_{dij} + obl_{dji}) + \sum_{e_{ij} \in OPT} w(e_{ij}) \times opt_{ij}.$$

Al ser conexo el conjunto de aristas obligatorias, la forma de obligar a que la solución sea un circuito es forzando la conservación del flujo en cada nodo:

$$\forall i \in V :$$

$$\begin{aligned} \sum_{j: e_{ji} \in OBL_A} obl_{uji} + \sum_{j: e_{ji} \in (OBL_E \cup OBL_{Einv})} obl_{dji} + \sum_{j: e_{ji} \in OPT} opt_{ji} = \\ \sum_{j: e_{ij} \in OBL_A} obl_{uij} + \sum_{j: e_{ij} \in (OBL_E \cup OBL_{Einv})} obl_{dij} + \sum_{j: e_{ij} \in OPT} opt_{ij}. \end{aligned}$$

Se pide además que una calle doble mano sea recorrida al menos una vez en alguno de los dos sentidos:

$$obl_{dij} + obl_{dji} \geq 1 \quad \forall i, j : e_{ij} \in E \wedge \text{required}(e_{ij}) .$$

Como resultado de resolver el problema en cada área se obtiene el conjunto de las aristas y arcos que la componen (obligatorias y optativas) con la cantidad de apariciones de éstas en el circuito solución.

2.6. Obtención de los recorridos

Para obtener los recorridos de las distintas zonas se genera un multigrafo dirigido $R = (V, A)$ (no mixto) en el cual cada arista se repite tantas veces como indica la variante asociada en dicha solución. Se sabe que este grafo

posee un circuito euleriano, se obtiene el mismo con una implementación del algoritmo de Euler. Luego (comparando con el subgrafo de la zona) se marca una de las apariciones de cada arista obligatoria (dirigida o no dirigida) como *de recolección* y el resto como *de paso*.

3. Resultados computacionales

Estudiamos a continuación los experimentos realizados. La implementación se realizó en el lenguaje de programación *Python*, se utilizó IBM CPLEX 12.5 como *solver* de programación lineal entera, con todos sus parámetros en los valores por defecto. Los experimentos se realizaron en una PC Intel Core i5 con 8Gb de memoria RAM en un sistema operativo Linux de 64bits.

3.1. Heurísticas de agregado

Se analizó la posibilidad de hacer crecer las áreas garantizando la factibilidad en cada paso. Para esto se consideró en primer lugar la heurística que agrega en cada iteración un circuito al área más cercana según los criterios de distancia *lejano* y *promedio* de la sección anterior. La generación de los circuitos se hizo mediante el algoritmo presentado en la Sección 2, y se obtuvieron los resultados que resumimos en el Cuadro 1. En el mismo se muestra para los distintos tamaños de grafo, medidos en cantidad de aristas, la cantidad de circuitos generados, el promedio de sus longitudes en kilómetros y en cantidad de aristas, la longitud máxima del mayor de ellos en kilómetros y la mínima en metros, junto con el tiempo consumido en computarlos. La principal desventaja observada en esta heurística es la dificultad para automatizar la delimitación de las zonas una vez generadas. Al no ser disjuntos los circuitos agregados, hay que definir a qué zona pertenecen las aristas que tienen en común, lo cual afecta significativamente los tiempos de recolección. Además, el grado de entrelazamiento de las zonas resulta en aristas obligatorias de una zona encerradas en el interior de otra. Este problema crece en dificultad junto con el número de zonas en contacto.

$\#aristas$	$\#circ.$	$ circuito $ [aristas]	$ circuito $ [km]	$ circuito _{max}$ [km]	$ circuito _{min}$ [mts]	$tiempo$ [min]
1885	2961	3.8888	0.2703	2.4588	0.8909	00:24.69
2077	3261	3.8749	0.2741	3.0915	0.8909	00:34.14
2329	3660	3.8694	0.2756	3.0915	0.8909	00:45.42
2536	3935	3.9532	0.2879	3.2355	0.8909	00:54.64
2743	4189	3.9957	0.2955	3.2355	0.8909	01:06.11
3274	4700	4.2834	0.3264	3.2712	0.8909	01:32.71
6982	12180	3.4756	0.2467	4.7809	3.8970	12:59.65

Cuadro 1: División de grafos en circuitos.

Como la modificación manual de las zonas resulta en un aumento notable de la varianza de sus recorridos, se reemplazará este criterio por el de agregar en cada iteración una arista o arco. Al agregar siempre una arista distinta, las zonas terminan siendo disjuntas, y lo único que resta es factibilizarlas mediante el agregado de los circuitos opcionales, resultando además sus contornos más simples.

Debido a que la finalización del algoritmo está dada por la imposibilidad de agregar más aristas a ninguna zona, el inconveniente ahora radica en la dependencia de la ubicación inicial de los centros de las áreas, ya que lo más probable es que un área se termine encerrando cuando aún quedan aristas por agregar. De este modo la similitud de las áreas generadas, definida como la diferencia de sus resultados finales de resolver el CPP, será inversamente proporcional a la cantidad de aristas sin asignar al momento del encierro de la primer zona.

En el presente análisis, la similitud entre las áreas se analizará hasta el momento en el que una de las áreas no pueda crecer más. Dejamos para un trabajo futuro la búsqueda de la ubicación óptima de los centros.

3.2. Heurísticas base

Para seleccionar el criterio de selección de la menor de las áreas analizamos en primer lugar las tres heurísticas base de manera independiente, esto es CPP, parcial y total.

La heurística CPP requiere resolver el CPP en cada iteración, por lo que fue evaluado únicamente sobre instancias pequeñas.

Las formas de las zonas generadas mediante este método son simples y uniformes, como se observa en la Figura 1 que representa un paso intermedio en la zonificación de 8 zonas nocturnas. En esta instancia, cuando las áreas cubren un promedio de 22 km cada una aproximadamente, el tiempo requerido para resolver el CPP es tan elevado que no es razonable continuar.

