

## INGENIERIA DE SISTEMAS

Volumen XIV

Nº1

Junio 2000

- ❑ Sistema para el Despacho de Vehículos de Emergencia de una Empresa Eléctrica en Chile 3  
*Andrés Weintraub • Julio Aboud • Claudio Fernández • Gilbert Laporte • Enrique Ramírez*
- ❑ Desarrollo de un Sistema Nacional de Información Ambiental: el caso chileno del SINIA 21  
*Samuel Varas • Raúl O’Ryan • M. Paz Soriano • Manuel Díaz • Gonzalo Asencio*
- ❑ Una Aplicación de Prismas de Factibilidad para Resolver el Problema de Asignación de Horarios en una Empresa de Doblajes de Películas 41  
*Carmen Ortiz • Roberto Opazo*
- ❑ Data Mining en la Empresa y en las Finanzas Utilizando Tecnologías Inteligentes 61  
*Richard Weber*
- ❑ Control de la Contaminación Atmosférica de las Fundiciones Estatales de Cobre Mediante un Sistema de Apoyo a las Decisiones 79  
*René Caldentey • Susana Mondschein*

## EDITOR

**Andrés Weintraub P.***Departamento de Ingeniería Industrial,  
Universidad de Chile*

## EDITORES ASOCIADOS

**Hernán Awad • Felipe Caro***Departamento de Ingeniería Industrial,  
Universidad de Chile*

## COMITÉ EDITORIAL

**Sergio Maturana***Universidad Católica de Chile***Lorena Pradenas***Universidad de Concepción***Tomislav Mandakovic***Universidad de Chile***Juan Carlos Sáez***Dolmen Ediciones***Susana Mondschein***Universidad de Chile***Ernesto Santibáñez***Universidad Católica de Valparaíso***Luis Llanos***CMPC Celulosa***Mario Tala***Ministerio de Obras Públicas***Miguel Nussbaum***Universidad Católica de Chile***Jorge Vera***Universidad Católica de Chile***Víctor Parada***Universidad de Santiago de Chile***Jorge Yutronic***KYBER*

---

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

---

## Instrucciones a los autores:

los autores deben enviar 2 copias del manuscrito que desean someter a referato a: Comité Editorial, Revista Ingeniería de Sistemas, Av. República 701, Santiago, Chile. Los manuscritos deben estar impresos en hojas tamaño carta, a doble espacio, deben incluir un resumen de no más de 150 palabras, y su extensión no debe exceder las 30 hojas. Detalles en [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)

---

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

---

Correo electrónico: [ris@dii.uchile.cl](mailto:ris@dii.uchile.cl)

Web URL: [www.dii.uchile.cl/~ris](http://www.dii.uchile.cl/~ris)

---

Representante legal: Pablo Serra Dirección: República 701, Santiago, Chile.

Diseño y Digramación: Magdalena Acevedo

Impresión: Dolmen Ediciones

---

# Sistema para el Despacho de Vehículos de Emergencia de una Empresa Eléctrica en Chile\*

---

**Andrés Weintraub**  
**Julio Aboud**  
**Claudio Fernández**

Departamento de Ingeniería Industrial,  
Universidad de Chile

**Gilbert Laporte**  
École des Hautes Études  
Commerciales de Montréal  
University of Montreal

**Enrique Ramírez**  
Chilectra Metropolitana

---

## Resumen

---

*Las empresas chilenas han empezado sólo recientemente a desarrollar e implementar sistemas computarizados en apoyo de la toma de decisiones relacionadas con la asignación de vehículos y sus itinerarios. Este trabajo presenta el problema de asignación e itinerario de los vehículos de reparación de la División de Servicios de Emergencia de Chilectra S.A. Se desarrolló un sistema computarizado basado en algoritmos heurísticos para el itinerario vehicular considerando las fallas de distintos tipos a las que debe responder Chilectra. Se usó un método de atenuación exponencial para pronosticar la ocurrencia de fallas y un método heurístico para el ruteo de vehículos. Una evaluación del desempeño del sistema indica una mejora de un 16% en la calidad del servicio, medido en términos del tiempo requerido para atender las fallas bajo condiciones normales, y una mejora de 53% en época de invierno.*

---

\* Este trabajo fue financiado en parte por Chilectra S.A. y Fondecyt, Proyecto 1960694 y por el Canadian Natural Sciences and Engineering Research Council, Proyecto OGP0039682.

---

## Introducción

---

A Chilectra le corresponde proporcionar energía eléctrica a la ciudad de Santiago. Su División de Servicios de Emergencia está a cargo de la atención de fallas eléctricas ocurridas en el Área Metropolitana. Existen Unidades de Emergencia separadas y autónomas que responden directamente a alguna de las sucursales de la empresa. El presente artículo se centra en la Unidad de Emergencia correspondiente a la Sucursal Cordillera de Chilectra, que atiende a las comunas de Providencia, Las Condes y Vitacura, tres de las dieciséis municipalidades de Santiago.

Actualmente, el proceso de despacho vehicular se efectúa en forma manual, es decir, existe un operador responsable de programar las actividades y las llamadas de servicio, basándose en su experiencia. Se han propuesto algoritmos heurísticos implementados mediante computadoras a fin de mejorar la tarea de despacho vehicular. La calidad del servicio, medida en términos de tiempo de respuesta, es la preocupación principal de la función de despacho.

A fin de representar todas las posibles demandas en una sola red, el área de la ciudad bajo estudio se representó en base a nodos que representan zonas que comprenden entre dos y cinco cuadras. Cada zona está representada por un nodo que agrupa la información sobre fallas y tiempos de traslado. Los tiempos de traslado entre diferentes nodos se calculan en base a modelos de tiempo de movilización urbana bajo diferentes escenarios de tráfico de acuerdo a la hora del día. Básicamente, una falla puede ocurrir en cualquier lugar, por ejemplo, en una casa o en una calle. Para incorporar esta información a la red definida, se desarrolló un sistema de codificación de direcciones geográficas relacionadas con los nodos de la red.

La construcción del sistema de toma decisiones requiere conocer la probabilidad de que una falla del sistema eléctrico que necesite servicio se produzca dentro de las próximas horas. Un modelo predictivo de demanda fue desarrollado utilizando un enfoque de atenuación exponencial basado en las estadísticas históricas sobre averías.

El método de solución desarrollado se basó en un algoritmo heurístico utilizado para generar ruteos de vehículos, incorporando el conocimiento de las eventuales demandas aleatorias y las prioridades de servicio según la urgencia de la falla, con el fin definir la asignación de tareas a los vehículos.

Por último, para evaluar el sistema propuesto, la calidad de los resultados fue estimada comparando dos enfoques: uno que utiliza el sistema de despacho propuesto, y el otro que utiliza el sistema operacional de la unidad de emergencia actual. Esta comparación se realizó mediante una evaluación de cada enfoque durante el período de una semana. Los resultados de las pruebas muestran una mejora en el tiempo de respuesta de alrededor de un 16% cuando se usa el sistema propuesto. La comparación fue realizada además para días lluviosos, cuando la frecuencia de averías aumenta en forma importante. En este caso el sistema computarizado logró mejorar el tiempo de respuesta en un 53%.

---

## Operación del Sistema de Emergencia de Chilectra S. A.

---

La tarea específica de la Sucursal Cordillera consiste en recibir y clasificar, de acuerdo a su grado de prioridad, las llamadas de servicio que hacen los usuarios del sistema eléctrico ubicados en tres comunas de la Región Metropolitana. Dichas llamadas se verifican posteriormente, en terreno, a fin de restablecer el servicio normal en lo posible. De no ser esto posible, se solicitan los recursos necesarios a la División de Servicios de Emergencia.

La sucursal trabaja en dos turnos de trabajo de ocho horas de duración cada uno, o sea, dieciséis horas por día. Las llamadas recibidas fuera de horas son atendidas por un departamento centralizado que cubre la ciudad completa. La central telefónica principal recibe las llamadas a medida que van llegando, las que son clasificadas según su grado de prioridad (se asigna un código de prioridad a cada una). Un operador telefónico recoge toda la información posible de cada llamada: ubicación geográfica, características de la avería incluyendo el grado de peligro que implica, el área afectada, etc. Se genera una solicitud de servicio que se ingresa en un terminal del sistema informático en línea de la sucursal. Dicha solicitud se envía a los despachadores. En el caso en que la falla es seria, el operador puede, además, llamar al despachador directamente para darle más detalles.

### **El Despachador**

El despachador asigna el trabajo a cada uno de los vehículos. Se utiliza un plano de las áreas asignadas para ubicar la avería. El criterio utilizado para asignar los trabajos consiste en cumplir con los requerimientos de prioridad y tiempo de espera, aprovechando tanto el conocimiento del despachador del área bajo su control, como su experiencia. En este proceso, el despachador hace uso de la siguiente información: la ubicación actual de cada uno de los vehículos, el tiempo necesario para completar todas las llamadas de servicio actualmente en proceso; el nivel de prioridad de los casos que se encuentran en la fila de espera y el tiempo que ha pasado desde que la llamada fue recibida. Toda esta información se encuentra disponible mediante consulta al terminal de computador, o bien, vía llamada radiofónica directa a los vehículos.

La asignación vehicular consiste en seleccionar y llamar por radio a cada vehículo para darle su próximo trabajo. El despachador toma su decisión de tal modo de obtener la mejor relación prioridad-tiempo de espera. Una vez asignada la llamada de servicio a un vehículo específico, la ruta de traslado es determinada por un inspector que acompaña el vehículo. Dicha decisión es tomada en base a su experiencia.

### **Vehículos de Servicio o Unidades de Servicio (US)**

Existen dos turnos de trabajo que funcionan desde las 07:00 hasta las 23:00 hrs. Cada turno funciona con aproximadamente cinco vehículos durante la temporada

baja (los meses de verano) y diez vehículos durante la temporada alta. Generalmente, el número de US se reduce a tres vehículos durante el turno nocturno, el que fundamentalmente atiende a los sectores con mayor número de averías. Los primeros treinta minutos de cada turno de las US se dedican a revisar la unidad, cargar materiales y equipo, etc. Las unidades de servicio deben llegar a la sede treinta minutos antes del fin de cada turno, con el objeto de desembarcar materiales de trabajo y entregar el vehículo al próximo turno. Cada turno dispone de un período de descanso de treinta minutos, el que puede tomarse en cualquier momento. Cuando no se encuentran respondiendo a llamadas, los vehículos de servicio trabajan en mantención preventiva dentro del sistema (Chilectra 1992, 1994).

---

## El Sistema propuesto

---

El siguiente enfoque fue desarrollado para apoyar el despacho de los vehículos de servicio de emergencia. Los siguientes elementos se definen como clave en el proceso de toma de decisiones:

### **Prioridad de Servicio**

Existen cinco niveles de prioridad. Las averías críticas, tales como un cable caído que presenta peligro, tiene Prioridad Uno. La pérdida de energía doméstica en hogares individuales tiene la prioridad mas baja. Las prioridades son de naturaleza dinámica debido a que ninguna solicitud de servicio puede postergarse indefinidamente, de tal modo que una falla de menor prioridad aumenta su prioridad mientras más tiempo esté en espera.

### **Tiempo de Traslado y Costos de Transporte**

Las solicitudes de servicio deben ser atendidas en el menor tiempo posible. Los costos globales de transporte, aún cuando son significativos, tienen menor importancia en este caso.

### **Probabilidad de Nuevas Solicitudes**

Es importante incluir información sobre posibles fallas en el futuro próximo, a fin de evitar la asignación de vehículos de manera tal que un área de probables fallas futuras pudiese quedar desprotegida.

### **Función Objetivo**

El sistema pretende optimizar la calidad de servicio medida en términos de tiempo de respuesta, tomando en cuenta los niveles de prioridad. Para la toma de decisiones

en tiempo real, hay que agregar un factor que considera la calidad de servicio esperada en el futuro. Este factor puede aproximarse considerando la distancia a que se encuentran los vehículos de una zona determinada, y el número de fallas esperadas para la misma.

En forma esquemática, hemos intentado minimizar la función de costos que refleja dichos factores para cada asignación vehicular. El costo se define como

$$\alpha \sum_{i \in I} P_i T_i + \beta \sum_t \sum_{j \in J} F_j K_{jt} \quad (1)$$

donde  $\alpha$  y  $\beta$  son parámetros por calcularse y que, respectivamente, ponderan la calidad de servicio para fallas conocidas versus la calidad de servicio esperada para las averías en el futuro. Si se desea, también puede agregarse una expresión que refleje el costo de traslado. La notación restante se define de la siguiente manera:

$I$  : Conjunto de fallas conocidas

$J$  : Conjunto de zonas

$P_i$  : Ponderación de factor prioridad para la falla  $i$

$T_i$  : Tiempo de servicio, incluyendo el tiempo de demora antes que un vehículo llegue al sitio de avería  $i$

$F_j$  : Demanda esperada (numero de averías) para zona  $j$

$K_{jt}$  : Factor de castigo que depende de la distancia de los vehículos de la zona  $j$  en el período  $t$

La siguiente sección describe cómo se determinaron estos valores. Algunos parámetros se determinaron mediante experimentación, mientras otros se obtuvieron a través de métodos algorítmicos que se detallan a continuación.

### **Ponderaciones en la Función Objetivo**

Los parámetros  $\alpha$  y  $\beta$  ponderan la calidad de servicio para las fallas conocidas y el factor de castigo debido a la mala ubicación de los vehículos en relación a averías futuras. Éste es un problema de objetivos múltiples (en este caso son dos) que es bastante complejo y que ha sido discutido extensamente en la literatura. Existen dos enfoques: el primero considera que la ponderación de los objetivos se define *a priori*. Este enfoque tiene la ventaja de ser de fácil aplicación. Sin embargo, tiene la desventaja de no conducir a una representación precisa en términos metodológicos, dado que las curvas de utilidad de los usuarios son generalmente complejas y no-lineales. El segundo enfoque es interactivo, siendo el usuario quien califica sus opciones de acuerdo a sus preferencias, para llegar de esta manera a una solución. Este enfoque es correcto en términos metodológicos, pero es de difícil aplicación. Nosotros elegimos un sistema sencillo de ponderaciones lineales tales como en la expresión (1). En dicha ecuación, los parámetros  $\alpha$  y  $\beta$  se calibraron mediante un

proceso de simulación, con el fin de obtener soluciones apropiadas para cualquier evento probable, las cuales luego se validaron para alcanzar un nivel de calidad de servicio global alto (Fernández, 1995).

### **Ponderación de las Prioridades**

Los factores de ponderación  $P_i$  asociados a las diferentes fallas de acuerdo a sus prioridades, son definidos por la empresa y presentados en el Cuadro 1.

<b>PRIORIDAD</b>	<b>PONDERACIÓN <math>P_i</math></b>
1	10
2	5
3	3
4	2
5	1

Cuadro 1

### **Factores ponderados $P_i$ por prioridad de avería**

La no-linealidad de la relación demuestra que la característica crítica de las fallas crece con el nivel de prioridad. Las ponderaciones han sido determinadas de acuerdo a una calibración realizada por la Sucursal Cordillera de Chilectra S.A. Los niveles de prioridad también varían según el tiempo de espera. Por ejemplo, una falla de Prioridad 5 se convierte en una de Prioridad 4 después de treinta minutos. El Cuadro 2 presenta las variaciones de prioridad a través del tiempo:

<b>Prioridad inicial</b>	<b>Variación de la prioridad</b>	<b>Tiempo transcurrido después de ocurrida la falla (minutos)</b>
2	1	240
3	2	120
4	3	60
5	4	30

Cuadro 2  
Prioridad de avería a través del tiempo



## **Tiempo de Traslado**

El valor de  $T_i$  ha sido definido como el tiempo requerido para responder a una falla. Dicho tiempo depende de las decisiones de despacho. Para administrar estas decisiones se ha definido un proceso dinámico. En cualquier momento existe un conjunto de fallas conocidas, a las cuales se debe asignar los vehículos. A medida que las averías van siendo atendidas, se las elimina del sistema, incorporándose nuevas fallas a medida que aparecen.

Se desarrollaron dos instrumentos para determinar el despacho de los vehículos: un algoritmo heurístico de despacho vehicular, y un diseño que representa la red de viajes y los tiempos de viaje.

---

## **Algoritmo heurístico para el despacho vehicular**

El primer método construye itinerarios de servicio para cada vehículo a fin de minimizar el tiempo total ponderado (costo) de los traslados. El servicio de las fallas de mayor prioridad se considera en primer lugar ya que las ponderaciones se asignan a las fallas de acuerdo a sus niveles de prioridad. El algoritmo de despacho contempla dos etapas.

### **Construcción de la Solución Inicial**

Para hallar una solución inicial, empleamos un enfoque de agrupar primero las fallas que se asignan a cada vehículo y luego rutear la secuencia en que cada vehículo visita las fallas. Este método tiene la ventaja de ir agrupando las solicitudes de acuerdo al sector geográfico en las zonas asignadas, lo que permite la construcción de itinerarios en el cual cada uno de los vehículos circula dentro de una área limitada, para que los vehículos no tengan que hacer traslados largos entre fallas muy distantes entre sí. Una vez asignados los puntos de demanda que cada vehículo debe atender, se construye un itinerario para cada vehículo. Para definir el itinerario de cada vehículo, se desarrolló un enfoque tipo "GENI" (Gendreau, et al 1992). Este método fue modificado para construir el itinerario tomando en cuenta que las solicitudes de servicio se relacionan con diferentes niveles de prioridad según el grado de peligrosidad y urgencia informados. Cabe notar que también se han propuesto algoritmos heurísticos alternativos para el problema de construir itinerarios dinámicos de vehículos (Powell, 1988). También Le Clerc y Potvin (1997) proponen un algoritmo de tipo heurístico basado en el enfoque genético para un problema dinámico de recoger y entregar material para cada cliente. Para nuestro problema, dado que la calidad de servicio es de naturaleza multi-atributo, se derivaron relaciones ponderadas aditivas, lineales y no-lineales, como funciones objetivo.

El método GENI se basa en la siguiente idea: una ruta se construye gradualmente, agregando un nodo (una falla) a la vez, el cual se selecciona de entre

un conjunto de nodos predefinidos como vecinos de los ya existentes en la ruta. Se inserta cada nodo adicional, probándose todas las combinaciones posibles de secuencias para encontrar la de menor costo. El algoritmo sólo inserta nodos entre los  $p$  vecinos más cercanos en la ruta itinerario parcialmente construida. Si el valor elegido para  $p$  crece, la solución obtenida tiende a ser mejor; sin embargo, el algoritmo requiere de un aumento importante de tiempo computacional.

En los Gráficos 1 y 2 se presentan dos formas de agregar un nodo  $v$  a un itinerario ya existente. Las dos formas involucran una re-optimización local del itinerario actual en torno a la posición donde se inserta el vértice  $v$ . En las inserciones Tipo I, se eliminan los arcos  $(V_i, V_{i+1})$ ,  $(V_j, V_{j+1})$  y  $(V_k, V_{k+1})$  se insertan los arcos  $(V_i, V)$ ,  $(V, V_j)$ ,  $(V_{i+1}, V_k)$ , y  $(V_{j+1}, V_k)$ . Esto significa que las trayectorias  $(V_{i+1}, \dots, V_j)$  y  $(V_{j+1}, \dots, V_k)$  se invierten. Se considera hasta  $p$  vecinos de cada  $v_i, v_j$  y  $v_k$  en la vecindad de  $v$  y se elige la selección que da la ruta de menor costo. Las inserciones Tipo II son algo más complejas. Resulta de la eliminación de los arcos  $(V_i, V_{i+1})$ ,  $(V_{l-1}, V_l)$ ,  $(V_j, V_{j+1})$  y  $(V_{k-1}, V_k)$  y su reemplazo por  $(V_i, V)$ ,  $(V, V_j)$ ,  $(V_l, V_{j+1})$ ,  $(V_{k-1}, V_{l-1})$  y  $(V_{i+1}, V_k)$ . Los caminos  $(V_{i+1}, \dots, V_{l-1})$  y  $(V_j, \dots, V_k)$  se invierten. Aquí se consideran  $O(p^4)$  casos en vez de  $O(p^3)$  como en el primer caso.