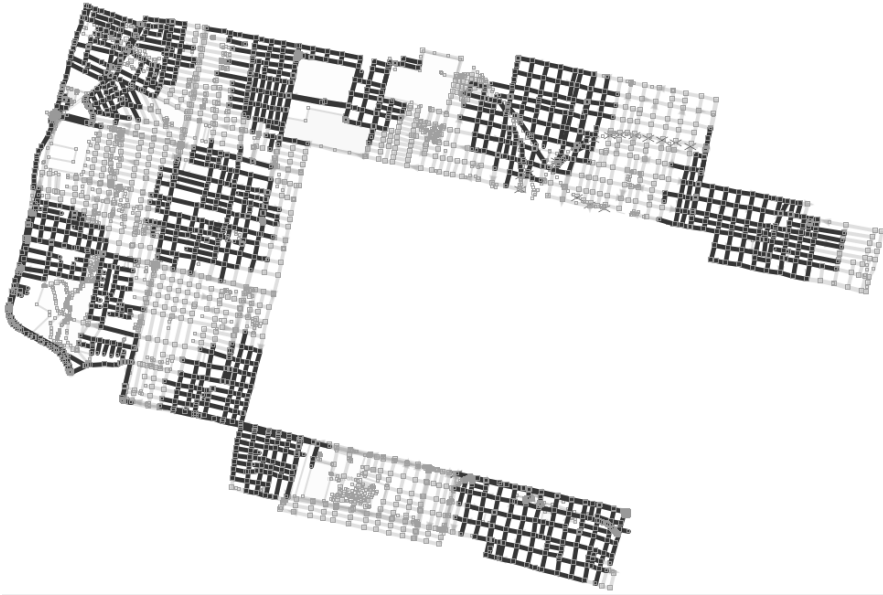


Figura 1: Ejemplo del crecimiento de las 8 zonas (en tono más oscuro) en las cuales está particionándose la región nocturna modificada, utilizando el criterio base CPP.

Los tiempos de ejecución de esta heurística en función de los tamaños de las zonas evaluadas son los representados en la Figura 2; en ella se observa el promedio del tiempo que tarda en resolverse la factibilización, la generación del modelo y el CPP para diversos tamaños de instancia. Estos datos provienen del proceso de zonificación de la región nocturna en 8 zonas. Como se puede ver, el tiempo requerido para resolverla crece exponencialmente con el tamaño de los subgrafos que representan las áreas crecientes, llegando a demorar salvo casos excepcionales, 1 minuto en promedio por cada una cuando éstas llegan a 20 km obligatorios. Nuevamente sin tener en cuenta los casos excepcionales (en los cuales llega a requerir varias horas para resolver una sola iteración), proyectando dicha curva para los tamaños de zona que se desea generar (entre 40 y 60 km) el tiempo consumido es inaceptable, por lo que no es posible utilizar este método en forma exhaustiva.

8 Areas - Tiempo de resolución de la heurística

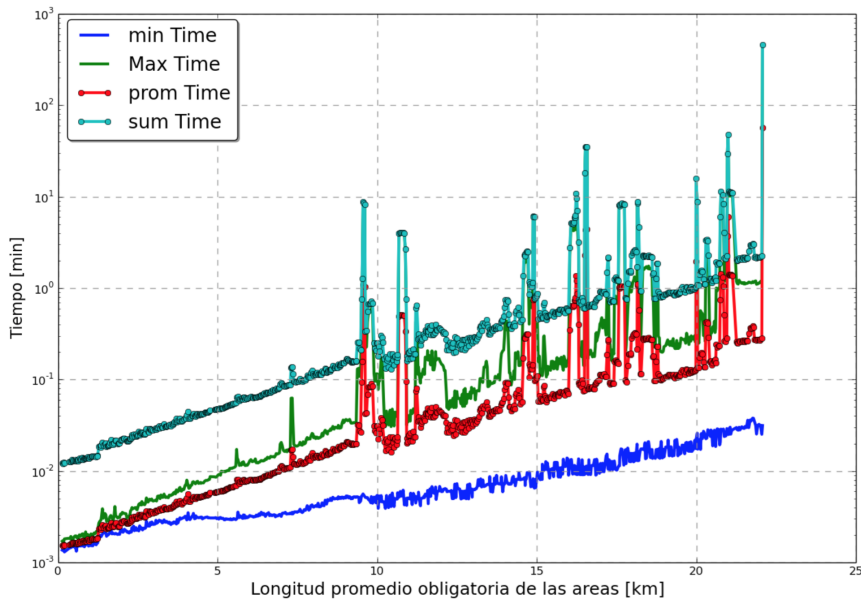


Figura 2: Variación del máximo, mínimo, promedio y suma de los tiempos que demoran en resolverse 8 zonas utilizando el criterio CPP, en función del tamaño de las áreas (medido como el promedio de la suma de los kilómetros de sus aristas obligatorias).

La segunda heurística base denominada *Parcial* se funda en la estrecha relación entre la longitud del recorrido y el peso del grafo analizado. En ella el criterio de selección de la menor de las áreas se basa en la sumas de los pesos de las aristas obligatorias de las mismas, lo cual permite resolver cada iteración sin necesidad de factibilizar el grafo, de generar el modelo ni de resolver el CPP. Aplicado sobre los mismos mapas genera zonas de contornos tan simples y uniformes como la heurística CPP pero, previsiblemente, en un tiempo varios órdenes por debajo de ésta.

En la Figura 3 se observa el costo temporal de una iteración a medida que aumenta el promedio del tamaño de las áreas. Dicha medición está hecha en un total de 8 zonas deteniendo la ejecución al bloquearse una de ellas luego de 20.96 segundos habiendo cubierto 283.68 km obligatorios de la región nocturna. Por un lado se observa el tiempo requerido en promedio, máximo y mínimo para resolver cada una de las áreas, y por otro lado la suma de resolver las 8 áreas.

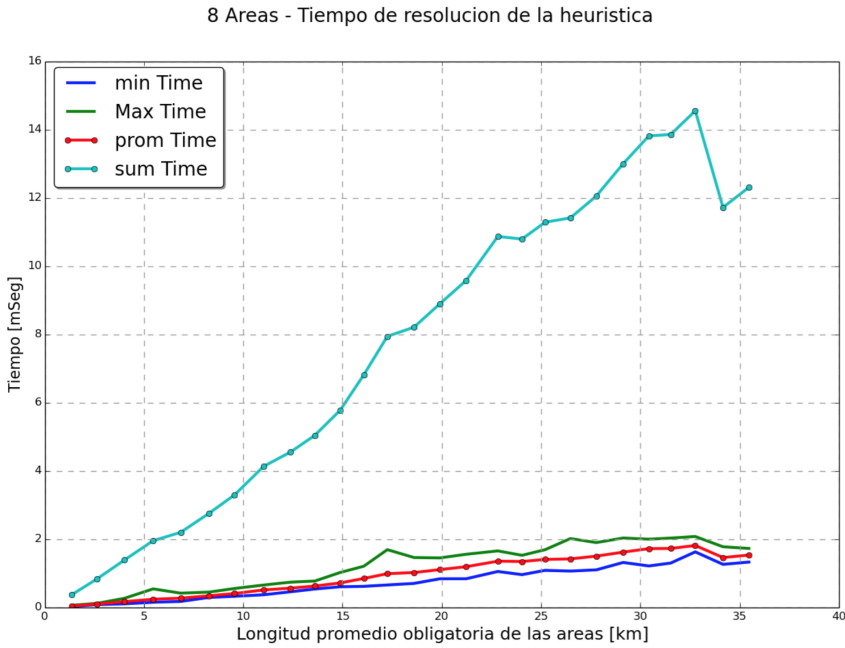


Figura 3: Costo temporal de resolver una iteración con la heurística base Parcial sobre 8 zonas en función del promedio de la suma de los pesos de las aristas obligatorias de las mismas.

La principal desventaja de esta heurística en comparación con la heurística CPP consiste en la disminución de la similitud de las áreas durante y al finalizar su generación. En la Figura 4 se observa cómo varía la máxima diferencia entre los tiempos de recolección para una zonificación de la región nocturna en 8 áreas utilizando la heurística Parcial, en función del promedio de los tamaños de las zonas generadas. La manera de calcular la variación de tiempo, para cada valor del eje x es restarle a las 8 áreas el tiempo consumido por la menor de ellas. Se grafica además el promedio de los valores resultantes y la media. Si bien al final se estabiliza alrededor de los 8 minutos, por momentos llega a superar los 15 minutos. A pesar de que este último resultado es hasta 4 veces peor que el de la anterior, la zonificación es resuelta en un tiempo extremadamente corto.

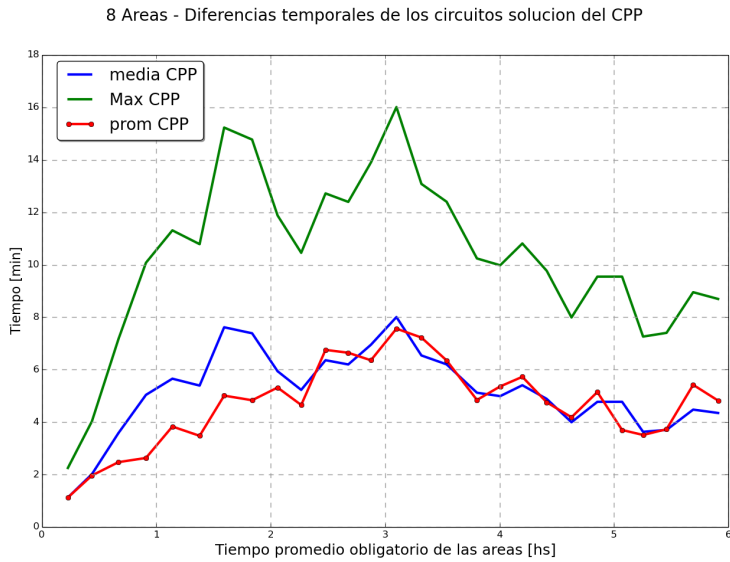


Figura 4: Detalle de la variación de la diferencia máxima en tiempo entre los recorridos de 8 zonas en función del tamaño que van adoptando mediante la inclusión de aristas utilizando el criterio base Parcial.

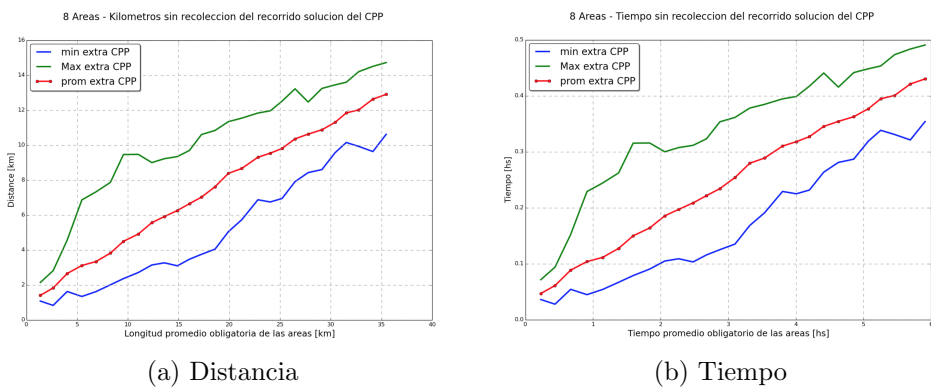


Figura 5: Diferencia de velocidades promedio de los camiones en aristas de recolección o de tránsito.

Se observó en los experimentos la estrecha relación entre el peso obligatorio del grafo (definido como la suma de los pesos de sus aristas) y la longitud del recorrido solución del CPP. La Figura 5 muestra el comportamiento típico observado en todas las instancias analizadas. En la Figura 5b se ilustra que a pesar de que la cantidad de kilómetros de tránsito es aproximadamente la mitad de los kilómetros de recolección, como puede verse en la Figura 5a, el tiempo que requiere recorrer los primeros es poco menos que la doceava parte de lo que demandan los segundos. De aquí se desprende que la diferencia de

velocidades de recolección juega un papel muy importante en la heurística a utilizar.

En la heurística *Total* el criterio utilizado para seleccionar la menor de las áreas tiene en cuenta, además de la suma de los pesos de las aristas y arcos obligatorios, el peso de los circuitos optativos agregados al factibilizar. Como se los recorre a una velocidad varias veces mayor se les asignó a las aristas de estos circuitos un peso menor que a los obligatorios.