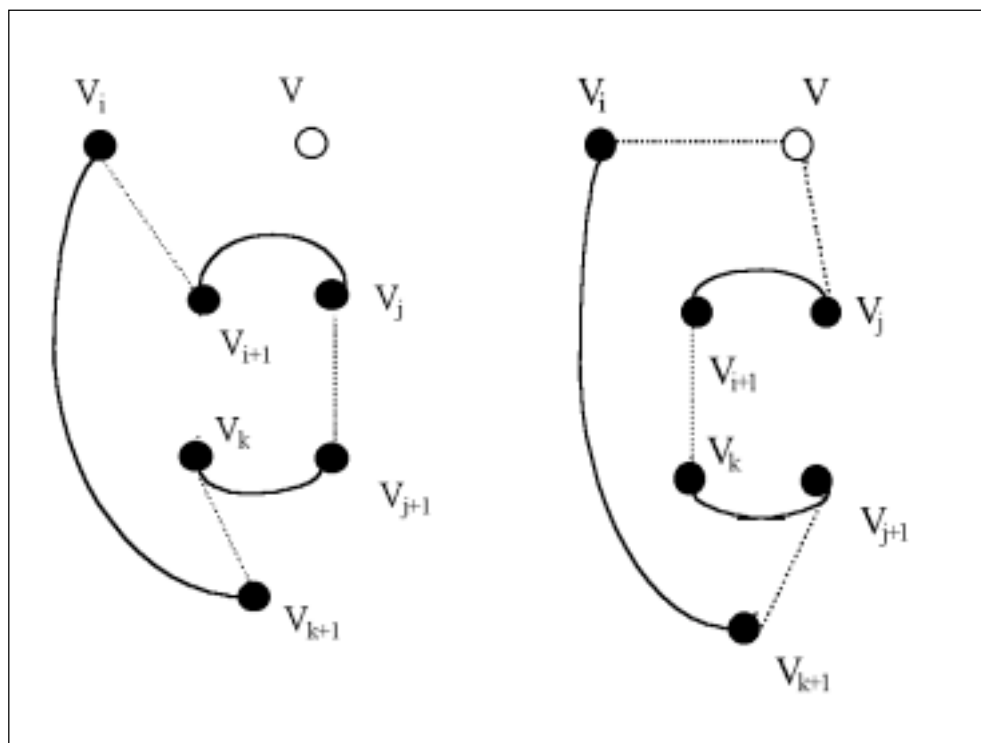


Gráfico 1  
Inserción Tipo I

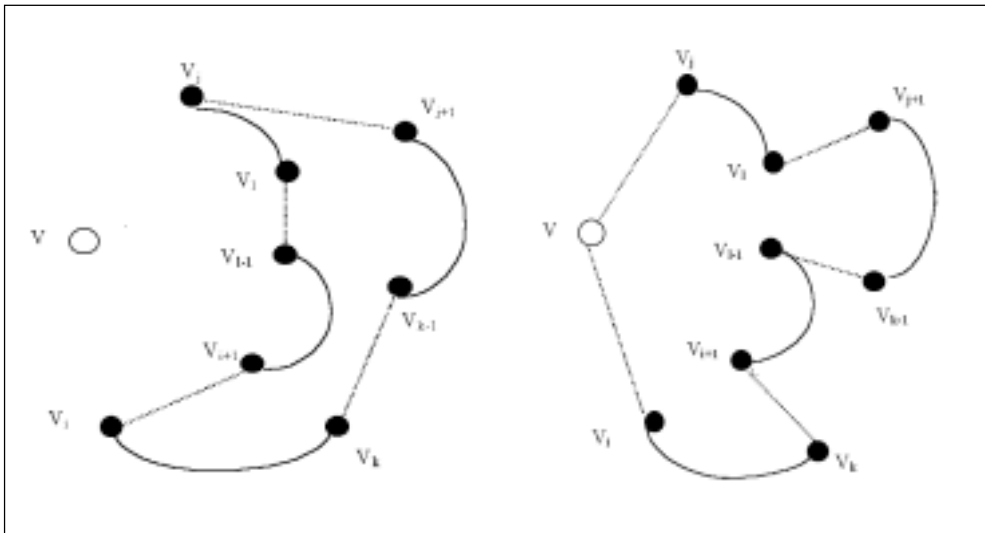


Gráfico 2  
Inserción Tipo II

**El algoritmo puede describirse de la siguiente manera:**

1. Crear un itinerario inicial con tres nodos seleccionados arbitrariamente. Definir una vecindad  $p$  para cada vértice, es decir el conjunto de sus  $p$  vecinos más cercanos.
2. Seleccionar un nodo  $v$  de costo mínimo que todavía no esté dentro del itinerario, considerando las dos direcciones posibles del mismo y los dos tipos de inserción. Actualizar la vecindad  $p$  de cada nodo.
3. Si cada uno de los nodos se encuentra en el itinerario, parar. Si no, volver al Paso 2.

Este algoritmo produce buenos resultados, ocupando poco tiempo computacional y ofreciendo una calidad de solución aceptable en comparación con los otros enfoques. El método se adapta especialmente a situaciones donde las nuevas demandas o vértices aparecen en forma dinámica y deben ser incorporadas a medida que se reciben, lo cual ocurre en el caso descrito pues las nuevas fallas van presentándose una a una en el transcurso del día.

Para este caso específico, se eligió valor de  $p$  de 4 para lograr un equilibrio entre la calidad de solución y el tiempo necesario para su ejecución. Además de la selección de un nodo de costo mínimo (ponderado), se efectuó un cambio para asegurar que todos los nodos (fallas) de mayor prioridad fueran insertados para ser visitados antes de los otros nodos en la ruta construida.

De este modo, el algoritmo considera la prioridad de los nodos así como el tiempo involucrado en atender cada solicitud. Adicionalmente, la función objetivo utilizada, descrita anteriormente, combina la minimización del tiempo de atención a las fallas (modificado de acuerdo a las prioridades) y la minimización del riesgo de dejar sin cobertura a zonas con niveles altos de solicitud de servicio de emergencia.

## **Obtención de la solución definitiva**

Una vez construidas las rutas iniciales, se busca mejorar la solución mediante un enfoque de equilibrio de carga. Este algoritmo de post-optimización empieza a operar una vez terminado el proceso de optimización GENI. La idea es evaluar la carga de cada vehículo y realizar una redistribución de la carga de trabajo entre los diversos vehículos. La carga de trabajo se mide en términos de la cantidad de solicitudes de servicio, así como los tiempos de traslado y de servicio del vehículo. Una vez traspasado el orden de servicio, la función objetivo se evalúa comparándose la solución antigua con la nueva. Este proceso se realiza en forma aleatoria para diferentes transferencias entre las cargas más pesadas y las más livianas. El proceso de post-optimización se realiza para un número determinado de iteraciones, o bien hasta que no se detecte mejora adicional. A fin de estimar la duración de las llamadas de servicio, es decir el tiempo que el vehículo ocupa en cada falla, se utilizaron los tiempos promedio de la empresa correspondientes a diferentes fallas. La llamada promedio dura alrededor de 30 minutos.

---

## Modelación de la red de transporte

---

El Servicio de Emergencia de la Sucursal Cordillera provee cobertura geográfica para tres comunas del área Metropolitana. Se construyó una red para el sector, representando el transporte urbano dentro del sector, de tal manera que los costos (tiempos) relacionados con los arcos de la red permitan una simulación aproximada de los tiempos de viaje entre cualquier par de puntos dentro de las zonas mencionadas.

Para la red se utilizó información proveniente del Modelo ESTRAUS (Sigdo-Coppers y CIS, 1989). Este modelo fue diseñado por la Subsecretaría de Transportes para representar las pautas de tráfico urbano en Santiago. El modelo estima los tiempos de viaje entre diferentes puntos de la ciudad, utilizando una representación de la ciudad compacta y aproximada. Se definen nodos o centroides como puntos que representan las zonas, y que generan o atraen viajes. Las intersecciones de las calles más importantes también se representan mediante nodos. Los arcos de la red conectan estos nodos entre sí, representando los corredores principales o uniones virtuales entre los centroides. La red que representa las tres comunas de este estudio contienen 370 nodos y 1200 arcos.

Se han efectuado estudios de simulación basados en modelos matemáticos dirigidos a evaluar los equilibrios de tráfico bajo diferentes escenarios. Los escenarios corresponden a diferentes situaciones de congestión, horas punta en la mañana y en la tarde, días laborales cuando hay mayor congestión y cuando los tiempos de viaje son más largos, y horas bajas y días festivos en que se contempla un menor grado de congestión. Dado que los datos ESTRAUS son estimaciones, se les considera adecuados como una primera aproximación de los tiempos de viaje. Los tiempos estimados pueden ser corregidos con tiempos reales medidos por los traslados efectuados en forma diaria. Para determinar el tiempo de viaje entre cada par de nodos dentro de la red, se empleó un algoritmo de camino más corto. Este algoritmo

se corre sólo una vez para cada escenario, almacenándose los tiempos de viaje para cada par de nodos. Estos tiempos se cambian únicamente en el caso de alteraciones en tiempo de traslado.

## Factor de castigo debido a ubicación de vehículos

Tal como se mencionó anteriormente, uno de los factores importantes en la toma de decisiones es la ubicación adecuada de los vehículos, dada la posibilidad de que se produzcan llamadas de servicio en el transcurso del día. A fin de determinar este factor, se requiere de algún tipo de conocimiento predictivo sobre cuándo y dónde ocurrirían las eventuales fallas. Existen estadísticas mensuales de las llamadas de servicio efectivamente atendidas en el Área Metropolitana para las diecinueve subzonas definidas en el presente estudio —un total del 14% de la ciudad de Santiago. En el Gráfico 3 se muestran las solicitudes mensuales correspondientes al período entre los años 1991 y 1994 inclusive. Nótese el efecto estacional que produce un aumento de las fallas durante el invierno debido a factores climáticos: la lluvia y el viento tienden a causar mayores fallas. Por otra parte, la baja producida durante los meses de verano se relaciona también con la reducción de la población de la ciudad. Una tendencia posible ascendente de fallas entre 1991 y 1993 relacionada con el crecimiento demográfico y la densidad habitacional se contrapone a mejoras efectuadas en la red eléctrica. Estos dos factores fluctúan en forma paulatina en los pronósticos diarios. Al examinar un mes típico, no se detectaron ciclos importantes relacionados con variaciones especiales de acuerdo al día o a la semana del mes. A nivel diario se nota una disminución de la demanda de servicio entre las 23:00 y las 07:00 horas, y un aumento entre las 21:00 y las 23:00 durante un día promedio en verano, tal como se muestra en el Gráfico 4.

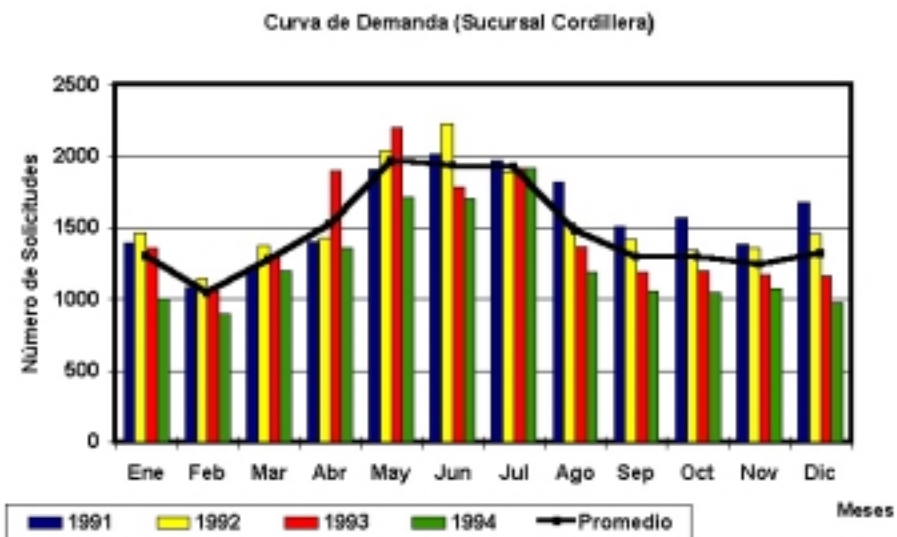


Gráfico 3. Solicitudes mensuales años 1991-1994

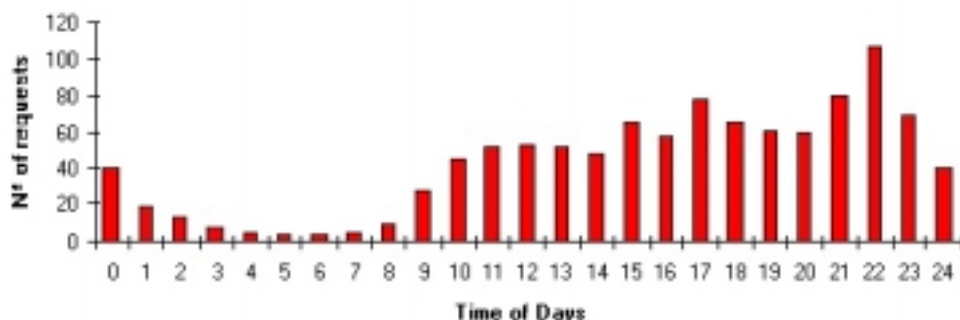


Gráfico 4: Número de solicitudes de emergencia en el transcurso del día

### Modelo de Atenuación Exponencial

Se desarrolló un modelo de atenuación exponencial para pronosticar las demandas diarias en cada zona. Este tipo de modelo supone que la demanda futura se puede predecir sobre la base de información anterior o histórica (Schroeder, 1994). De esta manera, la estimación de demanda para el día  $t$  se mediría como una combinación de la demanda prevista del día anterior ( $t-1$ ), que mide la historia de la demanda anterior y la demanda medida para el día  $t$ . Estos parámetros se ponderan por  $(1-\alpha)$  y  $\alpha$ , respectivamente, dependiendo de la importancia relativa otorgada a la historia pasada en relación al día  $t$ . Adicionalmente, se agregaron parámetros que incorporan los efectos de estacionalidad y tendencia.

Basándonos en los datos, hemos supuesto que las diecinueve sub-zonas son homogéneas con respecto a las funciones de estimación de la demanda. Para cada sub-zona  $j$ , la demanda diaria para el día  $t$  se estimó o según a las expresiones presentadas en el Apéndice.

Este modelo de pronóstico de demanda se corre diariamente para cada sub-zona, utilizando la información existente respecto de las fallas ocurridas.

---

## Implementación del modelo

---

El modelo está diseñado para ser implementado frecuentemente, cada vez que aparece una nueva solicitud de alta prioridad, o bien cuando aparecen varias solicitudes de baja prioridad. Esto correr el sistema aproximadamente cada quince minutos. Los pasos básicos del sistema son los siguientes:

### **1. Leer y preparar los datos al comienzo del día o al iniciarse el turno.**

En este paso se leen los archivos para recoger todos los datos necesarios para la operación del modelo. Estos datos incluyen todas las llamadas en la cola de espera esperando ser atendidos, junto con su ubicación y prioridad, además de la ubicación y las características de cada vehículo. Se efectúa una estimación de la demanda de cada zona basada en el proceso de atenuación exponencial ya descrito.

### **2. Incorporación de nuevas llamadas.**

Cuando se incorporan nuevas llamadas, el sistema requiere de información relacionada con las mismas, tales como ubicación, prioridad, etc. Una vez ingresada al sistema toda la información correspondiente a las nuevas llamadas, el sistema incorpora las nuevas solicitudes y las agrega a aquéllas ya existentes en la cola. Nótese que las prioridades pueden traspasar las llamadas que vienen llegando al primer lugar de la cola. Al comienzo del día o del turno, las llamadas pendientes del día o del turno anterior son atendidas primero. Durante la operación diaria, las nuevas llamadas se incorporan para actualizar la base de datos que se utilizará en los pasos siguientes. Por otra parte, la tarea de actualizar la base de datos incluye además la eliminación de las tareas ya completadas.

### **3. Asignación preliminar de llamadas a los vehículos.**

En este paso todas las llamadas del sistema se agrupan geográficamente de acuerdo a su proximidad, asignándoseles al vehículo más cercano, tomando en cuenta el número de vehículos disponibles y la cantidad de llamadas, de modo que la asignación se equilibra con respecto al número de llamadas por vehículo. En la primera asignación, hay que asegurarse que cada una de las llamadas sea asignada de acuerdo a su prioridad.

### **4. Construcción de rutas iniciales.**

La solución inicial al problema se construye en esta etapa, comenzando con la construcción de rutas iniciales para cada vehículo mediante el método GENI descrito anteriormente.

### **5. Solución final.**

La asignación final se obtiene en esta etapa, la que distribuye la carga uniformemente entre cada vehículo utilizando el enfoque de “equilibrio de carga” descrito anteriormente.

## **Programación de Itinerarios**

El sistema toma en cuenta todas las demandas conocidas en cada momento. Esto normalmente significa definir un horizonte de planificación de varias horas. Sin embargo, en un enfoque de horizonte movible, se actúa únicamente sobre las primeras decisiones. La consideración de todas las demandas conocidas ayuda a evitar una toma de decisiones de carácter miope.

## **Implementación del sistema**

Para evaluar el posible impacto de la instalación de este sistema en Chilectra, la empresa accedió a su implementación durante un corto período en Noviembre de 1994. La prueba se llevó a cabo durante dos semanas consecutivas con patrones de demanda similares. Durante la primera semana se utilizó el sistema tradicional manual. Durante la segunda semana, el sistema de despacho descrito arriba fue utilizado para asignar los vehículos en ambos turnos de unidades de emergencia. Hubo algunas dificultades en la implementación del sistema, ya que ésta coincidió con un cambio en el modo de funcionamiento del sistema de datos en línea para las llamadas recibidas referentes a fallas. Esta fue una de las razones por las que se empleó un período de prueba relativamente corto. En el Cuadro 3 se compara la duración promedio ponderado de las llamadas de servicio utilizando ambos métodos operacionales.

	<b>Operación manual (minutos)</b>	<b>Sistema propuesto (minutos)</b>
<b>Tiempo de servicio promedio ponderado</b>	31,2	27,0
<b>Desviación estándar</b>	18,0	13,2

**Cuadro 3**

**Tiempo de servicio promedio ponderado por prioridad del sistema manual y del sistema propuesto**

Se produjo una mejora de un 16% en el tiempo promedio ponderado. La desviación estándar de la simulación fue de aproximadamente el 50% de los valores medios. Los datos del Cuadro 3 se refieren a un día normal. Adicionalmente, se realizó una simulación para un día de alta frecuencia de falla, lo que es típico en días de mal tiempo en invierno (Junio-Agosto). La carga de un día normal constituye alrededor de 60 llamadas por día, incluyendo cinco llamadas de Prioridad 1. Durante un día de alto volumen, el número de llamadas supera las 150, de las cuales 15 son de Prioridad 1. La simulación (Villena, 1996) compara los resultados de la programación manual con el sistema descrito en el Cuadro 4 correspondiente a días lluviosos.



Indicador	Tiempo de (minutos)	
	Operación manual	Sistema propuesto
Duración promedio de trabajo	135	50
Tiempo promedio de traslado	13	19
Tiempo promedio total	148	69
Desviación estándar	144	30

Cuadro 4

#### Comparación de Tiempo de Servicio para Días de Emergencia (Días Lluviosos)

Tal como se observa en el Cuadro 4, el tiempo real ocupado en desplazarse es menor en el caso del sistema de asignación manual; sin embargo, la duración total de trabajo es menor en el caso del sistema computarizado, ya que el tiempo ocupado en busca de un vehículo disponible se ve reducido en forma importante. Esto puede explicarse por el hecho de que el sistema computarizado propuesto asigna los itinerarios considerando los tiempos globales y considerando posibles fallas futuras.

---

## Conclusión

---

Tal como se resume en la sección anterior, el nuevo sistema desarrollado muestra ser capaz de mejorar significativamente el tiempo de respuesta a las fallas, especialmente en situaciones de alta demanda, típicos de días lluviosos. Esta ventaja podría mejorarse aún más en vista de los problemas con la implementación del sistema computarizado enfrentados durante la primera semana de su operación. Dichos problemas estarían solucionados una vez instalado el sistema en forma permanente.

Por otra parte, puede mejorarse el sistema de información. Actualmente la información se ingresa al sistema manualmente. La información era manejada por tres personas antes de llegar a los vehículos de emergencia: del Cliente al Operador de la Central, del Operador de la Central al Despachador, del Despachador al Vehículo de Emergencia y del Vehículo de Emergencia al Cliente. El tiempo total de servicio dedicado al cliente involucra las siguientes tareas: asignación de tarea, traslado y reparación. El mejoramiento del proceso de información se refiere a la fase de asignación. El ingreso manual de la información frena el proceso notablemente. La asignación de ubicaciones geográficas a los nodos de la red también es un paso manual que produce una demora de aproximadamente cinco minutos en cada operación.

Eventuales mejoras en los procedimientos de recaudación de datos reducirían estos tiempos. El ingreso directo de datos, con el apoyo de un Sistema de Información Geográfica (SIG) permitiría el procesamiento automático de información, la que se alimentaría en forma directa al Sistema de Despacho, llevando así a una disminución importante del tiempo de cada etapa. En la actualidad, Chilectra ya ha construido una primera etapa de SIG para apoyar el proceso. La automatización del sistema de despacho permitiría que las órdenes de despacho pasen de los operadores de la central a un sistema automático, el cual generaría órdenes de despacho a los vehículos en forma continua y en línea.

El uso de Sistemas de Posicionamiento Global (GPS) instalados en cada vehículo, que indican su posición en todo momento, también aumentarían la automatización del proceso. De esta forma el contar con apoyos de información y comunicación en línea, junto con el sistema de asignación de itinerario computarizado descrito anteriormente, mejoraría notablemente la calidad de los sistemas de despacho de emergencia.