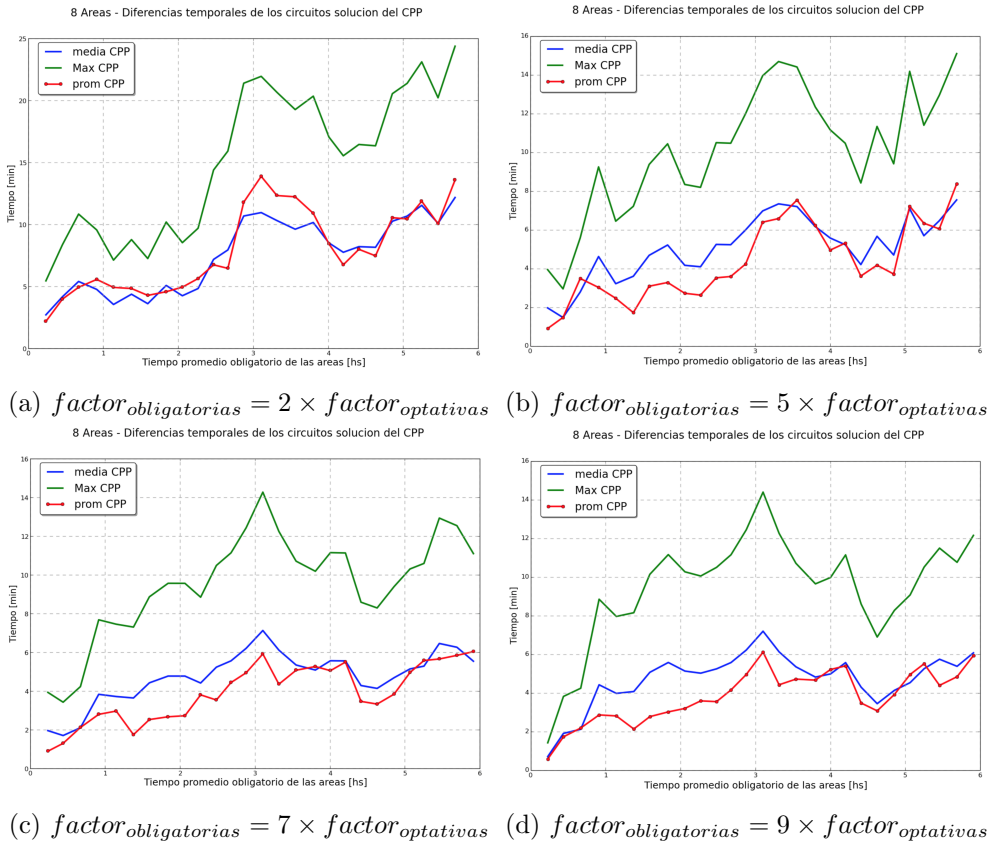


Figura 6: Detalle de la variación de la diferencia máxima en tiempo entre los recorridos de 8 zonas en función del tamaño que van adoptando mediante la inclusión de aristas utilizando el criterio Total con 4 valores distintos de factor.

A pesar de que para el cálculo del tiempo consumido por el recorrido resultante se utiliza un factor fijo de 5, se analizó asignarle distintos valores al mismo dentro de la heurística para ver si de ese modo mejoraba su comportamiento. Como se observa en los gráficos de la Figura 6, asignando menor prioridad al peso de los circuitos optativos mejora levemente la heurística al disminuir la máxima diferencia de los recorridos solución de las áreas en la segunda mitad de su generación, a la vez que empeora la primera. Por lo tanto, cuanto mayor prioridad se asigne a los circuitos optativos, mejores resultados

arrojará la heurística para grafos pequeños. Cuanta menor prioridad se les asigne, se tendrán mejores resultados sobre zonas ya más grandes, obteniendo resultados similares a la heurística anterior.

En lo que respecta a la forma de las áreas resultantes, el resultado fue el mismo que con las dos heurísticas anteriores, se generan contornos simples. La diferencia radica más en los tiempos de resolución que en los resultados. Si bien está dentro de lo razonable en contraposición de la primera (CPP), esta heurística demora bastante más tiempo que la segunda.

El Cuadro 2 sintetiza los tiempos de resolución de las tres heurísticas. La cantidad de aristas y los kilómetros hacen referencia al promedio de las áreas, mientras que los tiempos están dados por la suma de lo que consume la generación de todas ellas.

	zonas	aristas	longitud[km]	Parcial[seg]	Total[hs]	CPP[hs]
pequeñas	8			8,25	0 : 12 : 00,65	23 : 31 : 20
	9			9,21	0 : 16 : 21,98	89 : 57 : 23
	10	270	22,45	12,99	0 : 22 : 17,66	---
	11			14,51	0 : 27 : 39,91	---
	12			20,67	0 : 34 : 53,60	---
nocturnas	12	367	31,12	25,43	1 : 13 : 26,39	---
	11	400	33,92	25,41	1 : 09 : 04,08	---
	10	440	37,31	25,16	1 : 04 : 38,74	---
	9	489	41,47	23,49	0 : 59 : 08,61	---
	8	550	46,64	23,29	0 : 53 : 59,12	---

Cuadro 2: Tiempos de resolución de las tres heurísticas base.

A partir de estos resultados iniciales se exploraron heurísticas que simularan el criterio CPP para conseguir varianzas casi nulas y a la vez que utilizaran una combinación de los otros dos criterios para obtener resultados en corto plazo.

3.3. Heurísticas combinadas

La idea es utilizar alguno de los dos criterios veloces (Parcial ó Total), y cada N iteraciones corregir mediante el criterio CPP. Surgen así dos heurísticas que combinan Parcial con CPP, y Total con CPP, y utilizan las funciones de realimentación de la sección anterior.

Los cuatro gráficos de la Figura 7 muestran el detalle de la variación de la diferencia máxima en tiempo entre los recorridos de 8 zonas en función del tiempo promedio que lleva recorrer sus arcos y aristas obligatorias (a 6 km/h y $F_{obl}/F_{opt} = 5$) a medida que crecen éstas en número con el criterio Parcial utilizando 4 funciones de realimentación para $N = 100$. Fueron tomadas

muestras cada 100 iteraciones. Es importante notar que la cuarta función de realimentación logra una mejora notable bajando 2 minutos en promedio la diferencia máxima entre las áreas a lo largo de su generación.

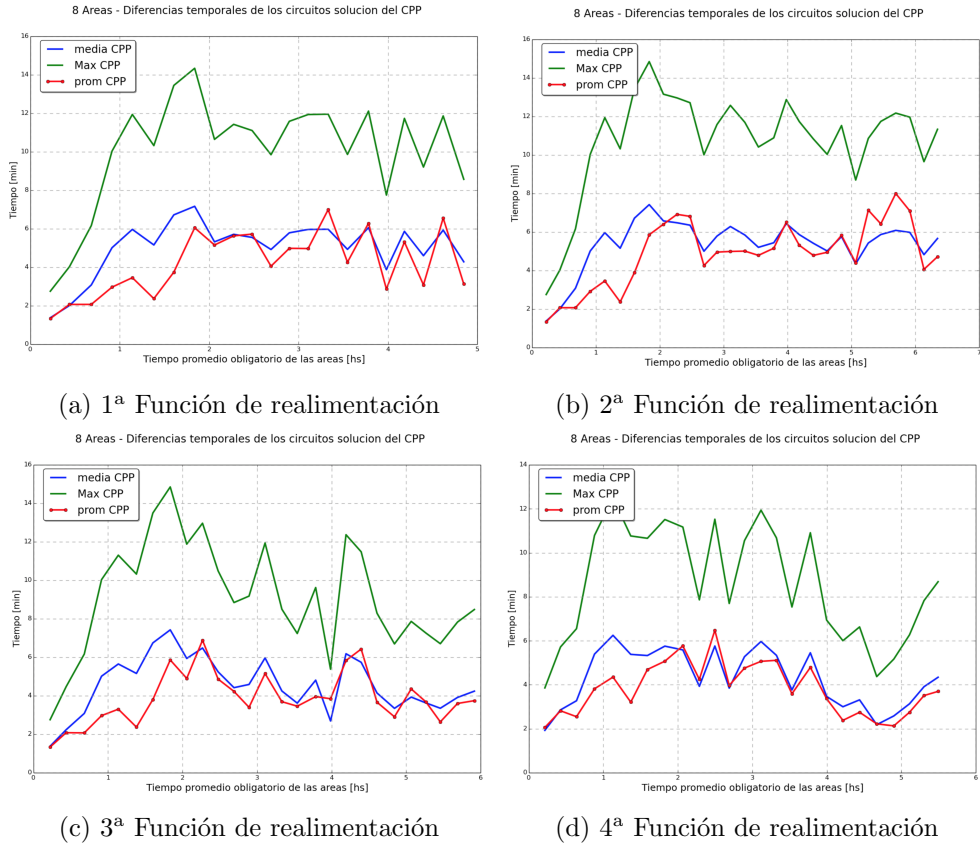


Figura 7: Detalle de la variación de la diferencia máxima en tiempo entre los recorridos de 8 zonas con el criterio Parcial.

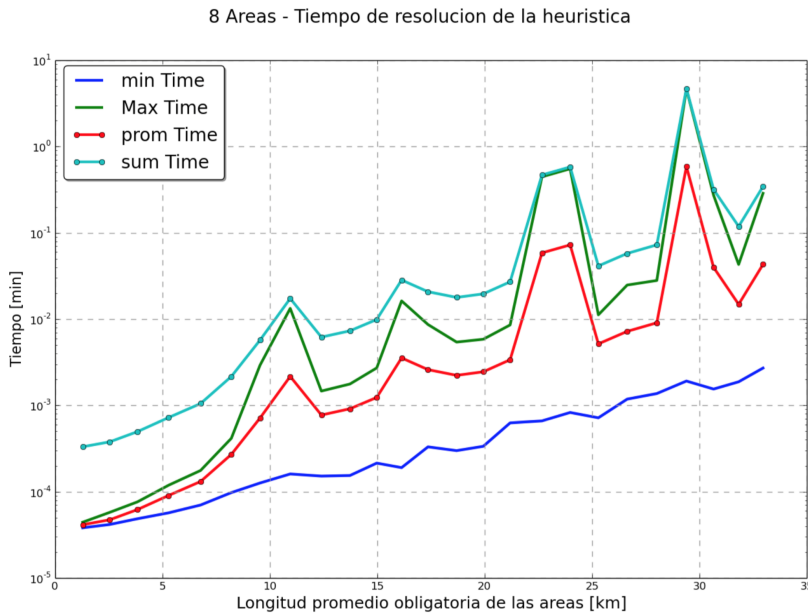
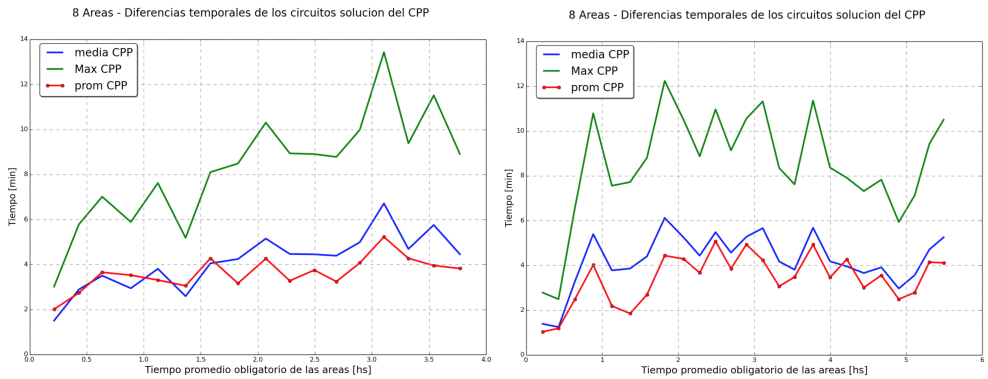


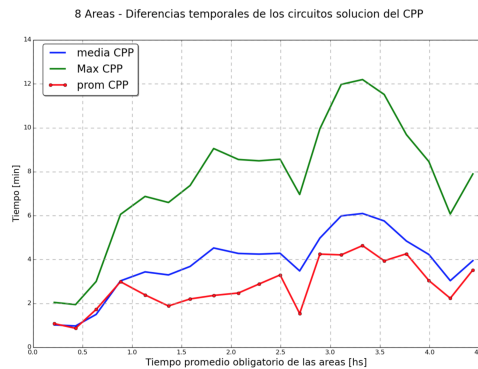
Figura 8: Variación del tiempo promedio requerido para el zonificado a medida que crecen las áreas utilizando la función de realimentación.

Los tiempos de resolución de estas heurísticas con realimentación se ven aumentados respecto de lo que era la heurística Parcial, como se observa en la Figura 8, siendo cada vez más grandes a medida que disminuye el N . Cada valor representa el promedio del tiempo empleado en resolver las 100 últimas iteraciones incluyendo la última en la que se resuelve el CPP. Este gráfico no está orientado a mostrar los valores máximos, sino dar una idea de la variación del tiempo promedio requerido para el zonificado a medida que crecen las áreas utilizando las funciones de realimentación.

Con respecto a las heurísticas con realimentación que utilizan el criterio Total en las iteraciones intermedias, sólo la cuarta obtuvo mejores resultados. Fijando $N = 50$ y variando la relación entre el factor multiplicador de las aristas obligatorias y las optativas se consiguieron las curvas que se muestran en la Figura 9. Se toman muestras cada 100 iteraciones. El tamaño de las zonas está representado por las horas promedio que demora recorrer las aristas y arcos obligatorios a la velocidad de recolección (6 km/h).



(a) $factor_{obligatorias} = 3 \times factor_{optativas}$ (b) $factor_{obligatorias} = 9 \times factor_{optativas}$



(c) $factor_{obligatorias} = 5 \times factor_{optativas}$

Figura 9: Detalle de la variación de la diferencia máxima en tiempo entre los recorridos de 8 zonas en función del tamaño que van adoptando mediante la inclusión de aristas utilizando el criterio Total con una función de realimentación para 3 valores distintos de factor.

Fijando el factor en 5, que es el que en promedio mantiene una diferencia menor, se varió N para intentar obtener mejoras en los resultados, pero sin éxito con valores mayores a 30, al igual que sucedía con la heurística anterior. Recién con $N = 20$ se ha logrado una mejora notable en cuanto a similitud de las áreas, pero con tiempos incluso mayores a los de la anterior heurística, por ende no razonables para los tamaños de zonas que requiere la región de San Miguel de Tucumán.