---

## Referencias Bibliográficas

---

- Chilectra S.A., 1992. «Informes de Gestión y Mantenimiento Correctiva», Gerencia de Explotación, Departamento de Servicios de Emergencia.
- Chilectra S.A., 1994. «Solicitudes de atención por fallas, provocadas en la red eléctrica para las áreas 5 y 6», Subdivisión Sistemas de Telecomunicaciones.
- Fernández, C., 1995 Desarrollo e implementación de un sistema de apoyo para el despacho de vehículos de emergencia en Chilectra S.A. basado en algoritmos heurísticos. Tesis de Grado de Magister en Ingeniería Industrial, Universidad de Chile.
- Gendreau, M., A. Hertz y G. Laporte 1992. New Insertion and Postoptimization Procedures for the Traveling Salesman Problem. *Operations Research*, 40, 1086-1094.
- Powell, W.B. 1988 A Comparative Review of Alternative Algorithms for the Dynamic Vehicle Allocation Problem. En *Vehicle Routing: Methods and Studies*, Golden and Assad eds, pp. 249-291. North Holland. Amsterdam.
- Le Clerc F. y J.Y. Potvin, 1997. Genetic Algorithms for Vehicle Dispatching. *International Transactions in Operations Research*, 5/6 pp. 391-400.
- Consortio Sigdo-Koppers Y Cis Ingenieros Consultores, 1989. Estudio de Evaluación y Desarrollo del Sistema de Transporte Urbano de la Ciudad de Santiago, Informe final.
- Schroeder, R., 1994. *Operations Management*. McGraw-Hill.
- Villena, O. 1996 Evaluación de Alternativas para el problema de Asignación en los Servicios de Emergencia y Mantenimiento. Memoria del Departamento de Ingeniería Industrial, Universidad de Chile.

## Apéndice:

---

La demanda diaria proyectada para el día t en una sub-zona dada se calcula mediante las siguientes expresiones:

$$A_t = \alpha D_t + (1 - \alpha) (A_{t-1} + T_{t-1})$$

donde

$D_t$  = Demanda real observada del día t

$\alpha$  = Factor de ajuste (*smoothing*)

$A_{t-1}$  = Demanda histórica proyectada a t-1

$T_{t-1}$  = Tendencia del día anterior

Para actualizar la tendencia definimos:

$$T_t = \beta (A_t - A_{t-1}) + (1 - \beta) T_{t-1}$$

donde

$\beta$  = Factor de ajuste (*smoothing*)

Para actualizar el coeficiente de estacionalidad, definimos:

$$R_t = \gamma (D_t / A_t) + (1 - \gamma) R_{t-L}$$

donde

$D_t$  = Demanda real observada del mes t

$A_t$  = Demanda histórica promedio del mes t

$R_t$  = Factor de estacionalidad (L, largo del ciclo, típicamente un año)

$\gamma$  = Factor de ajuste (*smoothing*)

Por último, la demanda proyectada de averías para cada periodo por zona se define como:

$$F_t = (A_t + T_t) R_t$$

donde

$A_t$  = Promedio estimado por t (definido por (2))

$T_t$  = Tendencia durante periodo t (definido por (3))

$R_t$  = Factor de estacionalidad para el mes t (definido por (4))

Los factores de ajuste (*smoothing*) fueron  $\alpha = 0,3$ , correspondiente a la demanda promedio,  $\beta = 0,2$  a la tendencia y  $\gamma = 0,1$  a la estacionalidad. Estos coeficientes demuestran que los eventos recientes tienen un impacto fuerte sobre las predicciones.



---

# Desarrollo de un Sistema Nacional de Información Ambiental: El caso Chileno del SINIA

---

**Samuel Varas\*,  
Raúl O'Ryan,  
M. Paz Soriano,  
Manuel Díaz y  
Gonzalo Asencio**

Departamento de Ingeniería Industrial  
Universidad de Chile

---

## Resumen

---

*En Chile existe una gran cantidad de información medioambiental diseminada entre los numerosos agentes públicos y privados. Esta información presenta un bajo nivel de coordinación en términos de esfuerzo de generación y mantención, y como resultado, la información medioambiental resulta ser incompleta y heterogénea. Por otro lado, los requerimientos de dicha información son tan variados, como tipos de usuarios, y su proceso de recolección es, generalmente, ad-hoc, tedioso y de alto costo. Esta situación entre proveedores y usuarios de información genera altos costos de búsqueda y recopilación de información, baja motivación para compartir y homogeneizar información existente, y pocos incentivos para generar información faltante. En los últimos años, con el desarrollo de Internet, ha surgido una plataforma tecnológica que permite reducir costos de transacción e integrar proveedores y usuarios distribuidos geográfica y culturalmente. Sin embargo, es necesario desarrollar todo un conjunto de mecanismos y aplicaciones que permitan construir una correcta integración entre proveedores y usuarios de información. Este paper aborda el tema de integración e intermediación de la información medioambiental chilena. Se presenta el diseño y desarrollo de un sistema nacional de información ambiental, desde un punto de vista práctico para que otros esfuerzos similares tomen en cuenta la experiencia obtenida en este trabajo.*

---

\* Contacto con el autor svaras@dii.uchile.cl Dirección: Av. República 701, casilla 2777, Santiago de Chile.

---

# 1 Introducción

---

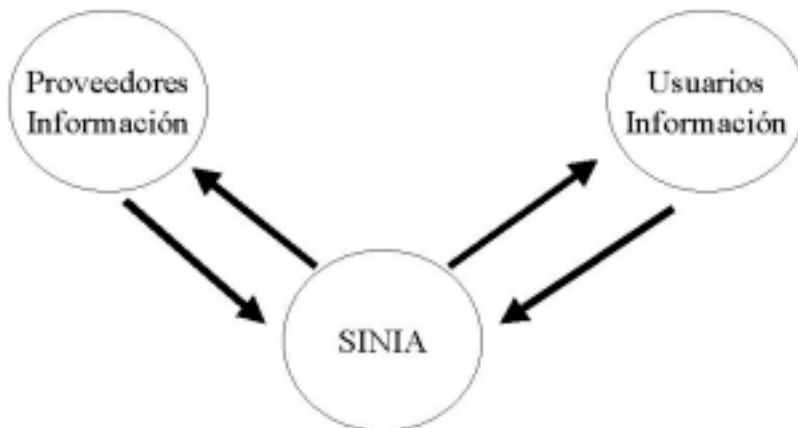
La información medioambiental en Chile se encuentra distribuida entre un gran número de instituciones públicas y privadas, universidades, bibliotecas, centros de estudios, etc. Esta característica dificulta su uso oportuno, y aumenta los costos de accederla. En efecto, quienes trabajan en el área gastan un esfuerzo enorme para determinar qué es lo que ya está hecho, donde está, y en que formato. Por otra parte, es común que mucha de la información no se actualice, o las series se descontinúen. Finalmente, aunque se sepa donde está la información, es difícil accederla en forma oportuna y completa, sin gastar un gran esfuerzo para lograr ubicar a la persona que pueda ayudar a ello. A pesar de que un determinado Servicio o Institución puede estar dispuesto a entregar cierta información a quien la solicite, las dificultades de llegar a la información y solicitar la autorización para su uso, pueden hacerla inaccesible en la práctica. Esto sucede, incluso para aquella información que no es confidencial o de acceso restringido. Simplemente no existe una cultura orientada a facilitar el acceso a la información, ni están implementados los sistemas requeridos para ello. Por ello, la Comisión Nacional del Medio Ambiente (CONAMA), de acuerdo a la ley 19.300 de bases generales del medioambiente, ha iniciado el desarrollo de un sistema nacional de información ambiental (SINIA), el cual es concebido como un sistema de libre acceso para todos los potenciales usuarios. Cabe describir brevemente las principales características de los usuarios y proveedores de información de este sistema.

Los usuarios de información ambiental son tan diversos como sus requerimientos. Por ejemplo: (a) existen usuarios a nivel escolar, profesional y académico, (b) con diversidad de preocupaciones, por ejemplo público versus privado, (c) con diversas escalas de preocupaciones, por ejemplo nivel nacional versus regional, entre otros. Algunos usuarios necesitan información técnica, antecedentes cuantitativos, mediciones o estadísticas, otros necesitan estadísticas generales, y otros querrían saber cómo denunciar a una empresa o bus que “échan” humo. Considerando esta característica de diversidad, la caracterización de los usuarios y sus requerimientos se hace imposible, al ser el SINIA un sistema de acceso público a la información medioambiental del dominio CL. Es clave, por tanto, generar un sistema de acceso fácil, de bajo costo y organizado de una manera tal que cada usuario pueda encontrar rápidamente lo que anda buscando.

Los proveedores de información por su parte, instituciones públicas y privadas, universidades, bibliotecas, entre otros, son bastantes más localizables y de menor número que los usuarios. Ellos pueden ser caracterizados desde el punto de vista de la frecuencia de generación de información y el nivel de cooperación existente entre los proveedores y el equipo SINIA como: (a) proveedores regulares (corresponden a aquellas instituciones que poseen sitios establecidos), y (b) proveedores esporádicos (corresponden a aquellos proveedores que incorporan o eliminan sitios y páginas en forma volátil). Estos proveedores de información crean, modifican y/o eliminan documentos o páginas webs, alterando el acceso y catalogación de sus documentos. Se requiere por tanto, un sistema de información

que, reconociendo la diversas de fuentes de información, pueda sistematizarlo de una manera simple y ponerla al alcance de los potenciales usuarios.

La actual situación de alta descoordinación y baja cooperación entre proveedores y usuarios de información, genera altos costos de transacción para la búsqueda y recopilación de información, baja motivación para compartir y homogeneizar información existente, y pocos incentivos para generar información faltante. En los últimos años, con el desarrollo de Internet, ha surgido una plataforma tecnológica que permite reducir costos de transacción e integrar proveedores y usuarios de la información medioambiental distribuidos geográfica y culturalmente. Sin embargo, es necesario desarrollar todo un conjunto de mecanismos y aplicaciones que permitan construir una correcta integración entre proveedores y usuarios de información. Este mecanismo de intermediación e integración es ejemplificado en la Figura 1.



**Figura 1: Mecanismo de Intermediación del SINIA**

CONAMA en un esfuerzo realizado en forma conjunta con instituciones, organizaciones y universidades, públicas o privadas, pretende precisamente, incorporar en un sistema de información, a todos aquellos agentes que estén relacionados con el tema medioambiental en Chile. Esta es una actividad inédita en Chile y poco común en el mundo, donde una institución pública utiliza recursos para mejorar el acceso a la información, de manera de reducir los costos de transacción en el ámbito de competencia. Este esfuerzo ha sido materializado en el Sistema Nacional de Información Ambiental (SINIA), que tiene como uno de sus objetivos básicos el poner al alcance y/o en conocimiento de los distintos usuarios de información ambiental, la información disponible en las distintas instituciones con competencia ambiental. El SINIA, desde un punto de vista conceptual, es un sistema de información medioambiental público, que intermedia

entre proveedores y usuarios, poniendo a disposición toda la información medioambiental nacional relevante. El valor agregado del SINIA corresponde a: (a) localización y/o generación de información en formato Internet, a partir de información medioambiental relevante existente en el dominio CL, y (b) la clasificación (edición de las diferentes categorías y palabras claves) de los contenidos que en dicho sistema son publicados.

El presente trabajo presenta el diseño y desarrollo del SINIA, desde un punto de vista práctico. La experiencia adquirida en este esfuerzo será útil para apoyar otros esfuerzos similares. El trabajo está organizado de la siguiente manera: en la sección 2 se presentan los trabajos relacionados, en la sección 3 se establecen los principios de diseño del sistema de información, en la sección 4 se presentan las principales áreas de desarrollo del trabajo, en la sección 5 se presentan los principales resultados, y finalmente, en la sección 6 las conclusiones.

---

## 2 Trabajos Relacionados

---

El trabajo desarrollado (un sistema de intermediación de información) tiene las siguientes características básicas: (a) categorización útil de búsqueda, (b) motor de búsqueda eficiente, (c) clasificación eficiente de la información, y (d) búsqueda y monitoreo de proveedores de información. Estas características requieren que se desarrolle un sistema de administración, de información y de control de gestión, de manera de proveer los servicios necesarios, tanto para los usuarios como los administradores del sistema. Es por ello, que al momento de analizar los trabajos relacionados es necesario hacerlo desde esta perspectiva, es decir, en base a las diferentes características mencionadas.

En el tema de definición de una taxonomía ambiental existe una serie de resultados, en particular fuentes relacionadas con el tema ambiental (Azqueta 1994, Canter 1998, Jain et al. 1995, Field 1995, Kopp y Smith 1993), sitios web internacionales (Environment Canadá, Environment Australia, Harvard University, Resources for the Future, The World Bank Group, United Nations Environment Programme (UNEP), y US EPA), y el Programa de Gestión y Economía Ambiental (PROGEA<sup>1</sup>). Sin embargo, ninguna de estas clasificaciones es completamente adecuada a la realidad de SINIA, ya que presentan orientaciones particulares, y en general, son de un alto detalle y especificidad. Es por ello, que se prefirió adecuar las diferentes taxonomías existentes en la literatura y de sitios web extranjeros, de una manera acorde con el tipo de información que existe en nuestro país.

En relación al tema de sistemas de administración y control de gestión, existe una gran cantidad de aplicaciones que integran los diferentes aspectos organizacionales (Sowa y Zachman 1992, Zachman 1987), metodologías orientadas a

---

1 [www.dii.uchile.cl/progea](http://www.dii.uchile.cl/progea)



coordinar esfuerzos en la provisión de servicios (Barros 1994, Action Tech. 1999, y prácticas de trabajo que deberían adoptarse en este tipo de proyectos (Cortada, 1998). Sin embargo, las características específicas de la institución (CONAMA) y del sector medioambiental chileno, requieren que se desarrollen procedimientos ad-hoc orientados a la realidad chilena.

Finalmente, en términos de herramientas de apoyo, en la actualidad existe un gran número de soluciones de software de apoyo a la búsqueda y clasificación de información, en particular, existen múltiples buscadores genéricos (Yahoo, Altavista, Infoseek, Hotboot, etc.) y otros orientados a la clasificación y recuperación de información ambiental (Environmental Protection Agency<sup>2</sup>, Environmental Australia Online<sup>3</sup>, The Green Lane<sup>4</sup>, The European Commission Environment<sup>5</sup>, United Nations Environment Programme<sup>6</sup>, The World Bank Group<sup>7</sup>, CNIE<sup>8</sup>, entre otros). Estos sistemas de búsqueda están siendo replanteadas, desde un mero motor de búsqueda de alta eficiencia (medida por su velocidad, indexación, etc.), a sistemas más orientados a ayudar a los usuarios en la búsqueda de información. Uno de los primeros fue Yahoo!, el cual ayuda a los usuarios a encontrar información, no sólo como un motor eficiente de búsqueda, sino como un eficaz sistema de clasificación que utiliza (muchas) personas para categorizar correctamente los cientos de miles de sitios en la WEB. Yahoo! es uno de los sitios más populares, quizás por su énfasis en el trabajo editorial de los sitios que maneja.

Muchas empresas u organizaciones han reconocido la importancia de la clasificación y categorización de datos a través de personas y no sólo en forma automática. En los años 1996-1997, los motores de búsqueda compitieron por el número de páginas indexadas y la velocidad de búsqueda, con lo cual se enfatizó la clasificación automática. Hoy día, la mayoría de éstos entienden que es completamente irrelevante si un motor de búsqueda entrega 100.000 o un millón de sitios como resultado de una búsqueda. En general, estos sistemas están reconociendo el límite de la cantidad masiva y falta de calidad de la información que ellos almacenan. Así, están preparando un número de estrategias para agregar "contexto editorial" a sus datos. Por ejemplo, Infoseek y Excite tiene "canales" o "guías" donde los editores (personas reales) identifican la calidad de los sitios en áreas particulares y guían a los usuarios a ellas, en vez de dejarlos que busquen solamente por palabras.

Las funciones de búsqueda y recuperación son asumidas en la mayoría de los productos y servicios. Sin embargo, el verdadero valor agregado de estos servicios se ha ido orientando a la clasificación y catalogación de información, lo que en definitiva es lo que atrae y mantiene a los usuarios. Los motores de búsqueda

---

2 URL [www.epa.gov/epahome/browse.htm](http://www.epa.gov/epahome/browse.htm)

3 URL [www.erin.gov.au](http://www.erin.gov.au)

4 URL [www.ec.gc.ca/issues\\\_e.html](http://www.ec.gc.ca/issues\_e.html)

5 URL [europa.eu.int/comm/environment/policy\\\_en.htm](http://europa.eu.int/comm/environment/policy\_en.htm)

6 URL [www.unep.org/unep/prdct.htm](http://www.unep.org/unep/prdct.htm)

7 URL [www.worldbank.org/html/extdr.thematic.htm](http://www.worldbank.org/html/extdr.thematic.htm)

8 URL [www.cnie.org](http://www.cnie.org)

están utilizando una variedad de tecnologías avanzadas para ayudar a los usuarios a obtener la información que ellos requieren, aún si no han preguntado por ella. Esta nueva tendencia generará una mayor especialización y preocupación por lo que los usuarios quieren y buscan en este tipo de sistemas.

---

### 3 Principios de Diseño

---

El SINIA se ha definido como un mecanismo de intermediación de la información medio ambiental chilena para los diferentes proveedores y usuarios de dicha información. Para llevar a la práctica este concepto establecido para el SINIA, se han desarrollado 6 principios generales de diseño que se presentan a continuación.

1. **Completitud de la Información:** el diseño deberá permitir incorporar (de diversas fuentes de información) la mayor cantidad de información medio ambiental relevante, y mantener actualizada su caracterización.
2. **Trasparencia de la Información:** el diseño deberá garantizar que la información medio ambiental este disponible y sea fácilmente accesada por cualquier usuario.
3. **Soportar la Heterogeneidad:** el diseño deberá soportar diferentes tipo de fuentes (antiguas) de información, paradigmas de procesamientos y arquitecturas heterogéneas (bases de datos, paginas estáticas o dinámicas, etc.).
4. **Flexibilidad:** el diseño deberá incorporar, mantener y/o eliminar en forma eficiente nuevas fuentes de información relevantes.
5. **Escalabilidad:** el diseño deberá proveer los mecanismos necesarios para soportar el actual y proyectado nivel de uso de sistema.
6. **Automatización:** el diseño deberá incorporar la mayor cantidad de automatización en las tareas que así lo requieran (tediosas y de bajo valor agregado), de manera que solo las actividades claves sean desarrolladas en forma no automatizada.

Estos principios son desarrollados en los tres niveles de manejo, procesamiento e integración de la información siguientes: (a) nivel de información a nivel de usuarios, (b) nivel de administración de información, y (c) nivel de obtención y procesamiento de la información. En la Figura 2 se caracterizan estos niveles de responsabilidad del manejo de la información.

En particular, el diseño se centró en la identificación, caracterización y modelamiento de los principales procesos que sustentan el SINIA. El concepto de proceso es entendido como un conjunto coordinado de acciones o actividades, que utilizando un conjunto de recursos, generalmente escasos, generan un producto que permite alcanzar los objetivos de la organización o cliente. La principal característica de los procesos es que permiten alinear los esfuerzos y recursos (personas, productos y tecnología) en torno a objetivos comunes, y por ende lograr una alta efectividad en la creación de valor a las organizaciones.

Nivel De Información	Servicio
Nivel usuarios	Los usuarios definen los requerimientos de información, y pueden manipular datos areas de interes especificas y ejecutar busquedas ad hoc a sus intereses.
Administración de Información	Este nivel deberá soportar requerimientos (escalables) de procesamiento de información, búsquedas, etc. Los administradores definen los responsables y dueños de la información y procesos que el sistema deberá proveer. Aquí se definen la políticas de replicación, almacenamiento y manejo temporal de la información.
Obtención de Información	Este nivel es responsable por replicar información (si es necesario), almacenarla y distribuirla a nivel nacional. Responsable por monitorear, recolectar y garantizar la calidad de la información (clasificación, localización y disponibilidad) desde las diferentes fuentes de información medioambiental relevantes. Este nivel es responsable por definir los mecanismos de obtención, niveles de calidad.

Figura 2: Niveles de Gestión de Información

---

## 4 Areas de Desarrollo

---

Establecido el marco para el SINIA, fue necesario desarrollar 3 actividades centrales para implantarlo: (a) desarrollo y mantención de la taxonomía, (b) mantención y generación de cooperación con instituciones, y (c) mantención de documentos (monitoreo, clasificación y publicación). A continuación se detalla cada una de ellas.

### **4.1 Desarrollo de una Taxonomía**

La taxonomía desarrollada incorporó dos criterios agregados de catalogación: (a) categoría ambiental y (b) palabras claves (descriptores). Esta subdivisión obedece a una necesidad tanto de búsqueda de información clasificada dentro del sistema, como de una estructuración clara y sistemática de ella, orientada a los distintos tipos de usuarios del sistema. Además, se requiere orientar la búsqueda agrupándola en temas que sean de interés nacional y que abarquen todo el abanico de tópicos e información relevante. Finalmente, esta taxonomía se ha desarrollado de una manera que facilita al interesado clasificar la información, ya que se considera un número bajo de categorías y algo no mayor, pero manejable, de palabras claves.

- **Criterio por Categoría Ambiental:** El criterio denominado categoría ambiental clasifica la información de acuerdo a temas ambientales relevantes. Estos temas se han estructurado de manera de abarcar la mayor cantidad de información ambiental. Específicamente, cada criterio se definió como sigue: con un mínimo de categorías, permitiendo una búsqueda fácil y sistemática. Considerando que muchos de los usuarios no son especialistas se ha optado por usar palabras simples, de uso común, y cuyo sentido sea realmente obvio.
- **Criterio por Palabras Claves (Descriptoros):** En forma complementaria, se proponen palabras claves, las que permiten, por una parte, clasificar en forma sistemática, y por otra, facilitar la búsqueda de información (a nivel de usuario) mediante descriptoros más especializados que las categorías. El criterio para definir estas palabras claves no ha sido el tratar de abarcar la mayor cantidad de temas, sino que englobar en algunas palabras aquellos temas más relevantes a juicio de los expertos ambientales del proyecto y de los profesionales de CONAMA.

A partir de estos criterios se definieron 30 categorías ambientales y 43 palabras claves lo que se presentan en el Anexo 1. Además para facilitar su uso por parte del clasificador, se desarrolló una definición por comprensión y extensión de cada una.

En el desarrollo se tomaron como base de información tres fuentes principales. La primera es la revisión de una serie de libros y documentos relacionados con el tema ambiental, principalmente el “Tesauro de Ingeniería Sanitaria y Ambiental” del CEPIS y el listado de términos utilizados en *The Green Lane*, de Environment Canada, de tal manera de identificar los tópicos más relevantes utilizados en el lenguaje ambiental. La segunda fuente de información han sido sitios web internacionales, a partir de los cuales se han determinado criterios de clasificación de la información, categorías ambientales, criterios de búsqueda de información y esquemas de presentación de dicha información resultante. La tercera fuente de información es el expertise adquirido por el equipo de trabajo del Programa de Gestión y Economía Ambiental (PROGEA) del Departamento de Ingeniería Industrial de la Universidad de Chile. Esta experiencia acumulada en la investigación y el desarrollo de proyectos del área medio ambiental es utilizada para ordenar la información de la literatura y de sitios web extranjeros, además de las distintas formas de clasificación y presentación, de manera que sea acorde con el tipo de información que existe en nuestro país.

#### **4.2 Cooperación con Proveedores de Información**

En esta área de trabajo se persigue que las instituciones públicas o privadas generen o pongan a disposición el material medioambiental relevante para que sea publicado por los responsables institucionales en el SINIA. Para ello fue necesario identificar la instituciones (públicas y privadas) y la información medioambiental relevante que poseen (en sus diferentes formatos, volúmenes, frecuencias de actualización, etc.). Sin embargo, identificar las instituciones poseedoras de la información ambiental relevante no asegura en ninguna medida

el éxito de un proceso de cooperación mutua para la difusión de información ambiental.

Ello debido principalmente a la reticencia habitual de las diferentes instituciones en la entrega de cualquier tipo de información a terceros. Por ello se decidió que era importante, al menos en una etapa inicial, firmar convenios de cooperación, que fijaran lo que se incluye en las páginas webs y un compromiso de actualización periódica. Esto llevó a la definición de ciertos aspectos claves a considerar para la priorización de instituciones y de la manera de abordar el contacto para iniciar el proceso de colaboración mutua en la difusión de información ambiental, a saber:

1. Conocer la situación actual: Se debe tener presente la situación actual de las instituciones para poder entrar a intervenir positivamente. Esto permite abrir una puerta hacia el interior de la institución, con miras al objetivo final que es la generación de los convenios.
2. Conocer las limitaciones: Es necesario tener presente las posibles restricciones que va a tener el proceso, por ejemplo, limitaciones del tipo legal, del tipo informático (confidencialidad / privacidad de la información, confiabilidad de la información), limitaciones de tipo operacional (responsable de la mantención y actualización de la información), limitaciones de tiempo de respuesta, etc.
3. Conocer la estructura organizacional: Informarse de la estructura de la organización (organigrama jerárquico y de funciones) para poder dirigir la información al sujeto objetivo que proceda, dándole una mayor efectividad a la gestión.
4. Identificar el material de trabajo existente: Registrar y recopilar todos los documentos, que ya existen sobre este proceso (oficios, documentación tipo de convenios, formularios de catastro, material de contacto -cartas- e-mail-faxes-memorándums).
5. Analizar la documentación existente: Este análisis es hecho bajo la nueva perspectiva de generación de convenios y con el fin de generar la nueva documentación de trabajo.

Una primera fase para el trabajo consistió en la realización de un catastro simple de la información medioambiental relevante. Esto implicó el desarrollo de caracterización organizacional, así como de las habilidades para recolectar, generar, procesar y administrar la información ambiental. Basados en estos antecedentes se realizó una priorización y selección de las instituciones a contactar. Se desarrolló además un formato estándar para recoger la información. En Anexo 2 se presenta dicho formato.

La segunda fase consistió en la evaluación de la voluntad o interés de la institución correspondiente en participar en proyectos de este tipo, considerado el factor clave del éxito del proyecto. Para aquellas instituciones más proclives a la cooperación, se realizaron los contactos formales y concertaron las reuniones respectivas. En tales reuniones se hizo una introducción general de lo que significa SINIA, sus alcances y proyecciones, y se realizó una invitación a participar en el

proyecto. Para dar incentivos a los participantes se facilitó poner la información en formato digital para desarrollar las páginas Web o las aplicaciones, según el caso, solicitadas por la institución.

Finalmente, y con aquellas instituciones que decidieron realizar un trabajo de cooperación, se acordó el trabajo conjunto con la institución y se procedió a la elaboración de catastros detallados de la información que sería llevada a la web, con lo cual es posible dimensionar el trabajo informático requerido. La recolección de la información y la creación de páginas web y aplicaciones son consecuencias del trabajo realizado y programado, con lo cual en esta etapa no debieran resultar imprevistos.

### 4.3 Operación y Gestión de Documentos

La operación del SINIA tiene por objetivo mantener actualizada la información publicada en SINIA con respecto al dominio CL, considerando los actuales y potenciales proveedores de información. Esto es modelado como un proceso de negocios, donde el ciclo de trabajo principal se descompone en 5 niveles : (a) monitoreo (búsqueda e identificación), (b) preclasificación, (c) clasificación, (d) modificación, y (e) publicación. Se debe tener en mente dos momentos básicos en el proceso de gestión de mantención, en un primer momento la administración de los documentos a mantener y por otro lado la confección, edición y publicación de la página web. La Figura 3 presenta el esquema del proceso basado en el modelo de ciclos de trabajo.

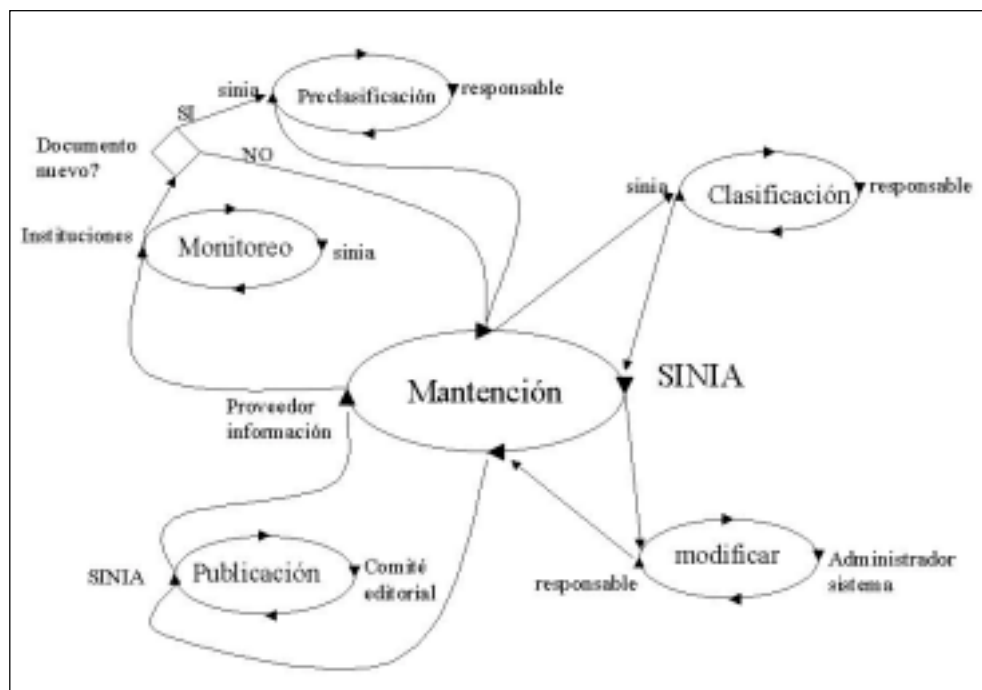


Figura 3: Proceso de Manejo de Documentos

El proceso de mantención se inicia con la oferta por parte SINIA o alguna institución en particular de publicar documentos (instituciones públicas o privadas) relacionadas con el medio ambiente. Este proceso es iniciado por un agente (persona o software) que identifica proveedores de información y/o documentos (nuevos, modificados y/o eliminados), para los cuales realiza un oferta de incluirlos en el SINIA. A continuación se documentan las principales actividades (ciclos de trabajo) de este proceso.

1. **Monitoreo:** En el proceso principal de Mantención entra el subproceso de Monitoreo en donde existe un agente, ya sea una persona o conjunto de personas o un agente de software o robot, que inicia el proceso de búsqueda de documentos. El monitoreo tiene como objetivo detectar e identificar documentos nuevos, modificados o eliminados susceptibles de ser incorporados en el SINIA, es decir, que posean información medioambiental relevante. En esta actividad se realizan las siguientes actividades: (a) identificación de documentos nuevos no existentes en el SINIA, que poseen información medioambiental relevante, y (b) detección de modificaciones o eliminación de documentos existentes en el SINIA.
2. **Preclasificación:** Si en el subproceso de monitoreo se detecta un documento nuevo, este documento se preclasifica por los responsables de SINIA según los criterios preestablecidos. Este subproceso se inicia por la petición de los responsables de SINIA de la búsqueda de documentos nuevos a un responsable de preclasificación, en lo cual se considera los criterios y políticas preestablecidos de preclasificación.
3. **Clasificación:** La clasificación se inicia a petición de los responsables de SINIA o proveedores de información, para que el documento sea clasificado en definitiva. El Comité de Evaluación, integrado por una o más personas, deberá evaluar el documento para clasificarlo en forma definitiva o rechazarlo en su incorporación al SINIA. Finalmente, el subproceso de clasificación termina con el resultado de aprobación o no del documento para su publicación. Esta decisión es comunicada al proveedor de información y a los responsables de publicación de documentos.
4. **Efectuar la Modificación:** En la ejecución del proceso se realiza la modificación de la clasificación del documento en la metabase de datos. En el caso, que sea un documento nuevo se procede a registrar el documento en forma adicional, en el caso de uno antiguo, se actualizan los cambios, eliminaciones y/o modificaciones de documentos.
5. **Publicación:** El proceso de publicación se inicia una vez terminado el proceso de mantención, y consiste en habilitar el documento para el acceso por parte de los usuarios o clientes del SINIA. En primer lugar se solicita el cambio de estado del documento, desde documento no activo a documento activo para el acceso de usuarios. Finalmente se realiza el cambio de estado del documento, quedando habilitado para que los usuarios realicen búsquedas sobre dicho documento.

Este proceso corresponde al proceso fundamental del SINIA, ya que de él depende la calidad de la información existente en el SINIA (corrección de la

clasificación y de los documentos incluidos), y la completitud de dicha información. Es importante hacer notar que la actividad de monitoreo es fundamental, ya sea automatizada (apoyada por algún software) o no (monitoreo realizado por personas), debido a que provee la sustentabilidad y genera confianza en los usuarios.

---

## 5 Resultados

---

El SINIA ha desarrollado una organización a cargo de CONAMA y ayudado por el Departamento de Ingeniería Industrial de la Universidad de Chile. En ella, un grupo de expertos apoya el desarrollo de la cooperación institucional, evalúa la operación del sistema, y genera los planes de desarrollo. Es importante hacer notar que el trabajo desarrollado a la fecha ha consistido en el poblamiento, apoyo a la cooperación institucional y construcción de la interface de usuarios. A continuación se presentan los principales resultados en lo referente a: (a) aplicación Internet, (b) cooperación institucional y (c) rediseño del SINIA.

### **5.1 Aplicación Internet**

En la actualidad se dispone de una aplicación Internet que provee la información medioambiental nacional en la dirección [www.sinia.cl](http://www.sinia.cl). Esta aplicación cuenta con cerca de 2000 documentos medioambientales pertenecientes a 360 instituciones. En forma adicional, el SINIA posee una interface a nivel nacional y 6 a nivel regional. La Figura 5 muestra la interface de consulta localizada en [www.sinia.cl](http://www.sinia.cl).

El resultado de las búsquedas es categorizado, en primera instancia, por las categorías ambientales que tienen relación con el patrón de búsqueda utilizado. Posteriormente, se entrega el total de documentos recuperados, ordenados y caracterizados en el siguiente orden: (a) regionales y (b) nacionales. En forma adicional, una vez ingresado un patrón de búsqueda (i.e., las palabras y opciones seleccionadas para buscar documentos en el SINIA), y estando en los resultados, es posible obtener los resultados de las diferentes regiones para el mismo patrón, seleccionando la región correspondiente en el nivel superior de la pantalla.

### **5.2 Cooperación Institucional**

En términos de la cooperación institucional se ha establecido un mecanismo de cooperación dependiendo del interés mostrado por la institución cooperadora. Así, se establecen tres niveles de cooperación dados por:



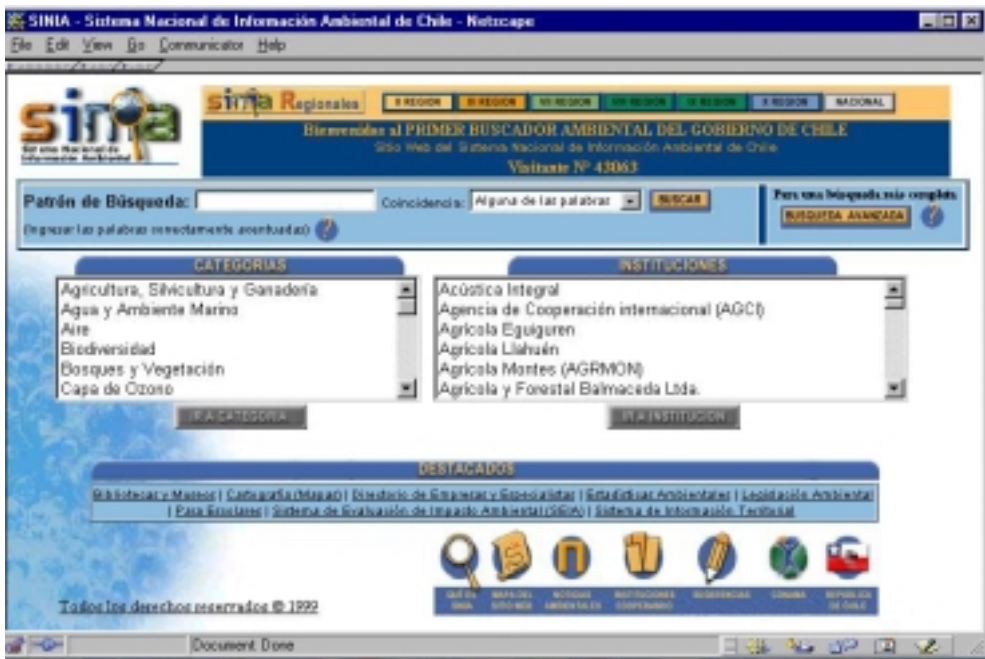


Figura 4: Página Principal SINIA

- Cooperación Marco: donde se establece el compromiso de cooperación en la generación de catastros de información medioambiental institucional. Este nivel de cooperación debe ser establecido al más alto nivel institucional, de manera de dar la mayor amplitud en la construcción del catastro.
- Cooperación de Desarrollo: este nivel de cooperación considera, una vez catastrada la información, que se deberán realizar aplicaciones institucionales, con apoyo del SINIA, que permita tener acceso vía Internet a la información medioambiental relevante. En forma conjunta, se deberán definir las responsabilidades de mantención y actualización de la información.
- Cooperación Amplia: este nivel de cooperación es de un carácter permanente, donde las instituciones se comprometen a actualizar periodicamente los catastros y las aplicaciones institucionales, y donde ellas poseen un acceso a la administración de su información en el SINIA.

A la fecha se han establecido un trabajo de desarrollo de aplicaciones con 6 instituciones, de las cuales 2 han establecido convenios institucionales de cooperación.

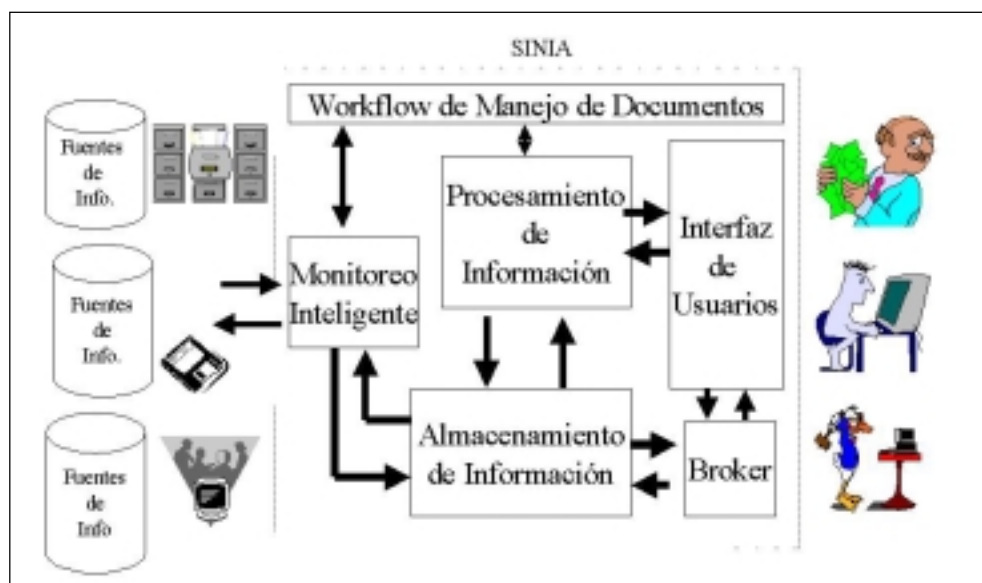


Figura 5: Arquitectura del SINIA

### **5.3 Rediseño del SINIA**

La tarea desarrollada hasta el momento se ha realizado en dos etapas. La primera consistió en el poblamiento de información medioambiental para una aplicación existente de SINIA, previa al proyecto actual. En una segunda etapa, y con el conocimiento de los problemas de la aplicación original, se desarrolló el rediseño del SINIA, desde un punto de vista de aplicaciones modernas en Internet.

El rediseño del SINIA comprendió dos áreas de trabajo, por un lado, el diseño de un sistema administrador descentralizado, y por otro, un sistema activo de suscripción de información medioambiental. Este diseño cumple con los principios establecidos en el presente trabajo, y sus principales características se entregan en la arquitectura siguiente y se esquematizan en la Figura 5. Los elementos de la arquitectura propuesta son descritos a continuación:

1. Fuentes de Información: el SINIA deberá interactuar con diversas fuentes de información, ya sea a nivel de tipo de información, como de mecanismos de almacenamiento (bases de datos, páginas webs estáticas y/o dinámicas, etc.).
2. Monitoreo Inteligente: la necesidad de mantener actualizado la información en el SINIA requiere de un monitoreo de las diversas fuentes de información. Este monitoreo deberá ser adecuado a los niveles de compromiso que existan con las diferentes instituciones proveedoras de información. Así aquellas que posean convenios, el monitoreo será mucho más localizado. Sin embargo aquellas instituciones u organizaciones que no posean convenios, el monitoreo deberá ser más extenso, ya que es difícil que dichas instituciones reporten los cambios realizados.

3. Almacenamiento de Información: la información recolectada (una vez depurada) permitirá realizar la clasificación y publicación de dichos documentos (meta - documentos) en el SINIA.
4. Servicio de Broker: los diferentes usuarios podrán interactuar con el SINIA mediante un BROKER (intermediario), cuya funcionalidad corresponde a la definición de grupos de interés y la comunicación de cambios relevantes a aquellos usuarios que lo requieran.
5. Procesamiento de Información: este módulo considera el procesamiento de requerimientos completamente estructurados. En particular, el modelo de toma de decisiones y control de gestión del SINIA (evaluación de convenios, consultas, etc.). En forma adicional, preguntas e informes frecuentes son incorporados en éste módulo.
6. Interfaz de Usuarios: los usuarios (en sus diversos tipos) requerirán interactuar con el sistema, para lo cual deberán definirse las interfaces, niveles de seguridad y acceso, localización geográfica, etc. de los usuarios.
7. Workflow: a este nivel se manejan los meta-documentos asociados a los documentos (información almacenada por el SINIA), aquí se definen los diferentes procesamientos, reglas para que los documentos sean clasificados, publicados, y revisados periódicamente.