Los centros de las áreas se ubicaron manualmente en posiciones tales que éstas se encerraran entre ellas lo más cerca de finalizar el algoritmo, logrando un 99% de cobertura para 8 zonas, un 98,45% para 9 zonas, y un 98.71% para 10 zonas utilizando la heurística Parcial. Con esas posiciones de los centros se han realizado 6 últimos experimentos: la zonificación de la región nocturna para 8, 9 y 10 zonas utilizando la heurística Parcial, y luego la heurística Total + CPP + Realimentación con un factor de 5 y $N = 100$. Estos resultados se

resumen en el Cuadro 3.

	Región Nocturna	Parcial	Total + CPP + Feed
8 zonas	Tiempo promedio de los recorridos	07:36:50.19	07:36:50.61
	Diferencia Máxima de tiempos	00:13:03.96	00:12:03.81
	Desviacion estándar tiempos	00:03:56.39	00:04:07.64
	Tiempo Total de Resolución	03:39:08.72	37:25:15.78
	Tiempo Generacion de Areas	00:00:19.26	36:00:24.75
9 zonas	Tiempo promedio de los recorridos	06:47:05.95	06:47:26.78
	Diferencia Máxima de tiempos	00:14:47.84	00:19:21.82
	Desviacion estándar de tiempos	00:05:26.14	00:06:04.90
	Tiempo Total de Resolución	00:41:29.77	11:14:37.74
	Tiempo Generación de Areas	00:00:24.74	10:29:34.58
10 zonas	Tiempo promedio de los recorridos	06:06:50.43	06:06:35.46
	Diferencia Máxima de tiempos	00:19:01.94	00:23:12.60
	Desviacion estándar de tiempos	00:06:24.89	00:06:42.64
	Tiempo Total de Resolución	00:25:46.15	02:39:45.96
	Tiempo Generación de Areas	00:00:39.81	02:30:07.90

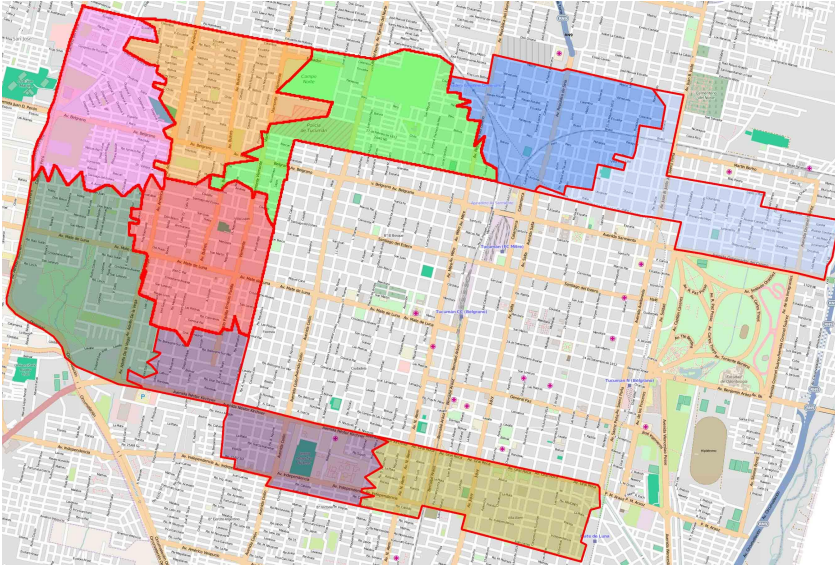
Cuadro 3: Comparación temporal en horas de la zonificación de la región nocturna en 8, 9 y 10 zonas utilizando las dos heurísticas con mejores resultados.

Si bien la zonificación en 8 áreas da como resultado recorridos menores a las 8 horas, al no tener en cuenta factores que pueden hacer demorar más tiempo a los camiones, éstos no parecen viables en la práctica. Por este motivo la zonificación que se utilizará es la que se observa en la Figura 10a en la cual se logra reducir la cantidad de camiones necesarios de 12 a 9, con lo que se contaría con 3 camiones más para cubrir la zona céntrica.

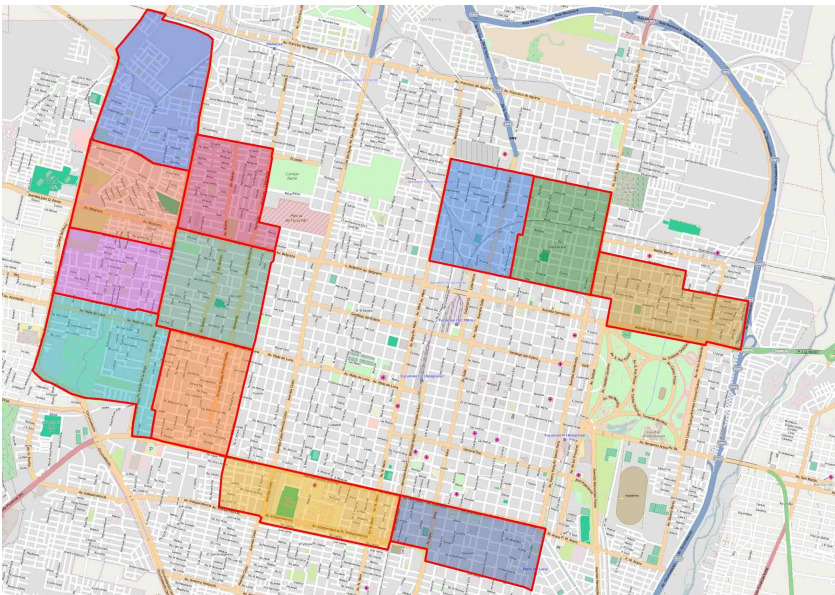
3.4. Calidad de las soluciones

Recordemos que la solución óptima según la definición del problema es una división del grafo en k subgrafos conexos, similares (en el sentido que un camión tarda aproximadamente el mismo tiempo en recolectar los residuos en cada una de ellas), tal que el promedio de los resultados del CPP en todos ellos sea el menor posible, con la restricción adicional de las formas sencillas de las áreas. Si bien no fue posible obtener resultados de *gap* de optimalidad para los experimentos, las soluciones obtenidas se consideran de calidad aceptable para los requerimientos del Municipio de San Miguel de Tucumán, ya que las zonas son lo suficientemente similares: la diferencia máxima entre tiempos de recorrido entre zonas oscila entre el 2.86 % para 8 zonas y el 6.33 % para 10 zonas, y el desvío estándar entre el 0.86 % y el 1.83 %, respectivamente (ver Cuadro 3). Por otro lado, los contornos de las zonas obtenidas resultan más complejos

que en la zonificación actual, pero son todavía aceptablemente sencillos (ver Figuras 10a y 10b). Puede observarse además que la región nocturna es ligeramente más grande en la zonificación propuesta por el algoritmo; incluimos este fragmento adicional en los experimentos (ver zona de color verde claro en la Figura 10a) para dar mayor flexibilidad al algoritmo.