---

## 6 Conclusiones

---

El trabajo realizado se orientó a tres áreas: (a) identificación, poblamiento y puesta en marcha del SINIA, (b) establecimiento de procedimientos y apoyo a la cooperación institucional, y (c) rediseño del SINIA y construcción de la interface de búsqueda. La estructuración del SINIA como un sistema administrativo ha sido el principal resultado del trabajo, debido a que permitirá su sustentación en el tiempo. En particular, para los temas destacados las conclusiones son las siguientes.

En términos de poblamiento, se localizaron, clasificaron y poblaron cerca de 2000 documentos medioambientales en Internet y 360 instituciones, lo cual a juicio de los autores, es una recopilación bastante exhaustiva de los documentos medioambientales en el dominio CL.

En el caso del desarrollo de niveles de cooperación institucional, se propuso un mecanismo de realizar dicha cooperación, estructurándola en un proceso de negocio. En este proceso se identifican las actividades de búsqueda, recopilación y estructuración de la información medioambiental, basado principalmente en visitas a terreno y el trabajo conjunto. Finalmente, se establecen niveles de cooperación que deberían desarrollarse con las diferentes instituciones, de acuerdo al grado de interés que demuestren.

Finalmente, y en relación con el rediseño, se propone una estructura del SINIA como un sistema administrativo y de información, donde no sólo los aspectos

tecnológicos son abordados, sino también los procedimientos organizacionales. El (re)diseño propuesto incorpora los aspectos tradicionales, así como aspectos de sistemas activos, proveyendo un servicio de suscripción de información para los usuarios. En forma paralela, se diseñaron los principales procedimientos administrativos y de gestión que el sistema deberá apoyar. Por último, se abordó el mejoramiento de la aplicación computacional de búsqueda, incorporando la dimensión geográfica en ella, y haciendo más fácil su uso por parte de los usuarios.

\*\*\*\*

### **Agradecimientos**

Los autores agradecen a SINIA y CONAMA por el contrato 23-22-002/9 y 23-22-001/9 del proyecto Programa 1999 de Clasificación, Poblamiento y Rediseño del SINIA, que hicieron posible el desarrollo de este trabajo.

---

## Referencias Bibliográficas

---

- Azqueta, D., 1994. Valoración Económica de la Calidad Ambiental. McGraw Hill.
- Barros, O., 1994. Reingeniería de Procesos de Negocios: un planteamiento metodológico. Dolmen, Santiago.
- Canter, L., 1998. Manual de Evaluación de Impacto Ambiental. McGraw Hill.
- Cortada, J., 1998. Best Practices in Information Technology: How Corporations Get the Most Value from Exploiting their Digital Investments. Prentice Hall PTR.
- Field, B., 1995. Economía Ambiental: una Introducción. McGraw Hill.
- Action~Technologies Inc. Internet workflow applications, designing web-based applications to support higher customer expectations. White Paper, 1999.
- Jain, R. et als., 1995. Environmental Assessment. McGraw Hill.
- Kopp, R. and K. Smith, 1993. Valuing natural resources, the economics of natural resources damage assessment. Resources for the Future.
- Mulligan, D., 1996. Environmental Management in the Australian Minerals and Energy Industries: Principles and Practices. UNSW Press.
- Sowa, J. and J. Zachman, 1992. Extending and formalizing the framework for information systems architecture. IBM Systems Journal, 31.
- Zachman, J., 1987. A framework for information system architecture. IBM Systems Journal, 26.

## Anexo 1

### Taxonomía del SINIA

El criterio denominado categoría ambiental clasifica la información de acuerdo a temas ambientales relevantes. Estos temas se han estructurado de manera de abarcar la mayor cantidad de información ambiental, permitiendo una búsqueda fácil y sistemática.

Agricultura, Silvicultura y Ganadería	Gestión Ambiental Pública
Agua y Ambiente Marino	Industria y Medio Ambiente
Aire	Legislación y Normas Ambientales
Biodiversidad	Minería
Bosques y Vegetación	Monitoreo y Vigilancia
Capa de Ozono	Participación Ciudadana
Ciudad y Medio Ambiente	Patrimonio Ambiental
Clima y Cambio Climático	Residuos Sólidos
Comercio Internacional y Medio Ambiente	Riesgos Ambientales
Desarrollo Sustentable	Ruido
Ecología	Salud y Medio Ambiente
Economía Ambiental Impacto Ambiental Educación Ambiental	Sistema de Evaluación de Suelos y su Administración
Energía y Combustibles	Sustancias Tóxicas y Peligrosas
Estadísticas Ambientales	Transporte

**Tabla A1. Categorías Ambientales**

En forma complementaria a las categorías ambientales, se proponen palabras claves, las que permiten clasificar en forma sistemática y luego facilitar la búsqueda de información (a nivel de usuario) por parte de los antecedentes que no han sido definidos y englobados en forma clara por las categorías ambientales y pensando en una manera más amistosa de entregar información. El criterio para definir estas palabras claves no ha sido el tratar de abarcar la mayor cantidad de temas, sino que englobar en algunas palabras aquellos temas más relevantes a juicio de los expertos ambientales del proyecto y de los profesionales de CONAMA.

AGRICULTURA, SILVICULTURA Y GANADERÍA	GESTIÓN AMBIENTAL PÚBLICA
Agua y Ambiente Marino	Industria y Medio Ambiente
Acuerdos Internacionales	Etnias
Aguas Subterráneas	Fiscalización
Ambiente Acuático	Gestión del Agua
Análisis Costo Beneficio	Indicadores Ambientales
Areas y Especies Protegidas	Instrumentos Económicos
Bosque Nativo	Lluvia Acida
Calentamiento Global	Monumentos Naturales
Cartografía y Mapas	Normas ISO
Censo	Ozono
Conservación de la Naturaleza	Paisaje
Contaminación Ambiental	Pesticidas
Control Sanitario	Planes de Descontaminación
Convenciones y Tratados	Población
Desarrollo Urbano	Producción Limpia
Desastres Naturales	Reciclaje:
Desierto	Recursos Naturales
Dispersión y Difusión	Residuos Sólidos y Líquidos
Ecosistemas	Tecnologías Ambientales
Emergencias Ambientales	Tipos de Clima
Erosión	Turismo
Especies Biológicas	Uso de Suelos
Especies Migratorias	

Tabla A2. Palabras Claves

## Anexo 2

### Formulario de Catastro de Información

A continuación se detallan los dos documentos utilizados en el trabajo con instituciones para la realización del catastro. Primero, el esquema de administración de tareas a desarrolladas con las intituciones y los tipos de trabajos establecidos. Posteriormente, se muestra el formulario de catastro de información institucional.

**ESQUEMAS DE TRABAJO INSTITUCIONALES**

N° N°	INSTITUCIÓN	ESQUEMA DE TRABAJO		OBSERVACIONES
		DESARROLLO DE APLICACIONES	CREAR PÁGINAS CLASIFICAR SITIO WEB	

**Esquema de Trabajo Institucional**

## CATASTROS DE INFORMACION

**INSTITUCION:**

**CONTACTO:**

**FECHA PRIMERA REUNION:**

**Catastro:**

**Recopilación de información:**

**Responsable: Gonzalo Asencio**

Características de los datos a obtener		Características técnicas e informáticas				Observación						
Título del Documento o grupo de documentos.	Descripción General	Acceso	TAGs	Datos Originales	Salida		Actualización	Nombre del Archivo	Software	Hardware	Conectividad	Volumen

Figura 6: Formulario de Catastro de Información



---

# Una Aplicación de Prismas de Factibilidad para Resolver el Problema de Asignación de Horarios en una Empresa de Doblajes de Películas

---

**Carmen Ortiz\***

Escuela de Ingeniería Industrial  
Universidad Adolfo Ibáñez

**Roberto Opazo\*\***

Empresas VyReal

---

## Resumen

---

*El problema de asignación de horarios es un clásico de la literatura en el área de la investigación operativa. Normalmente se presenta aplicado a colegios o universidades con grandes simplificaciones, lo que hace que las soluciones obtenidas sean de dudosa aplicabilidad.*

*En este trabajo se define una versión aplicada del problema para una empresa de doblajes de películas, con la intención de adaptar la definición de un modelo a las necesidades de la empresa y no la empresa a las limitaciones del modelo.*

*El resultado es un sistema computacional fácil de utilizar, eficiente y que genera soluciones mejores de las que una persona es capaz de encontrar. Además, se propuso una estructura de datos especializada en la administración de restricciones de horario, que puede ser utilizada en la implementación de muchas heurísticas en el área.*

---

\* E-mail: cortiz@uai.cl

\*\* Roberto Opazo es académico jornada parcial del Departamento de Ciencias de la Computación de la Universidad de Chile. E-mail: roberto@opazo.cl

---

## 1. Introducción

---

Una empresa de doblajes de películas recibe películas en un idioma, por ejemplo inglés y las entrega dobladas a otro idioma, por ejemplo, castellano. Para esto cuenta con un grupo de actores, los que deben grabar con su voz la traducción del texto original de los personajes de la película.

Para una empresa de doblajes mediana, la cantidad de personajes es muy superior a la cantidad de actores disponibles, por lo que se requiere organizar el trabajo de doblaje de los actores, considerando que cada actor sólo puede estar en un estudio de grabación en cada momento del tiempo y, adicionalmente, debido a que los actores además tienen otros trabajos cada actor entrega una disponibilidad de horario a la empresa.

Existen varias formas de organizar a los actores: una de ellas consiste en reunir a todos los actores que doblan la voz de los personajes que participan en cada escena y reproducir sincronizadamente los diálogos doblados en el idioma objetivo; otra alternativa consiste en capturar la voz de cada actor separadamente y luego, mediante un proceso de mezcla, conseguir la secuencia de sonidos que corresponde a la escena doblada. También es posible trabajar con una combinación de ambos métodos, situación que corresponde al caso que se describe en este trabajo.

Una programación horaria para actores de una empresa de doblajes, consiste en indicar las horas de inicio y término en que los actores asisten a cada estudio para doblar los parlamentos de los personajes que se les haya asignado. Además, debe indicarse las líneas de texto que cada actor debe doblar en estas entradas a los estudios de doblaje.

No es posible determinar en forma exacta el tiempo que utilizará un actor para grabar los parlamentos de un personaje, pues, en la práctica, intervienen una serie de factores impredecibles en la grabación. Por ejemplo, un actor puede requerir ensayar viendo la cinta original antes de grabar, puede equivocarse realizando la grabación o puede estar en desacuerdo con el director en la forma de interpretar al personaje. Además, estos tiempos varían entre los actores y también, algunos personajes son más fáciles de doblar que otros, por ejemplo un dibujo animado que no mueve la boca es fácil de doblar, porque los sonidos son más fáciles de coordinar con el movimiento de imagen en pantalla. Esta situación es difícil de administrar, pues si el tiempo de grabación de un actor es una variable estocástica, entonces todo el problema de programación horaria lo es.

Históricamente, la programación de actores se ha realizado en DINT (Doblajes Internacionales) en forma manual, asignando un tiempo aproximado para el doblaje de cada personaje. En la implementación de la heurística que se describe en este trabajo, se mantuvo esta simplificación que evita la variable estocástica en el problema.

Por restricción de la empresa, cada película debe grabarse completamente en un sólo estudio, no pudiendo grabarse alguna parte en otro estudio, o suspender su grabación para grabar partes de otra película. Esto se debe a que los directores son citados a los estudios para dirigir películas específicas y existen complicaciones técnicas asociadas al cambio de película, como tiempos de inicialización de las cintas correspondientes.

A cada actor se le puede asignar la grabación de uno o más personajes en todas las películas que la Unidad de Operaciones decida hacerlo participar. En casos excepcionales, puede exigirse que dos o más actores asistan concurrentemente a grabar en un estudio. Esto puede deberse a características especiales del personaje o de la escena. Por ejemplo, un personaje puede requerir doblarse con dos voces en paralelo (como Dos Caras, el enemigo de Batman), o la escena puede consistir en el grito de varios personajes. En esta última situación, no es posible aprovechar el proceso de mezcla de la película para juntar los gritos, pues sólo se dispone de dos pistas de sonido libre para juntar grabaciones hechas en los estudios. Se usan otras dos pistas para incluir los sonidos de música y efectos especiales.

Los actores no forman parte de la planta fija de personal de la empresa, por lo que no es posible contar con ellos en cualquier horario de cualquier día. Para administrar la información de las disponibilidades de horario, los actores llenan un formulario en el cual indican los horarios en que pueden asistir a grabar. Eventualmente, los estudios también pueden no estar disponibles, debido a diferentes razones, como mantenciones, horarios de almuerzo o la realización de trabajos que no pueden ser planificados por Operaciones.

Para organizar la asignación de parlamentos en horarios de grabación, la película es dividida en unidades elementales de guión, llamadas "loop". Cada loop consiste en 4 líneas seguidas del mismo personaje, sin importar si éstas se ven interrumpidas por la intervención de otros personajes.

Actualmente, la empresa cuenta con 4 estudios de grabación y más de 50 actores, lo que es una empresa mediana a pequeña en relación con el tamaño de las empresas españolas que se dedican a la misma actividad, pero es la empresa más grande en Chile en el rubro.

Una forma de abordar este problema es mediante programación matemática. En el apéndice 1 se muestra un modelo entero mixto, de variables enteras y continuas, que describe una versión simplificada del problema considerando sólo un estudio de grabación.

Para administrar la disponibilidad de horario de los actores, la jornada de trabajo se divide en bloques de 30 minutos cada uno, denominados módulos. En el caso de DINT, la jornada de trabajo se extiende desde las 8:00 a las 22:00 hrs. Por lo que hay 28 módulos diarios. El modelo posee gran cantidad de variables binarias, lo que en general, hace crecer exponencialmente el tiempo requerido para su resolución, independientemente del algoritmo y del software que se use. Podría hacerse el esfuerzo de preprocesar el modelo para reducir el número de variables, aprovechando simetrías y otras propiedades. Con ello se lograría reducir a la mitad el número de variables de secuenciamiento, sin embargo, aún con esta mejora, el número total de variables enteras para una instancia pequeña del problema (mostrada en el apéndice) resulta ser de 12.825, si se considera un horizonte de planificación de 1 día, 50 actores, cada uno con un promedio de 10 módulos disponibles y un máximo de 3 entradas por día. Así, dada la dimensión del modelo, la utilización de software comercial capaz resolver problemas mixtos de optimización quedó descartada.

Este problema se asemeja a una combinación entre los problemas de Asignación de Maquinaria y los problemas de Asignación de Horarios de Clases. La principal diferencia con los problemas clásicos de Asignación de Maquinaria es que en este caso se requiere especificar restricciones de horario. Y la principal diferencia con las formulaciones previas del problema de Asignación de Horarios de Clases es que el tiempo de asignación es una variable continua y no un módulo de largo fijo, por ejemplo, todas las clases duran 90 minutos.

Los problemas de Asignación de Maquinaria han sido estudiados en una amplia gama de trabajos previos, indicándose en muchos de ellos formas eficientes de encontrar soluciones exactas y heurísticas, un resumen clasificado de estos problemas puede verse en Pinedo (1995). En cambio, el problema de Asignación de Horarios en Escuelas, en casi todas las experiencias previas ha sido formalizado con tantas simplificaciones, que no resulta práctico para las necesidades de situaciones reales (Kitagawa y Ikeda, 1987), una de las pocas buenas simplificaciones se presenta en Ortiz (1989).

En Opazo (1999) se demostró que el Problema de Asignación de Horarios aplicado a una Empresa de Doblajes de Películas pertenece a la clase NP-Completo, por lo que se resolvió utilizar un método heurístico de solución.

A continuación se describe la heurística desarrollada en este trabajo.

## 2. Heurística

Esta heurística es capaz de administrar las restricciones de horario de los actores y de los estudios, combinándolas con criterios de planificación tales como considerar las preferencias de los actores por algunos módulos de grabación (unos prefieren la mañana, otros la tarde, etc.), mantener la duración (tiempo) de las entradas a grabar en cada estudio en rangos razonables, considerar el tiempo que le toma a un actor trasladarse de un estudio a otro, entre otras consideraciones.

La Ilustración 1 muestra la estructura general de la heurística:

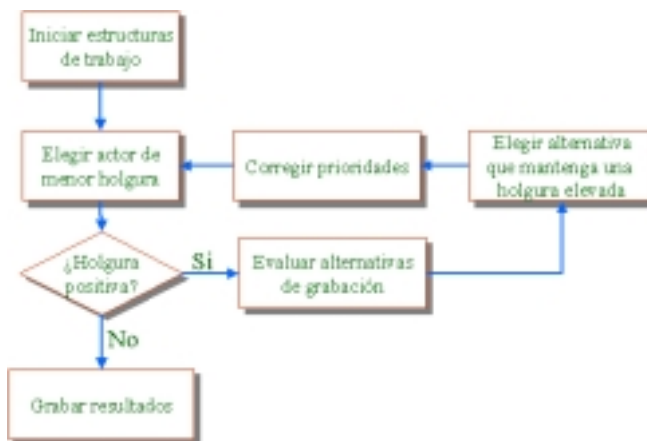


Ilustración 1: Estructura General de la Heurística

El método se articula en torno a un ciclo en el que en cada iteración se elige el elemento (actor) con mayor prioridad (menor holgura) y se asigna considerando sus preferencias y los espacios de tiempos menos demandados, con esto se busca mantener siempre a los elementos pendientes por ser asignados con la mayor holgura posible:

$$\text{holguraActor} = \frac{\sum \text{alternativasReales} \times \text{duracionAlternativa}}{\sum \text{tiempoPersonaje}}$$

Las alternativas reales de grabación son calculadas para cada actor en cada estudio, sumando los minutos en que se cumplen 3 condiciones: el actor puede grabar, el estudio está disponible y además el actor tiene personajes con texto pendiente asignados al estudio en evaluación. Este valor es dividido por la suma de minutos requeridos para doblar todos los parlamentos de personajes asignados al actor.

La prioridad de cada actor se define de modo que tengan mayor prioridad los actores con menos holgura, por ejemplo, si un actor debe doblar la voz de dos personajes, para lo que necesita 49 minutos (esto se calcula basándose en la cantidad de loops en el parlamento de cada personaje y la velocidad de grabación definida por el usuario para cada personaje) y tiene una disponibilidad de horario que incluye 5 módulos, pero en uno de esos módulos el estudio está en mantención y en otro módulo el estudio tiene asignada una película para la cual el actor no tiene que doblar voces, entonces sus alternativas reales son sólo 3 y en consecuencia la holgura del actor se define como  $3 \cdot 30 / 49 = 1,84$ . De esta manera un actor con menos disponibilidad tiene menos holgura y a más loops por grabar también disminuye la holgura.

La complejidad de la heurística es de orden  $O(J^2)$ , en que  $J$  representa el número de actores que deben doblar personajes en la película que se está programando. El ciclo principal de la heurística se ejecuta una vez para cada entrada a grabar de algún actor (el promedio de entradas por actor ( $p$ ) varía entre 2 y 4), por lo que este ciclo se ejecuta  $p \cdot J$  veces. Por otra parte, la operación más costosa en tiempo dentro del ciclo es la de corrección de prioridades, ya que se debe revisar para cada actor la forma en que la asignación elegida afecta su prioridad, la que podría ocupar espacios temporales en los que el actor tenía disponibilidad de horario y por lo tanto si estos tiempos los usa otro actor, entonces disminuyen las alternativas reales de grabación, esta operación requiere  $O(J)$  operaciones.

El diseño de la heurística es simple, pero su implementación podría ser muy compleja si no se utiliza una herramienta especializada. La complejidad en la implementación se deriva de la necesidad de responder en cualquier momento, durante la asignación, preguntas como las siguientes:

¿Cuáles actores pueden trabajar en un horario particular en un estudio?

Esto requiere revisar los datos de las disponibilidades de horario, los personajes asignados al actor (debe tener al menos uno con parlamentos pendientes en la

película asignada al estudio en dicho tiempo) y las grabaciones programadas para el actor.

¿Cuáles son los tiempos de horario más demandados?

Un espacio de tiempo está demandado por cada actor que tiene disponibilidad, no ha sido asignado y tiene personajes en la película correspondiente.

¿En qué horarios pueden trabajar juntos todos los actores de un grupo dado?

¿Cuáles son todas las alternativas de grabación para un actor particular?

---

### 3. Uso de Prismas de Factibilidad

---

Como se ha dicho, la principal dificultad de este procedimiento es la administración de las alternativas de asignación de cada actor en todo momento del tiempo durante su ejecución.

Por ejemplo, inicialmente un actor puede ser asignado en un módulo en el cual se cumplen las siguientes condiciones:

- El actor tiene disponibilidad de horario.
- El estudio tiene disponibilidad de horario.
- El actor tiene trabajo pendiente para la película que se está grabando en ese momento en el estudio.

En un punto intermedio de la asignación la situación se vuelve más compleja, puesto que también hay que considerar que:

- El actor puede tener completamente asignada la carga de trabajo correspondiente a la película que se está grabando en el módulo del estudio.
- El módulo del estudio puede haber sido asignado parcial o totalmente a otro actor.
- El actor puede haber sido asignado en ese horario a otro estudio. Si así fuera, hay que considerar además, que el actor demora una cantidad de tiempo en trasladarse de un estudio a otro.

Para administrar esta información de un modo fácil y eficiente, se creó una estructura de datos denominada “prisma de factibilidad” y una operación de proyección sobre este prisma.

Un prisma de factibilidad es una secuencia de cristales. Cada cristal representa el estado de un intervalo de tiempo y está identificado por:

- Un inicio ( $t_{Ini}$ ) y un fin ( $t_{Fin}$ ), medidos en unidades de tiempo (minutos).
- Un color (Color), expresado como un número entero. Si el número es positivo, se entiende que el cristal es transparente, es decir factible o disponible. En

cambio un color 0 o negativo, indica que el cristal es opaco o infactible. Los colores de los cristales pueden representar distintas cosas, por ejemplo la preferencia del actor por grabar durante ese tiempo o la demanda de actores por un espacio de tiempo particular de grabación.

- Una caja de información asociada (INF), la cual contiene datos asociados al estado del cristal (espacio temporal). En el caso de la heurística presentada en este trabajo, se decidió que las cajas de información puedan almacenar una identificación de película, una identificación de personaje, una identificación de estudio y un valor real llamado carga, que indica el porcentaje de sobrecarga al cual puede estar sometido el cristal. Si este valor es 1, entonces el trabajo asignado en el cristal puede desempeñarse a una velocidad normal, pero si la carga es, por ejemplo, 1.2, entonces el cristal tiene asignado un 20% de trabajado adicional con relación al que se lograría grabando a velocidad normal, por lo que el actor debe trabajar más apurado. Uno de los parámetros de la heurística controla la sobrecarga máxima que se le puede asignar a un cristal.

Un cristal puede ser imaginado como se muestra en la Ilustración 2



Ilustración 2: Cristal de Factibilidad

Un prisma de factibilidad es una secuencia de cristales contiguos. Entonces, puede verse como se representa en la Ilustración 3.

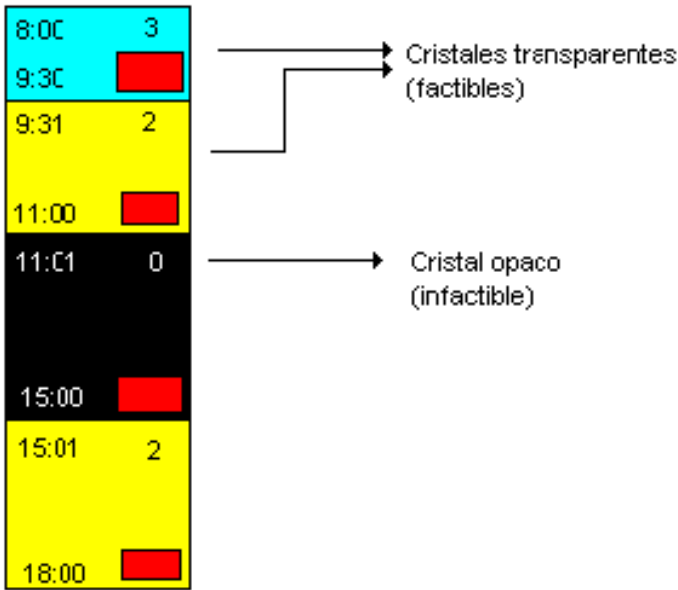


Ilustración 3: Prisma de Factibilidad

Los prismas son usados en la heurística para administrar la siguiente información

- **Prismas de Actor:** Las preferencias de los actores son almacenadas en 5 prismas (uno para cada día de la semana). Inicialmente estos prismas tienen cristales negros en todos los módulos en que el actor no puede grabar. En los otros módulos el prisma tiene cristales cuyo color indica la preferencia del actor por el módulo, a mayor número (más claro), mayor es la preferencia. La caja de información está inicialmente vacía.
- **Prismas de Estudio:** Cada vez que un actor es asignado a grabar en un estudio, se marca un cristal desde el instante de inicio hasta el final de la grabación con color opaco (-1) y la caja de información se llena con todos los datos del personaje y estudio en el que se trabajará. Los prismas de estudio inicialmente tienen cristales negros en todos los módulos en que el estudio no está disponible y cristales con color cuyo número se calcula como la cantidad de actores que pueden utilizar ese módulo para grabar al menos uno de sus personajes asignados y la caja de información guarda información sobre la película que debe grabarse en ese cristal.
- **Asociaciones actor-estudio:** Para cada par (actor, estudio), se definen 5 prismas (uno para cada día de la semana), que tienen cristales negros en los intervalos de tiempo en que el actor no tiene carga de trabajo pendiente para la película que está asignada en ese momento en el estudio y tiene color transparente (1) cuando sí tiene personajes asignados.

Existen dos operaciones que pueden ser realizadas sobre un prisma de factibilidad:



- **Creación del prisma:** Inicializarlo como un sólo cristal completamente opaco. Esta operación puede tener distintas interpretaciones dependiendo del tipo de prisma. Si es un prisma actor, entonces se tiene un actor sin disponibilidades (ni preferencias) de horario; si es un prisma estudio, entonces el estudio está cerrado todo el día; y si es un prisma actor-estudio, entonces el actor no tiene ningún personaje que grabar de la película asignada al estudio en cada momento del tiempo. En todos estos casos la estructura de datos y la operación usada es la misma, lo que cambia es el ente dueño del prisma.
- **Pintado del prisma:** Indicar el tiempo de inicio y término en que el prisma debe tener un color determinado y una caja de información con valores específicos. Esta operación crea un nuevo cristal en el prisma que se está pintando, con el valor correspondiente en el color y caja de información. Si después de pintar un cristal o una parte de un cristal, el prisma queda con dos cristales contiguos de las mismas características (color y caja de información), entonces estos cristales son unidos, generando un nuevo cristal más grande (desde  $t_{Ini}$  del primer cristal hasta  $t_{Fin}$  del segundo) con el color y caja de información indicado. Esto evita la fragmentación de cristales, con lo que se protege el tiempo de respuesta de la heurística.

Además se define una operación, llamada Proyección de Prismas de Factibilidad, que toma como entrada dos o más Prismas de Factibilidad. Al proyectar dos o más prismas de factibilidad, se obtiene otro prisma, cuyos cristales están definidos por el color que generaría un rayo de luz al intentar pasar por todos los cristales de la línea que le corresponda. Un ejemplo se muestra en la Ilustración 4.

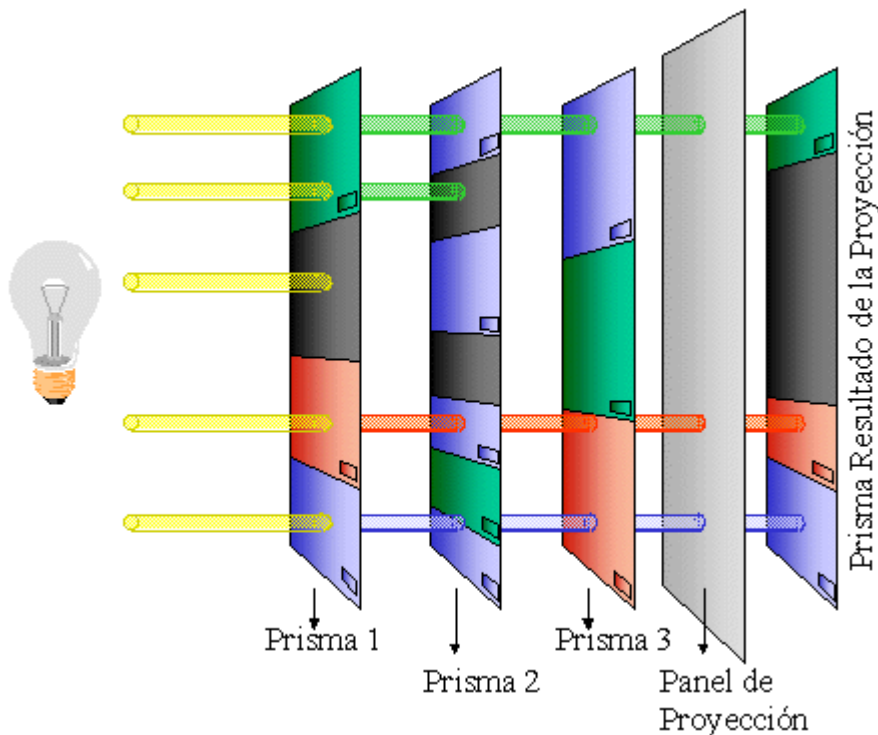


Ilustración 4: Proyección a través de prismas de Factibilidad

En la figura anterior sólo se proyectaron 3 prismas, pero podría ser cualquier número de éstos. El prisma resultante es una secuencia de cristales.

En el prisma resultante aparecen cristales negros en todos los intervalos de tiempo en los que alguno de los prismas proyectados haya tenido un cristal opaco. El color y la información de los cristales transparentes del prisma resultante, están determinados por el cristal del primer prisma en la lista a proyectar. Si después de la proyección quedan en posición contigua 2 cristales de las mismas características, estos son fundidos en un sólo cristal que abarque todos los intervalos de tiempo involucrados. El poder de la operación de proyección queda más claro al ver la forma en que puede usarse para responder algunas de las preguntas fundamentales de este proceso, planteadas previamente:

¿Cuáles actores pueden trabajar en un horario particular en un estudio?

Primero se construye un prisma auxiliar que sólo tiene transparente el horario por el cual se desea preguntar, luego, para cada actor se proyecta el prisma auxiliar, con el prisma actor y el prisma actor-estudio correspondiente, si el prisma resultante tiene un cristal transparente en el horario consultado, entonces la asignación es factible.

¿Cuáles son los horarios más demandados?

La demanda por cada intervalo de tiempo en un estudio está dada por el color del cristal con que se representa dicho intervalo en el prisma estudio.

¿En que horarios pueden trabajar juntos todos los actores de un grupo dado?

Para conocer las disponibilidades comunes de un grupo de actores, basta con proyectar todos sus prismas actor.

¿Cuáles son todas las alternativas de grabación para un actor particular?

Las alternativas de grabación de cada actor están dadas por el resultado de proyectar el prisma actor, el prisma estudio y el prisma actor estudio correspondiente para cada estudio.

---

## 4. Calibración de la Heurística

---

La solución que entrega la heurística depende del valor que se asigne a varios parámetros de entrada.

Estos parámetros pueden ser agrupados en dos grandes categorías: aquellos que deben ser definidos por el usuario, en función de sus requerimientos (como la duración de los módulos y el tiempo de traslado entre estudios) y los que tienen una influencia en el comportamiento de la heurística, pero cuya magnitud ideal es difícil de determinar por el usuario, como el valor de premios o pesos de diferentes criterios. Sin embargo, el usuario puede indicar relaciones cualitativas indirectas de estos

parámetros, como por ejemplo definir si es más importante que las programaciones usen al máximo los estudios o respetar las preferencias de los actores por los horarios de los módulos.

La heurística implementada considera 7 parámetros directamente definidos por el usuario, entre los que se cuentan el tiempo mínimo, máximo y óptimo de duración de las entradas a grabar de un actor, la duración de los módulos para entregar las disponibilidades de horario (30 minutos), el tiempo de traslado entre estudios (5 minutos, se asume constante para todos los viajes entre estudios), la cantidad de entradas diaria máxima para los actores (3 entradas máximo en un día) y la máxima sobrecarga (10% por sobre lo que se graba normalmente en un intervalo de tiempo).

Para definir el valor a usar en los otros parámetros (los que no puede definir el usuario directamente) se decidió utilizar un criterio simple, que permite reducir a un valor escalar la calidad de una solución a una instancia del problema de asignación de horario. Con este criterio simple o función de desempeño, es posible modificar el valor de los diferentes parámetros para estudiar su impacto en el resultado de la función objetivo.

La función de desempeño analizada fue la tasa de utilización promedio de los estudios, ya que la maximización de ésta es el principal objetivo de la empresa. En forma complementaria se obtuvo el número de entradas promedio por actor, índice que debería ser lo más pequeño posible a fin de no tener las entradas demasiado fragmentadas. Es decir, una solución es mejor mientras mayor sea la tasa de utilización de los estudios y menor sea la cantidad de entradas a grabar de los actores.

Para analizar el impacto de cada una de estas variables en la calidad de la solución entregada por el programa, se experimentó con un ejemplo real de programación de estudios, en el cual se requiere programar un horizonte de dos días, con tres estudios. Las 4 películas ingresadas suman 173 personajes a doblar, con 1678 loops<sup>1</sup> en total, para lo que se cuenta con 35 actores diferentes asignados al reparto de estas películas.

En el funcionamiento interno, la heurística utiliza varios criterios de asignación, estos criterios no siempre son complementarios y se requiere determinar cuáles reglas tendrán mayor importancia con respecto a las otras en la elección de una asignación de horario parcial (una entrada a grabar de un actor), para esto se asignó un valor numérico que se calcula para cada criterio de selección y un peso (número que se multiplica por el nivel de cumplimiento de la regla).

Existen 2 tipos de parámetros estudiados que el usuario no puede definir directamente, esto es por lo difícil que resulta predecir su efecto en las soluciones generadas.

---

1 Cada loop representa 4 líneas de parlamento del mismo personaje.

Parámetros de pesos relativos: Son usados para modificar la importancia de cada criterio con respecto a los otros, como el peso de cuan demandado es cada intervalo de tiempo, peso de las preferencias por módulos de los actores y peso del respeto por el número máximo de entradas

Parámetros usados para calcular el valor numérico que indica el grado de cumplimiento de un criterio: Premio por no dejar espacios vacíos al asignar a un actor, premio por llenar un espacio vacío al asignar a un actor, premio por conectar una entrada de un actor con otra del mismo actor, castigo por violar el tiempo mínimo de duración de una entrada y número de actores demandando un espacio temporal que vuelve crítica o urgente su asignación.

Para analizar la sensibilidad de la heurística con respecto a cada uno de estos parámetros, se hicieron pruebas preliminares en las que se evaluaron todas las combinaciones de valores para cada parámetro, pero considerando pocas alternativas de valor por parámetro con el fin de mantener controlado el número de escenarios diferentes. Esto permitió conocer los intervalos en que cada parámetro puede afectar en la calidad de la solución encontrada.

Luego se analizaron separadamente (dejando fijos los otros valores) cada uno de los parámetros a calibrar, pero esta vez en un rango grande de valores, partiendo por los parámetros que demostraron mayor varianza en la etapa previa.

En los gráficos de las Ilustraciones 5 y 6 se muestra el valor de la tasa de ocupación de los estudios (curva de arriba) y número de entradas generadas en promedio por actor (curva de abajo).

Puede apreciarse que la función objetivo definida (tasa de utilización) no es una curva suave, es decir, si al aumentar el valor de un parámetro tiende a mejorar la calidad de las soluciones obtenidas, esto no significa que siempre que el parámetro aumente, la solución mejora. En algunos gráficos su comportamiento es tipo zigzag, esto se explica por el “Efecto Dominó” que se desencadena al cambiar alguna de las decisiones de asignación en etapas tempranas de la ejecución de la heurística.

Sin embargo, en todos los gráficos es posible distinguir una tendencia, sobre la base de la cual se eligieron los valores definitivos para cada parámetro. Se presentan a continuación dos gráficos elegidos:

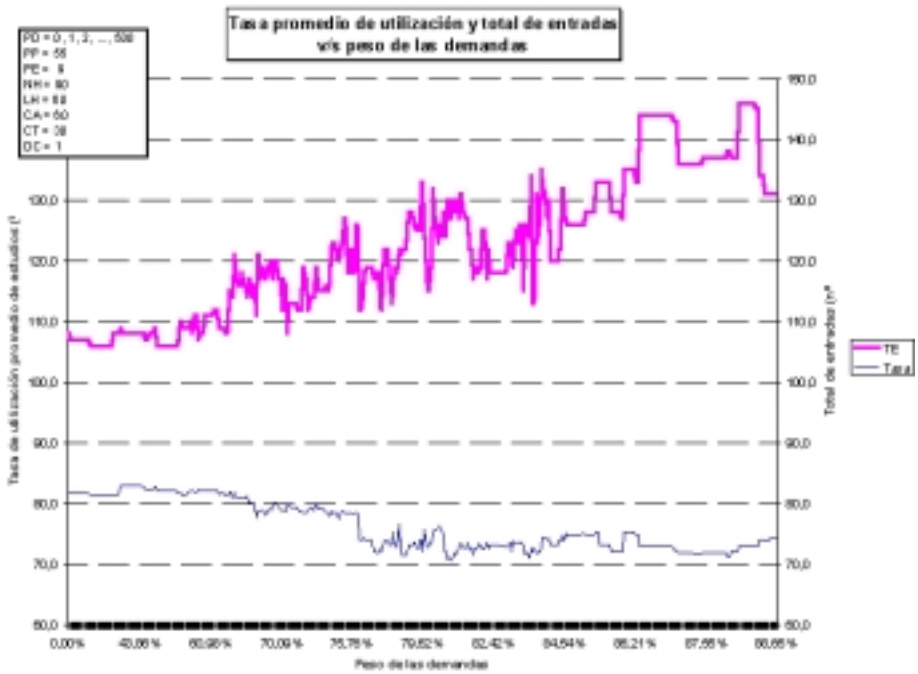


Ilustración 5: Ejemplo de gráficos de calibración de la heurística

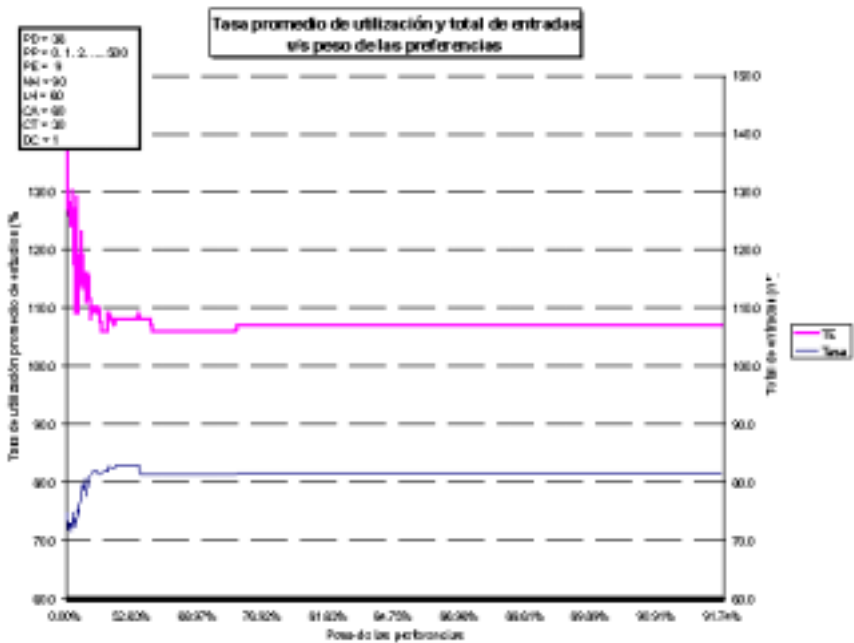


Ilustración 6: Ejemplo de gráficos de calibración de la heurística

---

## 5. Conclusiones

---

El problema de asignación de horarios, aplicado a una empresa de doblajes de películas, es un desafío difícil de enfrentar, debido a complicaciones propias del proceso en el cual está inmerso y a sus características combinatoriales, que hacen complejo encontrar soluciones en forma automatizada.

En general, se ha estudiado el problema tanto desde una perspectiva teórica, como práctica. Desde el punto de vista teórico, se han construido modelos de programación lineal con variables continuas y discretas los que resultan ser tan grandes, en términos del número de variables enteras y de restricciones, que se hace infactible usarlos para realizar programaciones en situaciones reales. Por ello se decidió implementar una heurística especializada.

La heurística mencionada, fue implementada para la empresa DINT (Doblajes Internacionales) y fue probada junto con la Jefa de Operaciones de la Empresa.

La ejecución computacional de la programación de estudios permite realizar la asignación varias veces al día, antes de optar por una planificación definitiva. La ventaja de esto es que cada ejecución permite identificar a los actores que tienen disponibilidades de horario muy restringidas, negociar con ellos un aumento en su disponibilidad y así mejorar la programación de los estudios. La velocidad en la ejecución de la heurística, es el factor determinante en la factibilidad de este tipo de negociaciones. Con el nuevo sistema la empresa es capaz de agregar una disponibilidad de horario y volver a ejecutar la heurística en menos de 5 minutos.

La ejecución de la heurística toma 2 minutos aproximadamente, en un PC Pentium de 200 Mhz, con 64 Mb en memoria Ram y 2 Gb de espacio en disco.

El sistema fue conectado de un modo semiautomático tanto con el sistema que genera información de entrada para la programación de estudios (digitación de guiones), como con el sistema que recibe los resultados de la operación de doblaje en los estudios (contabilidad). La definición clara de estas interfaces de entrada y salida del sistema permitió concentrar todos los esfuerzos en el tema central en estudio y entregar una solución que no afectó a las otras áreas de la organización.

Los Prismas de Factibilidad resultaron ser una estructura de datos muy útil para manejar situaciones complejas de disponibilidad de horario. Esta estructura, también podría usarse para implementar heurísticas en otros problemas relacionados, como asignación de horarios en colegios y universidades, planificación de vuelos en aerolíneas, asignación de personal a ambulancias o empresas de servicios en general, etc. Algunos de los problemas mencionados son más simples que el problema resuelto en este trabajo, debido a que en ellos el tiempo a asignar se divide en unidades discretas de tiempo, normalmente llamadas módulos.