(a) Posible zonificado de la región nocturna en 9 zonas resuelto con el algoritmo desarrollado en el presente trabajo con la heurística total + CPP + feedback con $N = 100$, y un factor de 5.



(b) Zonificado actual de la región nocturna de San Miguel de Tucumán según los datos provistos por el Municipio.

4. Conclusiones

En este trabajo se estudiaron algoritmos heurísticos para resolver el zonificado de la recolección de residuos del municipio de San Miguel de Tucumán. Buscamos minimizar la cantidad de camiones recolectores, y dividir el municipio en zonas similares, definiendo similitud entre zonas como el parecido entre los resultados de resolver el problema del cartero chino mixto en cada uno de los subgrafos asociados.

La idea central de los algoritmos desarrollados consistió en partir de un conjunto de subgrafos nulos del grafo inicial e ir agregando a los mismos conjuntos de aristas hasta formar el particionamiento.

Se desarrollaron dos heurísticas base, que demostraron dar muy buenos resultados en muy corto tiempo. Se intentó mejorar estos algoritmos por medio de funciones de realimentación basadas en la resolución del problema del cartero chino mixto cada cierta cantidad de pasos, y mediante la variación de sus parámetros. Esta estrategia aportó mejoras con respecto a las heurísticas iniciales.

Mediante las heurísticas desarrolladas, se ha logrado cumplir con el requerimiento más importante: la disminución de la cantidad de zonas de la región nocturna. Al reducir de 12 a 9 esta cantidad, se liberan tres camiones para mitigar la problemática del área delimitada por las cuatro avenidas. Con respecto a los requerimientos sobre los contornos de las zonas, fue posible obtener zonas comparables a las utilizadas hasta el momento.

Como trabajo a futuro, podrían desarrollarse heurísticas para mejorar las formas de las zonas generadas que redistribuyan las aristas y arcos de las intersecciones buscando mantener en la medida de lo posible los resultados del CPP.

Adicionalmente, podrían modificarse las posiciones de los centros de las áreas para generar zonas más equilibradas que cubran toda la región sin bloquearse mutuamente en el crecimiento. En este trabajo, los centros de las zonas a generar son asignados en forma manual; este comportamiento podría automatizarse, esto es, iterativamente resolver el zonificado con alguna heurística sencilla e ir variando las posiciones de los centros intentando generar espacio para la primer zona que se bloquea en la ejecución anterior.

Finalmente, en este trabajo no se han tenido en cuenta las reglas de tránsito al momento de armar los recorridos: podrían agregarse restricciones de giros explotando los nodos y conectando sólo las aristas asociadas a las calles que permitieran la circulación entre ellas.

Los resultados de este trabajo fueron entregados al municipio para que se evalúe su implementación en el caso real.

Referencias

- [1] C. Arribas, C. Blazquez, y A. Lamas. Urban solid waste collection systems using mathematical modelling and tools of geographic information systems. *Waste Management & Research*, (28):355–363, 2010.
- [2] F. Bonomo, G. Durán, F. Larumbe, y J. Marengo. A Method for Optimizing Waste Collection Using Mathematical Programming: A Buenos Aires Case Study. *Waste Management & Research*, 30(3):311–324, 2012.
- [3] G. Braier, G. Durán, J. Marengo, y F. Wesner. An integer programming approach to a real-world recyclable waste collection problem in Argentina. *Waste Management and Research*. Aceptado para su publicación.
- [4] V. Bucarey, F. Ordóñez, y E. Bassaletti. Shape and balance in police districting. In *Applications of Location Analysis*, páginas 329–347. Springer International Publishing, 2015.
- [5] N. Chang, H. Lu, y L. Wei. Gis technology for vehicle routing and scheduling in solid waste collection systems. *Journal of Environmental Engineering*, (123):901–933, 1997.
- [6] A. Drexel y K. Haase. Fast approximation methods for sales force deployment. *Management Science*, (45):1307–1323, 1999.
- [7] J. Edmonds. The chinese postman problem. *Operations Research*, (5):88–124, 1965.
- [8] J. Edmonds y E. L. Johnson. Matching, Euler tours and the chinese postman. *Mathematical Programming*, (5):88–124, 1973.
- [9] D. Eisenstein y A. Iyer. Garbage collection in Chicago: a dynamic scheduling model. *Management Science*, (43):922–933, 1997.
- [10] J. Ferland y G. Guénette. Decision support system for a school districting problem. *Operations Research*, 38(6):15–21, 1990.
- [11] The OpenStreetMap Foundation. <http://www.openstreetmap.org>, 2015.
- [12] M. Garey y D. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman and Company, 1979.
- [13] B. Golden, A. Assad, y E. Wasil. Routing vehicles in the real world: applications in the solid waste, beverage, food, dairy, and newspaper industrie. In Toth P. y Vigo D., editors, *The Vehicle Routing Problem*, página 245–286. SIAM, 2002.

- [14] S. Hanafi, A. Freville, y P. Vaca. Municipal solid waste collection: An effective data structure for solving the sectorization problem with local search methods. *INFOR*, (37):236–254, 1999.
- [15] P.E. Hart, N.J. Nilsson, y B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics*, (4):100–107, 1968.
- [16] M. Hojati. Optimal political districting. *Comput Oper Res*, 23(12):1147–1161, 1996.
- [17] J. Kalcsics, S. Nickel, y M. Schröder. Towards a unified territory design approach – Applications, Algorithms and GIS Integration. *Top*, 13(1):1–56, 2005.
- [18] H. C. Kappauf y G. J. Koehler. The mixed postman problem. *Discrete Applied Mathematics*, (1):89–103, 1976.
- [19] B. Kim, S. Kim, y S. Sahoo. Waste collection vehicle routing problem with time windows. *Computers and Operations Research*, (33):3624–3642, 2006.
- [20] J. Liebman. Models in solid waste management. In S. Gass y R. Sisson, editors, *A Guide to Models in Governmental Planning and Operations*. Sauger Books, 1975.
- [21] J. Male y J. Liebman. Districting and routing for solid waste collection. *Journal of the Environmental Engineering Division*, (104):1–14, 1978.
- [22] R. Minciardi, P. Puliafito, y R. Zoppoli. A districting procedure for social organizations. *European Journal of Operational Research*, 8:47–57, 1981.
- [23] M. Mourao y M. Almeida. Lower-bounding and heuristic methods for a refuse collection vehicle routing problem. *European Journal of Operational Research*, (121):420–434, 2000.
- [24] L. Muyldermans, D. Cattrysse, D. van Oudheusden, y T. Lotan. Districting for salt spreading operations. *European Journal of Operational Research*, 139(3):521–532, 2002.
- [25] T. K. Ralphs. On the mixed chinese postman problem. *Operations Research*, (14):123–127, 1993.
- [26] D. West. *Introduction to Graph Theory*. Prentice Hall, 2nd edition, 2000.
- [27] J. Yeomans, G. Huang, y R. Yoogalingam. Combining simulation with evolutionary algorithms for optimal planning under uncertainty: an application to municipal solid waste management planning in the regional municipality of Hamilton-Wentworth. *Journal of Environmental Informatics*, (2):11–30, 2003.