En la práctica la separación de los conceptos color de un cristal y caja de información resulta artificial, pues dos cristales son diferentes ya sea porque sus colores son distintos o porque sus cajas de información lo son. Entonces resulta

más natural y mejora la lúdica de los Prismas de Factibilidad definir el color de un cristal como un concepto más abstracto, que incluye todos los datos relevantes de una caja de información.

La operación de Proyección Simple entre Prismas de Factibilidad es sumamente poderosa para determinar las disponibilidades comunes entre distintos elementos que puedan manejar algún tipo de gestión de horario de trabajo, sin embargo la regla que indica que el color y la caja de información del prisma resultante están definidos por los valores que estos elementos tengan en el primer prisma usado en la proyección, hace que se pierda información útil contenida en los otros prismas. Esta operación podría aumentar su poder si se definiera una “Proyección Compuesta”, en la cual el programador pudiera ver el prisma resultante indicando en cada acceso con respecto a cual de los prismas usados en la proyección original se deben definir los colores y cajas de información de los cristales del prisma resultante.

La heurística aquí presentada también puede usarse en conjunto con otras metaheurísticas que requieren de algún método para obtener una solución inicial e intentan mejorar la solución encontrada, por ejemplo, se podría utilizar búsqueda tabú.

También sería interesante comparar el desempeño de esta heurística con otras, como búsqueda tabú, algoritmos genéticos, etc.

---

## 6. Referencias Bibliográficas

---

- Pinedo, 1995. *Scheduling, Theory, Algorithms and Systems*. PRENTICE HALL, New Jersey/USA.
- Kitagawa y Ikeda, 1987. An existencial problem of a weight-controlled subset and its application to school timetable construction. *Discrete Mathematics*, 72, Hiroshima/Japón, Pág. 195.
- Ortiz, 1989. *Asignación de Horarios en un Régimen Curricular Flexible*. Tesis de Magíster en Ciencias de la Ingeniería, Mención Industrial, Universidad de Chile, Stgo/Chile.
- Opazo, R., 1999. *Una Solución al Problema de Asignación de Horario para una Empresa de Doblajes de Películas*. Universidad de Chile, Stgo/Chile.

---

## 7. Apéndices

---

### **Apéndice 1: Modelo Matemático**

El modelo que se presenta fue desarrollado para analizar la posibilidad de una solución basada en software ya existente, capaz de resolver problemas de programación lineal entera, como Cplex, Lindo, Maple, etc.

Para facilitar la comprensión del modelo, todas las ecuaciones se presentan en forma simplificada, suponiendo un problema con sólo una película, un estudio y un día de planificación.

El modelo busca maximizar la tasa de utilización de los estudios. Por simplicidad, no se incluye la consideración de preferencias de los actores en la función objetivo.

### Definiciones Previas

**Loop:** Como ya se dijo, los guiones son divididos en unidades elementales que en este caso corresponden a cuatro líneas consecutivas del guión para el mismo personaje. En caso de que el texto del personaje sea interrumpido por otros personajes, entonces las líneas del loop no quedan seguidas, pero cada loop debe contener cuatro líneas, a menos que el personaje ya no tenga más intervenciones en la película.

**Sesión:** Entrada a grabar a un estudio por parte de un actor. Las sesiones tienen definida una hora de inicio y una de término. En cada sesión un actor entra a grabar a un estudio durante el tiempo programado e intenta grabar todos los loops asignados.

**Módulo:** Para administrar la disponibilidad de horario de los actores, la jornada de trabajo es dividida en bloques de 30 minutos denominados módulo. Indices: Se usará el índice  $j$  para actores;  $e$ , entradas y  $m$ , módulos.

### Parámetros del modelo

$E =$	número máximo de sesiones para actores, 3 diarias en la situación actual.
$L =$	Duración de los módulos (30 minutos).
$M =$	número total de módulos.
$T_{\max} =$	Duración máxima de una entrada (180 minutos).
$T_{\min} =$	Duración mínima de una entrada (15 minutos).
$d_{mj} =$	1 si el actor $j$ tiene disponibilidad en el modulo $m$ , 0 si no tiene.
$W =$	Valor muy positivo
$J =$	Número total de actores, aproximadamente 50.
$P_j =$	Tiempo total que debe grabar el actor $j$ . Se calcula como la suma de los tiempos de los personajes asignados al actor $j$ .

### Variables de decisión

$t_{je} =$	Instante en que el actor $j$ entra a grabar la sesión $e$
$x_{je} =$	Tiempo de grabación del actor $j$ en la sesión $e$
$y_{je} =$	1 si el actor $j$ realiza la sesión $e$ , 0 si no.
$z_{j_1 e_1 j_2 e_2} =$	1 si el actor $j_1$ graba su entrada $e_1$ antes que el actor $j_2$ grabe $e_2$ , 0 si no.
$h_{jem} =$	Variable auxiliar binaria para implementar restricciones excluyentes.



### Función objetivo

El principal requerimiento que debe considerarse para elegir una solución factible, es que ésta sea la de mínimo tiempo de grabación para completar las películas. Esto se obtiene indirectamente al exigir que la tasa de utilización de los estudios sea la máxima posible. Con este criterio, se obtiene un beneficio adicional: es posible maximizar, aún cuando en el período de planificación no se pueda terminar de procesar toda la carga de trabajo asignada a los estudios y actores.

Por lo que la función objetivo puede escribirse como

$$\max \sum_{j \in \{1, \dots, J\}} \sum_{e \in \{1, \dots, E\}} x_{je}$$

Esta función objetivo podría ajustarse para considerar las preferencias de los actores, como criterio de selección, en caso de obtener 2 o más soluciones que arrojen el mismo valor para la función objetivo. O dar premios al valor de la función objetivo, si la selección tiene duraciones de las entradas cercanas a un valor óptimo.

### Restricciones en la duración de las entradas

El tiempo de grabación debe ser menor o igual que el máximo permitido

$$x_{je} \leq T_{max} \quad \forall j, e$$

El tiempo de grabación debe ser mayor o igual que el mínimo permitido si la entrada se realiza ( $y_{je} = 1$ ) y cero en caso de que la entrada no se realice ( $y_{je} = 0$ ).

$$W(1 - y_{je}) + x_{je} \geq T_{min} \quad \forall j, e$$

$$x_{je} \leq W y_{je} \quad \forall j, e$$

$$t_{je} \leq W y_{je} \quad \forall j, e$$

### Restricciones de trabajo asignado

$$\sum_{e \in \{1, \dots, E\}} x_{je} \leq P_{je} \quad \forall j \in \{1, \dots, J\}$$

Los actores no pueden grabar durante más tiempo del que se les asignó.

### Restricciones de secuenciamiento

En cada estudio, los actores graban sus parlamentos independientemente. En un momento posterior, cuando se edite la película, estos parlamentos serán mezclados de forma que puedan escucharse diálogos coherentes. Esto significa que las asignaciones de actores a estudios deben hacerse de modo que en cada entrada un actor termine todo el trabajo asignado antes de que entre el siguiente. Es decir, para toda entrada  $e_1$  del actor  $j_1$ , esta entrada se graba completamente antes o después que cada entrada  $e_2$  del actor  $j_2$ .

$$\begin{aligned} t_{j_1 e_1} + x_{j_1 e_1} &\leq t_{j_2 e_2} + W(1 - z_{j_1 e_1 j_2 e_2}) \\ t_{j_2 e_2} + x_{j_2 e_2} &\leq t_{j_1 e_1} + W z_{j_1 e_1 j_2 e_2} \\ &\forall e_1, e_2 \in \{1, \dots, E\} \end{aligned}$$

### Restricciones de disponibilidad de horario

Las entradas de los actores, deben ser en un momento del tiempo que tenga intersección vacía con los módulos en los que el actor no puede grabar ( $d_{mj} = 0$ ).

Para implementar este tipo de restricciones se deben generar restricciones para cada par módulo - actor en que el actor no tenga disponibilidad de horario. Dado un  $m$  y un  $j$  que cumplan este requisito, se debe exigir que la sesión del actor termine antes que el módulo sin disponibilidad, o comience después.

$$\begin{aligned} x_{je} + t_{je} &\leq L(m - 1) + W(1 - h_{jem}) \\ Lm &\leq t_{je} + W h_{jem} \quad \forall j, e \\ &\forall m, j \mid d_{jm} = 0 \end{aligned}$$

**Restricción de término de las últimas entradas**

Cada entrada debe terminar antes del fin de la jornada de trabajo.

$$x_{je} + t_{je} \leq ML \quad \forall j, e$$

**Dominios de variables**

$$x_{je}, t_{je} \in \mathbb{R}_0^+ \quad \forall j, e$$

$$y_{je}, z_{j_1e_1j_2e_2}, h_{jem} \in \{0, 1\}$$

En el caso que se estudió, se asumió que existen 50 actores (J = 50) que cada actor en promedio tiene 10 módulos sin disponibilidad y un máximo de 3 entradas diarias por actor (E = 3). La cantidad de variables binarias resultó ser:

Variables Binarias	Fórmula Base	Fórmula Optimizada	Cantidad Base	Cantidad Optimizada
Entradas ( $y_{je}$ )	J*E	J*E	150	150
Precedencia ( $z_{j_1e_1j_2e_2}$ )	J2*E2	(J*E)(J*E-1)/2	22.500	11.175
Auxiliares ( $h_{jem}$ )	J*E*10	J*E*10	1.500	1.500
Total			23.700	12.825



---

# Data Mining en la Empresa y en las Finanzas Utilizando Tecnologías Inteligentes

---

**Richard Weber<sup>1</sup>**

Departamento de Ingeniería Industrial,  
Universidad de Chile

---

## Resumen

---

*El desarrollo de software y hardware computacional ha resultado en la generación de enormes cantidades de datos almacenados en muchas empresas y organizaciones. Estos datos contienen información valiosa, que puede ser usada para mejorar la competitividad de estas instituciones. Data mining (minería de datos) se ha convertido en un área de investigación y desarrollo, en el cual se están proponiendo técnicas que apuntan a encontrar la información oculta en los datos. Este trabajo presenta aplicaciones de técnicas de data mining que han sido realizadas por un departamento de marketing en una institución bancaria.*

---

## 1. Introducción

---

El marketing juega un papel crucial en la mayoría de las empresas, tanto en las áreas de negocios y finanzas, como de comercio, servicios y telecomunicaciones. Entre las metas respectivas se incluyen las de encontrar nuevos clientes y mantener los actuales reconociendo y satisfaciendo sus necesidades. Hoy en día el desarrollo de *software* y *hardware* computacional permite a las empresas recoger y almacenar enormes cantidades de datos descriptivos de sus clientes y su comportamiento. El área de investigación y aplicación en la que se desarrolla estrategias de marketing fundadas en tales bases de datos se llama *database marketing* (Furness 1994; Nash 1993).

El presente trabajo considera aplicaciones de técnicas de *data mining* para el *database marketing*, tratando en mayor detalle del uso de tecnologías inteligentes tales como la lógica difusa (*fuzzy logic*) y las redes neuronales (*neural networks*).

---

1 Dirección: Departamento de Ingeniería Industrial, Universidad de Chile, República 701, Santiago, Chile. E-mail: rweber@dii.uchile.cl

En la Sección 2 se presentan las definiciones básicas junto con un resumen panorámico de *data mining*. La sección 3 presenta las tecnologías inteligentes usadas en esta publicación. En la sección 4 se presentan dos sub-tareas de un proyecto de *database marketing* realizado por una institución bancaria, junto con sus correspondientes resultados. La sección 5 presenta un resumen del trabajo, haciendo alusión además a futuros desarrollos y aplicaciones de los enfoques de *data mining*.

---

## 2. Descubrimiento de Conocimientos en Bases de Datos y Data Mining

---

*Data mining* entrega diversas técnicas para encontrar patrones en grandes conjuntos de datos. Este enfoque multidisciplinario combina los resultados e intuiciones provenientes de varias ramas científicas tales como la estadística, el aprendizaje de máquina (*machine learning*), tecnologías difusas (*fuzzy technologies*), y redes neuronales (véase por ejemplo, Adriaans y Zantinge 1996).

En la literatura existen muchas definiciones para caracterizar expresiones como “descubrimiento de conocimiento en base de datos” (KDD: *Knowledge Discovery in Databases*) y “*data mining*” (véase por ejemplo Frawley *et al.* 1991), donde se hacen las siguientes aseveraciones:

*“El descubrimiento de conocimiento en bases de datos es el proceso no trivial de identificar patrones en datos que sean válidos, novedosos, potencialmente útiles y, por último, comprensibles”* (Fayyad 1996, p. 21)

*“... data mining se refiere al acto de extraer patrones o modelos a partir de los datos ...”* (Fayyad 1996)

Una representación frecuente de un proceso típico de KDD, contempla los siguientes nueve pasos (véase por ejemplo Fayyad 1996, p. 23)

1. Desarrollar una comprensión del dominio de la aplicación
2. Crear un conjunto de datos objetivo
3. Limpieza y pre-procesamiento de los datos
4. Reducción y transformación de los datos
5. Elegir la tarea de *data mining*
6. Elegir los algoritmos de *data mining*
7. *Data mining*
8. Evaluar el resultado del *data mining*
9. Consolidar el conocimiento descubierto

En el paso 2 hay que seleccionar un sub-conjunto de los datos disponibles que normalmente están almacenados en una base de datos o en un *Data Warehouse* (bodega de datos). En una base de datos como sistema tradicional de almacenamiento

se puede captar datos de transacciones de un negocio. En cambio un *Data Warehouse* como sistema analítico apoya la toma de decisiones ofreciendo varias vistas distintas a los datos almacenados (Anahory y Murray 1997). Ambos sistemas sirven como fuente de un conjunto de datos objetivo.

En muchas aplicaciones reales los datos en una base de datos o también en un *Data Warehouse* contienen varias formas de errores, como por ejemplo valores faltantes, valores fuera del rango del atributo considerado o valores no plausibles (Guynes et al. 1996). Esa razón hace necesario el paso 3 de limpieza y pre-procesamiento de los datos (Famili et al. 1997).

En el paso 5 se elige la tarea de *data mining* específica para un análisis más detallado. Aquí se debe decidir si el propósito es, por ejemplo, la agrupación de objetos, la regresión o el modelaje de dependencia. Sobre la base de esta decisión, los más importantes algoritmos de *data mining* deben ser seleccionados (paso 6), los que se usan en la búsqueda real de patrones en los datos (paso 7).

Este trabajo presenta una aplicación del proceso de KDD para un banco donde, en el paso 5, han sido elegidas las tareas de segmentación de clientes y análisis de respuesta. El propósito de la segmentación de clientes es la agrupación (*clustering*) de los clientes, mientras que el análisis de respuesta apunta al modelaje de la dependencia existente entre las medidas de marketing y el comportamiento de respuesta posterior. En la aplicación considerada en el presente trabajo, se ha realizado la segmentación de clientes utilizando la llamada agrupación difusa (*fuzzy clustering*), haciendo uso además de técnicas de redes neuronales y selección automática de atributos, para aumentar las tasas de respuesta.

---

### 3. Fundamentos de Tecnologías Inteligentes

---

En las siguientes sub-secciones se presenta tecnologías difusas y redes neuronales, respectivamente. Ambos métodos se utilizan para el *data mining* en el proyecto descrito en esta publicación.

#### **3.1 Tecnologías Difusas**

La idea básica de la teoría de conjuntos difusos (*fuzzy set theory*) es que a diferencia de la teoría clásica de conjuntos un elemento puede ser miembro de un conjunto con un grado de pertenencia que normalmente es un número entre 0 y 1 (Zadeh 1965). La noción de un conjunto difuso se define de la siguiente manera:

Sea  $X$  un conjunto convencional de elementos. Un conjunto difuso se define como  $\tilde{A}$  sobre  $X$  de tal manera que:

$$\tilde{A} := \{(x, \mu_{\tilde{A}}(x)) ; x \in X\}$$

La expresión  $\mu_{\tilde{A}}(x)$  se denomina grado de pertenencia del elemento  $x$  al conjunto difuso  $\tilde{A}$ , llamándose  $\mu_{\tilde{A}}$  la función de pertenencia, que normalmente asume valores entre 0 y 1.

A base de estas definiciones elementales de un conjunto difuso, en el pasado se han desarrollado muchas teorías y aplicaciones en el área de lógica difusa, un resumen de las cuales se puede encontrar en (Zimmermann 1996). Un método difuso, de especial interés para el presente trabajo, es el algoritmo de agrupación *fuzzy c-means*. Una presentación detallada del *fuzzy c-means* se encuentra en (Bezdek 1981), y las aplicaciones de la agrupación difusa se describen, por ejemplo, en (Meier et al. 1994; Strackeljan, Weber 1999).

El algoritmo *fuzzy c-means* asigna un conjunto de objetos, caracterizados por sus respectivos valores de atributos, a un número determinado de clases. Como resultado del *fuzzy c-means*, cada objeto tiene un grado de pertenencia a cada clase, representada por su centro de clase. Básicamente, el algoritmo *fuzzy c-means* se realiza aplicando los siguientes cuatro pasos (para más detalles véase, por ejemplo, Bezdek et al. 1999):

### Paso 1: Inicialización

Utilizamos la siguiente notación

- número de clases a encontrar:  $c$
- número de objetos a agrupar:  $J$
- vector de atributos del objeto  $j$ :  $x_j, \quad j = 1, \dots, J$
- grado de pertenencia del objeto  $j$  a clase  $i$ :  $\mu_{ij}, \quad i = 1, \dots, c; j = 1, \dots, J.$

Sea  $\Theta^{(0)}$  una matriz ( $c \times J$ ) con el elemento  $\mu_{ij}$  en posición  $(i,j), i = 1, \dots, c; j = 1, \dots, J.$

Esta matriz se inicializa en forma aleatoria con la siguiente restricción:

$$\sum_{i=1}^c \mu_{ij} = 1, \quad \forall j = 1, \dots, J$$

### Paso 2: Cálculo de Centros de Clase

Dados los valores de pertenencia  $\mu_{ij}$ , los centros  $v_i$  de cada clase  $i$  están dados por:

$$v_i = \frac{\sum_{j=1}^J (\mu_{ij})^m x_j}{\sum_{j=1}^J (\mu_{ij})^m}, \quad \forall i = 1, \dots, C$$



El parámetro  $m$ , utilizado en la formula anterior, se llama difusor (*fuzzifier*) y determina el grado de difusión (*fuzziness*) para las clases encontradas ( $1 < m < \infty$ ). Para  $m$  “cercano a 1” se calcula una solución con clases no-difusas (*crisp*); mientras mayor sea  $m$  más difusa se hace la solución.

### Paso 3: Actualización de valores de pertenencia

Dados los centros de clase calculados en el paso 2, los valores de pertenencia  $\mu_{ij}$  son actualizados utilizando la siguiente fórmula:

$$\mu_{ij} = \left( \sum_{k=1}^c \left( \frac{d_{kj}}{d_{ij}} \right)^{\frac{2}{m-1}} \right)^{-1} \quad \forall i = 1, \dots, c \\ \forall j = 1, \dots, J$$

El valor  $d_{ij}$  es la distancia entre el objeto  $j$  y el centro de clase  $i$  ( $v_i$ ). En el cálculo de esta distancia se utilizan los centros de clase  $v_i$  obtenidos en el paso 2.

### Paso 4: El Criterio de Detención (*Stopping*)

Los pasos 2 y 3 se repiten en forma iterativa hasta cumplir con el siguiente criterio de detención:

$$\|\Theta^{(t+1)} - \Theta^{(t)}\| \leq \epsilon$$

donde  $\Theta^{(t)}$  es la matriz de los valores de pertenencia en la iteración  $t$  y  $\epsilon$  es un umbral a ser determinado por el usuario. Las siguientes dos condiciones deben cumplirse para asegurar la convergencia del algoritmo *fuzzy c-means*:

$$\sum_{i=1}^c \mu_{ij} = 1, \quad \forall j = 1, \dots, J$$

$$\mu_{ij} \in [0, 1]; \quad i = 1, \dots, c; \quad j = 1, \dots, J$$

En términos de resultados, el algoritmo *fuzzy c-means* rinde centros de clase  $v_i$  para las  $c$  clases, así como los valores de pertenencia de cada objeto a cada clase  $\mu_{ij}$ .

### Evaluación de las clases encontradas

En la literatura se han propuesto los siguientes parámetros para evaluar la estructura de clase encontrada por la agrupación difusa (Windham 1981):

coeficiente de partición (cp):

$$\text{cp} = \sum_{j=1}^J \sum_{i=1}^C \frac{(\mu_{ij})^2}{J}$$

entropía de partición (ep):

$$\text{ep} = -\frac{1}{J} \sum_{j=1}^J \sum_{i=1}^C \mu_{ij} \log_e \mu_{ij}$$

exponente proporcional (exp):

$$\text{exp} = -\log_e \prod_{j=1}^J \sum_{i=1}^{m_k} (-1)^{i+1} \binom{c}{i} (1 - \mu_{ij})^{c-1}$$

El valor  $m_k$  se define como el entero más grande que sea menor o igual a  $\frac{1}{\mu_k}$ , donde

$$\mu_k := \max \{ \mu_{ik}; i = 1, \dots, c \}$$

Las siguientes relaciones se cumplen para los parámetros definidos anteriormente:

$$\frac{1}{c} \leq \text{cp} \leq 1 \quad 0 \leq \text{ep} \leq \log_e c \quad 0 \leq \text{exp} \leq \infty$$

En la aplicación de la segmentación de clientes que se describe en el capítulo 4.1, los clientes son los objetos considerados, que se describen mediante atributos tales como por ejemplo edad e ingreso. En este capítulo se mostrará también cómo los parámetros de evaluación pueden ser usados para encontrar un número adecuado de clases.