Instituto Sistemas Complejos de Ingeniería

INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA

ISCI
Instituto Sistemas
Complejos de Ingeniería

El Instituto Sistemas Complejos de Ingeniería (ISCI) trabaja en un continuo que va desde el desarrollo de metodologías de frontera hasta aplicaciones innovativas en Sistemas Complejos de Ingeniería, en los cuales interactúan infraestructura, comportamiento humano y gestión.

GRUPOS DE INVESTIGACIÓN

- Recursos Naturales: Minería, Forestal y Acuicultura
- Transporte
- Localización y Logística
- Consumer Analytics
- Smartcities
- Data Science
- Organización Industrial
- Energía

PATROCINA:



INSTITUCIONES ALBERGANTES:



ANEXO DE ESTADÍSTICA
PUNTO MATEMÁTICO
UNIVERSIDAD DE CHILE



INSTITUCIÓN ASOCIADA:



contacto@isci.cl
www.isci.cl

Programas de Postgrado y Postítulos DII

DOCTORADO

Doctorado
en Sistemas de Ingeniería



Sólida formación en herramientas metodológicas necesarias para identificar, analizar, modelar y resolver problemas complejos en sistemas de ingeniería

Contacto:
562-29784017 | doctorado@sistemasdeingenieria.cl
Informaciones y postulación en línea en:
www.dsi.uchile.cl

fcfm FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

MAGÍSTERES

 INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE



MGO | Magíster
Gestión de Operaciones

Formar profesionales de excelencia en investigación de operaciones, quienes podrán enfrentar problemas complejos en gestión de operaciones, integrando herramientas matemáticas, económicas y tecnológicas.

fcfm FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Informaciones en: www.mgo.uchile.cl
Contacto: 562-29784073

Programas de Postgrado y Postítulos DII



MAGCEA
MAGÍSTER EN ECONOMÍA APLICADA

Busca formar profesionales y académicos de gran capacidad analítica y sólida base en economía

www.cea-uchile.cl | infocea@dii.uchile.cl | 562-29784073



MBE
Master in Business Engineering

Magíster en Ingeniería de Negocios
con Tecnologías de la Información

Los líderes de hoy
comprenden cómo la
tecnología lleva a las
empresas al éxito.

A Quién está Dirigido

Ejecutivos y profesionales que deseen liderar o ejecutar proyectos innovadores de diseño integral y sistémico de los negocios orientados a mejorar su competitividad.

Metodología

Este es un Magíster integrador, conformado por un conjunto de cursos de gestión, modelos analíticos aplicados, diseño de negocios, arquitectura y procesos, tecnologías de información de base y diseño de aplicaciones, y de inducción de habilidades de innovación.

Además de las evaluaciones tradicionales por medio de controles y exámenes, una parte fundamental del trabajo de los alumnos será el desarrollo de un proyecto de innovación en el negocio de la empresa auspiciadora -donde ejecutará su residencia-, el cual se llevará a cabo durante todo el programa, en los cursos obligatorios del mismo.

Duración:

3 semestres académicos más un semestre para dar término al Proyecto de Grado.

Horario:

Martes o jueves vespertino, viernes de 14:30 a 18:00 horas
y sábados de 8:30 a 11:45 horas.

Informaciones:

Coordinadora: Ana María Valenzuela
(56 2) 978 4835 / anamaria@dii.uchile.cl

www.dii.uchile.cl



Programas de Postgrado y Postítulos DII



MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS-MGPP



LÍDERES DE EXCELENCIA PARA AMÉRICA LATINA

MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS

ADMISIÓN 2018

Postulaciones en línea en:
www.mgpp.cl

(562) 2978-4067 / (562) 2978-4043
mgpp@dii.uchile.cl

Inicio Horario Diurno

Fines de Mayo de 2018

Inicio Horario Ejecutivo

Julio de 2018



Magíster en Gestión y Políticas Públicas
Acreditado 7 años, desde octubre de 2011
y hasta octubre de 2018



Universidad de Chile - Acreditada
7 años, en todas las Áreas (Docencia de
Pregrado, Docencia de Postgrado,
Investigación, Gestión Institucional,
Vinculación con el Medio), desde diciembre
de 2011 y hasta diciembre 2018

Programas de Postgrado y Postítulos DII



MBA UNIVERSIDAD DE CHILE
INGENIERÍA INDUSTRIAL

PRIMEROS EN GESTIÓN

TALLERES DE HABILIDADES DIRECTIVAS
QUE TE PREPARAN PARA CARGOS DE ALTO LIDERAZGO

ENFOQUE DEL MANAGEMENT SCIENCE PARA
TOMAR MEJORES DECISIONES DE GESTIÓN

MBA EXECUTIVE

24 o 20 meses

Viernes (tarde) y sábado cada 2 semanas

MBA PART TIME

20 o 16 meses

3 a 4 veces a la semana (tardes)

8

ANOS
UNIVERSIDAD DE CHILE
FUNDACIÓN DE 1833
Mayo 2012 a mayo 2020



MBA UChile
Ingeniería Industrial | Universidad de Chile

Admisiones: +56 2 229784002

Programas de Postgrado y Postítulos DII




**Generación
de Redes**



**Carácter
Internacional**



**Cuerpo Docente
de Excelencia**

 +562 2978 4020

 mbamin@dii.uchile.cl

www.mbamin.cl



fcfm

FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Programas de Postgrado y Postítulos DII

EDUCACIÓN EJECUTIVA



DIFERÉNCIATE CON FORMACIÓN DE EXCELENCIA



Diplomados

En diferentes áreas y formatos



Cursos de Especialización

En temas más específicos y actualizados

www.eeuchile.cl

diplomas@dii.uchile.cl | +56 2 2978 4002