### **3.2 Redes Neuronales**

El término “red neuronal” (NN, por *Neural Network*) se usa para denotar modelos matemáticos de las funciones del cerebro humano, que pretenden expresar las propiedades del procesamiento paralelo masivo y de la representación distribuida existente en el cerebro. Las redes neuronales artificiales aprenden a partir de la experiencia, y es esta característica la que ha suscitado el interés actual en estos métodos. El aprendizaje ofrece una poderosa alternativa a la programación. Se han propuesto diferentes estrategias de aprendizaje para redes neuronales tales como el aprendizaje supervisado y no-supervisado (Rojas 1996).

Un ejemplo de los paradigmas de aprendizaje supervisado es el algoritmo de retropropagación (*backpropagation algorithm*) (Rojas 1996). En el aprendizaje supervisado la red se enfoca en un conjunto de entrenamiento representativo que consiste en pares vectoriales. Uno de los vectores se aplica a la entrada de la red, el otro se usa de objetivo, representativo de la salida deseada. El entrenamiento se logra al ajustar los pesos de las conexiones de la red para minimizar la diferencia entre las salidas deseadas y las salidas computadas.

El aprendizaje no supervisado, muchas veces llamado auto-organización, solamente requiere de vectores de entrada para adiestrar la red. Durante el proceso de entrenamiento, las ponderaciones de la red se ajustan a fin de que las entradas similares produzcan salidas similares. Esto se logra mediante el algoritmo de entrenamiento que extrae regularidades estadísticas desde el conjunto de entrenamiento. Los mapas de atributos auto-organizadores (*self-organizing feature maps*) introducidos por Kohonen pertenecen a la clase de algoritmos de aprendizaje no-supervisado (Kohonen 1995).

En la mayoría de los casos, una estructura bi-dimensional (mapa de atributo) se usa para redes Kohonen, sin embargo, también es posible encontrar estructuras uni-dimensionales (cadenas) o tri-dimensionales (cubos). Cada neurona tiene un vector referencial, que consiste en las ponderaciones  $\mu_{ik}$  entre las neuronas  $i$  y el vector de entrada  $k$  ( $i=1, \dots, N$ ;  $k=1, \dots, M$ ), donde  $N$  es el número de neuronas y  $M$  es el número de entradas. A continuación se presenta el algoritmo de Kohonen, así como sus parámetros de afinamiento.

1. Inicialización de todas las ponderaciones  $\mu_{ik}$  con valores aleatorios
2. Presentación al mapa de un vector de entrada  $e = (e_1, \dots, e_M)$
3. Cálculo del estado de excitación de cada neurona dado el vector de entrada  $e$
4. Búsqueda de la neurona ganadora (neurona con mayor excitación): neurona  $j$
5. Adaptación de las ponderaciones neuronales  $\mu_{ik}$ , de acuerdo a la siguiente formula:

$$\mu_{ik}^{new} = \mu_{ik}^{old} + \epsilon \cdot r_{ij} (e_k - \mu_{ik}^{old}) \quad i, j = 1, \dots, N, k = 1, \dots, M$$

6. Volver al paso 2.

donde,

$\mu_{ik}$ : ponderaciones neuronales;  $i = 1, \dots, N$ ;  $k = 1, \dots, M$

$r_{ij}$ : valores de retroalimentación entre la neurona ganadora (neurona  $j$ ) y la neurona  $i$  ( $i = 1, \dots, N$ )

$\epsilon$ : tasa de aprendizaje

Los valores de retroalimentación indican dependencias geométricas (*neighborhood dependencies*) entre las neuronas. Estas se calculan mediante una función gaussiana:

$$r_{ij} = e^{-\left(\frac{d^2}{\sigma^2}\right)}$$

donde

$\sigma$ : radio de aprendizaje

$d$ : distancia euclidiana entre la neurona ganadora (neurona  $j$ ) y la neurona  $i$ .

La tasa de aprendizaje  $\epsilon$ , junto con el radio de aprendizaje  $s$ , debe ser determinado por el usuario.

En el gráfico siguiente se presenta un mapa de atributo bi-dimensional con  $N$  neuronas. Entre la neurona  $i$  y el valor de atributo  $k$  de un objeto, se considera la ponderación  $\mu_{ik}$  ( $i = 1, \dots, N$ ;  $k = 1, \dots, M$ ).

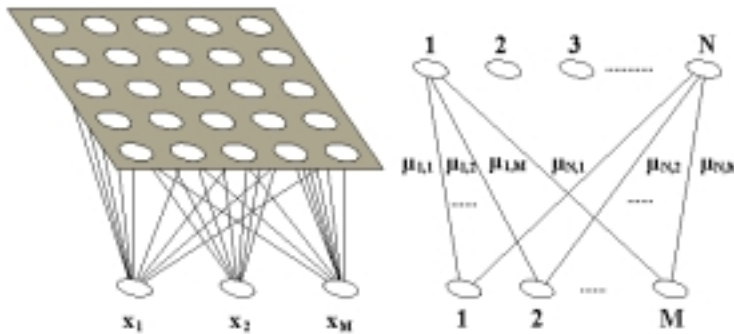


Gráfico 1: Mapa de atributo de Kohonen

En la aplicación descrita en el capítulo 4.2, se han usado mapas de atributos auto-organizadores para analizar las respuestas a *mailings*.

## 4. Aplicaciones de Data Mining en Instituciones Financieras

---

La siguiente sub-sección describe una aplicación de la agrupación difusa para la segmentación de clientes. Basándose en los segmentos encontrados, se aplican redes neuronales para un análisis de respuesta que está descrito en la sub-sección 4.2.

### **4.1 Segmentación de clientes utilizando la agrupación difusa**

La segmentación de clientes tuvo su origen en el campo de marketing e investigación de mercados, y entrega una división analítica de todos los potenciales clientes de un mercado de venta de acuerdo a diferentes criterios. Esto resulta en la formación de grupos de clientes, o segmentos de ellos, internamente homogéneos y externamente heterogéneos, planteando así actividades de marketing enfocadas a estos segmentos.

Tanto las instituciones financieras, como muchos departamentos comerciales, se han dado cuenta que un enfoque decisivo para el mejoramiento de la calidad de consulta y servicios reside en la segmentación de clientes en distintos grupos objetivos conformados por clientes similares (Angstenberger et al. 1998). Los productos y los servicios correspondientes a un segmento de clientes específico, deben ser entregados en la forma y al nivel en que el cliente perteneciente a este segmento espera y está dispuesto a pagar. Para aplicar tal enfoque se requiere conciencia de la relevancia de la segmentación orientada al cliente y la identificación de segmentos de clientes distintos mediante el uso de métodos adecuados (Weber 1996).

El análisis de agrupación tradicional no-difusa (*crisp cluster analysis*) es el método más usado tanto para la segmentación de mercado (Hruschka 1986), como para la segmentación de clientes.

Mediante el uso del análisis de agrupación, se descubrirán ciertos segmentos de clientes (del conjunto entero de todos los clientes). Cada cliente será asignado a una determinada agrupación, en forma concreta (no difusa) y completa. Estas agrupaciones se encuentran separadas la una de la otra por umbrales claros y precisos. Si los valores de atributos de cierto cliente sobrepasan estos umbrales, el cliente será asignado a otra agrupación. Si los valores de atributos de un cliente alcanzan exactamente un umbral determinado, este cliente es asignado opcionalmente a una u otra agrupación en forma completa, aún cuando él o ella pertenecería en el mismo grado en ambas agrupaciones.

Lo anterior revela las limitaciones del análisis tradicional de agrupaciones. La aplicación de técnicas de agrupación difusa aquí presentadas fue realizada para un banco alemán, cuyos productos eran, por ejemplo, cuentas corrientes, tarjetas de crédito y esquemas de inversión. La ventaja de la agrupación difusa en esta aplicación es que los clientes son asignados a diferentes segmentos con diferentes grados de pertenencia, lo que entrega un mejor conocimiento del perfil de pertenencia de cada cliente y, por lo tanto, sus necesidades en relación con los productos ofrecidos.

En el proyecto descrito en esta publicación, los datos específicos de clientes proporcionados por un banco alemán han sido analizados para la segmentación. En primer lugar, se seleccionarán los atributos para la segmentación y se analizarán en busca de correlaciones.

La selección de los atributos relevantes para la segmentación es de la mayor importancia para el análisis de agrupación, ya que este paso determina en forma definitiva qué criterios llevan a la agrupación final de los datos de clientes. Existe una enorme variedad de atributos disponibles, clasificables en atributos demográficos, tales como edad, género y situación familiar; y atributos socioeconómicos como educación, profesión, ingreso y propiedad. Los atributos finalmente seleccionados deberían mostrar un potencial discriminatorio suficiente sin correlación. Más aún, deben ser cuantificables (para la aplicación de métodos matemáticos), debiendo tener también el mismo nivel de escala. En el cuadro siguiente se presenta una matriz con los cinco atributos seleccionados en sus columnas. Cada fila representa un cliente. La base de datos completa, que se ha investigado, contiene alrededor de 300.000 clientes.

<b>Código cliente</b>	<b>Edad (años)</b>	<b>Ingreso (\$ por mes)</b>	<b>Propiedades (\$)</b>	<b>Crédito (\$)</b>	<b>Margen de contribución (\$ por año)</b>
101126	35	681,0	0,0	-1177,0	-5,0
607226	60	2218,0	6500,0	0,0	131,0
273974	53	537,0	6577,0	0,0	177,0
827866	56	4397,0	21333,0	0,0	-50,0
626508	38	2813,0	6667,0	0,0	-32,0

**Cuadro 1: Atributos seleccionados para algunos clientes**

El análisis de correlación debería comprobar que los atributos seleccionados son en gran medida independientes y que ningún atributo será sobreponderado en el análisis de agrupación.

	Edad	Ingreso	Propiedades	Crédito	Margen de C.
<i>Edad</i>	1	0,274	0,330	-0,024	0,282
<i>Ingreso</i>	0,274	1	0,437	-0,305	0,194
<i>Propiedades</i>	0,330	0,437	1	0,025	0,336
<i>Crédito</i>	-0,024	-0,305	0,025	1	-0,213
<i>Margen de C.</i>	0,282	0,194	0,336	-0,213	1

Cuadro 2: Resultado del análisis de correlación

Por otra parte, la normalización hace que los datos sean comparables, ya que originalmente los datos pueden tener diferentes dimensiones y diferentes rangos de escala, llevando (sin la normalización) a diferentes ponderaciones de atributos en el proceso de agrupación.

### Resultados y beneficios del análisis de datos difuso para la segmentación de clientes

Tras haber determinado los parámetros para el algoritmo *fuzzy c-means*, se modifica el número de agrupaciones. Para cada número de agrupaciones entre  $c = 2$  y  $c = 10$ , se aplica el algoritmo *fuzzy c-means*. Los valores de pertenencia de todos los clientes a las agrupaciones 1 a  $c$  son calculados y presentados. Adicionalmente, se entrega el centro de todas las agrupaciones (clases)  $c$ .

Clase	Edad	Ingreso	Propiedades	Crédito	Margen de C.
A	32,8	1.946,92	6.315,78	-4.509,91	21,92
B	59,28	1.951,87	9.518,03	-3.667,27	62,94
C	47,58	3.905,84	29.317,29	-13.816,90	171,15
D	10,45	135,03	2.607,43	-467,65	6,18
E	75,49	1.552,54	21.957,89	-1.983,58	203,71
F	41	3.921,11	12.661,52	-8.144,57	68,48

Cuadro 3: Centros de clases para  $c = 6$  agrupaciones

Las medidas de validez de agrupación, descritas anteriormente en la Sección 3.1, se usan para evaluar diferentes soluciones de agrupación. Estas medidas de validez de agrupación dan una pista matemática del “mejor” número de agrupaciones que pueden ser encontradas. Usando por ejemplo la entropía de partición (ep) el número más adecuado se determina como 6 en la aplicación demostrada (véase gráfico 2). Los resultados de la segmentación de clientes, descritos aquí, son verificados mediante entrevistas posteriores con expertos de la industria bancaria involucrados en este proyecto.

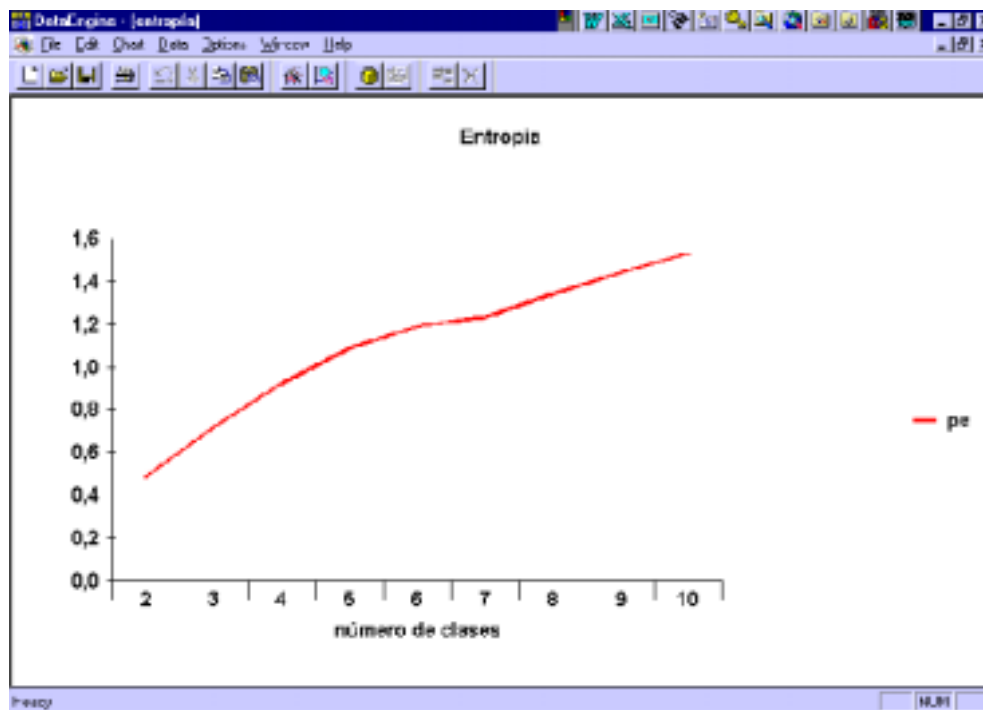


Gráfico 2: Entropía de partición para c=2 a c=10 agrupaciones

A diferencia del enfoque o perspectiva estática resultante de las técnicas tradicionales de agrupación, el análisis de agrupación difusa proporciona grados continuos de pertenencia de los objetos a las agrupaciones. Para cada cliente, se calcularán los valores de pertenencia en todas las clases, llevando a una posición más diferenciada de cada cliente entre las clases. Por ejemplo, los clientes que muestran valores característicos de diferentes segmentos se tratarán con actividades de marketing de acuerdo a sus valores de pertenencia. Como resultado, el experto de marketing del instituto financiero es capaz de ofrecer a sus clientes productos más acordes y los servicios correspondientes.

Además de esto, la agrupación difusa proporciona ventajas en cuanto a los aspectos dinámicos (Joentgen et al. 1999; Weber 1999). La evolución de un cliente a lo largo del tiempo se reconocerá en forma inmediata, a base del cambio de los valores de pertenencia a diferentes clases. Investigaciones periódicas de los datos



de clientes pueden ofrecer pistas precoces del movimiento de un cliente entre las clases y, como consecuencia, para actividades de marketing apropiadas.

## **4.2 Análisis de Respuesta**

Para uno de los segmentos de clientes encontrados, se ha realizado un análisis de respuesta a fin de incrementar la tasa de respuesta de las campañas de marketing realizadas para promocionar ciertos productos. En este caso específico, se consideran los *mailings* que promocionan las tarjetas de crédito.

En el banco correspondiente a este proyecto, la selección de personas a ser incluidas en un *mailing* se ha realizado hasta ahora en forma manual, configurando un perfil del cliente prototipo, y comparando cada registro de la base de datos con este prototipo mediante una consulta a la base de datos (*database query*). Aquellas personas que cumplen con los requerimientos pre-determinados son incluidas en el *mailing*. La composición manual del perfil a ser igualado introduce cierta arbitrariedad al proceso. Tomando en consideración la cantidad de información disponible sobre cada cliente, resulta evidente que es un problema no-trivial establecer un *query* para determinar clientes como, por ejemplo:

```
SELECT "Cliente"          FROM "Base de Clientes"  
  
WHERE "Edad < 25 AND Trabajo = Empleado AND Monto de crédito < 5.000"
```

Sería útil establecer automáticamente el perfil del cliente prototipo. A continuación se muestra cómo una red neuronal ha sido aplicada para realizar esta tarea. Se describe la elaboración de un clasificador neuronal utilizado para extraer de una base de datos aquellas personas que con más probabilidad adquirirán una tarjeta de crédito en respuesta a un *mailing*.

El análisis se basa en aproximadamente 9.000 registros de clientes seleccionados desde un sub-segmento especial denominado "jóvenes adultos". Los registros de datos incluyen datos personales, como edad, género y trabajo, así como información detallada sobre el uso de productos del cliente. Fueron considerados 18 productos distintos, desde cuentas corrientes y de ahorro, hasta planes de inversión y valores. En forma adicional, se dispuso de una clasificación del lugar de domicilio de cada persona incluida en la base de datos. En total, alrededor de 180 campos de datos (atributos) se encontraban disponibles para cada cliente en la base.

A partir de un *mailing* anterior, la respuesta de aproximadamente 2.500 personas ya era conocida (en términos de la compra o no del producto publicitado). La tasa de respuesta a este *mailing* anterior, calculado como la razón entre el número de compradores reales que responden al *mailing* y el volumen del mismo, fue de un 8,2%.

El análisis estadístico preliminar de los datos reveló que las distribuciones de los campos individuales de datos de los compradores eran muy similares a las de

los no-compradores. Por tanto, el perfil de un comprador típico no se pudo establecer a partir de las distribuciones estadísticas. Lo que significa que no era suficiente observar atributos aislados del cliente. Claramente, la configuración del perfil del cliente tenía que considerar la combinación de distintos *ítems* de información sobre el cliente. Sin embargo, la naturaleza de esta combinación de información era desconocida, además de qué información habría que incluir realmente.

Por esta razón, se decidió establecer la relación entre el perfil del cliente por un lado y su tendencia a comprar un producto por otro lado usando una red neuronal, ya que ésta tiene la capacidad de representar el tipo de relación no-lineal y no-homogénea requerida. Sin embargo, a fin de usar sólo los atributos más importantes, el siguiente paso del procedimiento de análisis consistió en reducir el espacio de atributos considerados. Esto también debería entregar una indicación de los atributos que realmente son relevantes.

En el Gráfico 3 se presenta un esbozo del procedimiento de análisis recién descrito. Los atributos más relevantes se seleccionan haciendo uso de una modificación del método descrito en Behr et al. (1997). Este procedimiento tiene la capacidad de considerar inicialmente un espacio de atributos de muchas dimensiones, para luego seleccionar las combinaciones de atributos que dan las tasas de re-clasificación más altas posibles. El procedimiento entrega un listado de combinaciones de atributos junto con una estimación de la tasa de re-clasificación para cada combinación.



Gráfico 3: Esquema del procedimiento de análisis global

La etapa de selección de atributos dio, como resultado, combinaciones de dos a cinco atributos, los que fueron analizados en forma más detallada. Las redes neuronales de tipo Kohonen (1995) (mapas de atributos auto-organizadoras) fueron aplicadas a un sub-conjunto de los datos de respuesta, siendo evaluadas mediante un conjunto distinto para cada una de las combinaciones de atributos seleccionados.

## **Resultados**

Los resultados fueron comparados con los resultados reales del *mailing*, calculándose las tasas de respuesta correspondientes. Estas se presentan en el siguiente cuadro.

<b>Numero de atributos</b>	<b>Tasa de respuesta</b>	<b>Tamaño del grupo objetivo seleccionado</b>
2	20.0%	350
3	18.3%	460
4	17.4%	530
5	17.6%	500

**Cuadro 4: Tasa de respuesta de los clasificadores neuronales**

Al comparar los resultados del análisis presentados en el Cuadro 4 con la tasa de respuesta del 8,2% obtenida anteriormente, resulta obvio que ésta fácilmente podría ser duplicada al realizar el proceso de selección en la forma descrita. Esto implicaría una reducción dramática de los costos del *mailing* y, al mismo tiempo, el volumen del *mailing* seleccionado por la red neuronal sería considerablemente menor que el seleccionado manualmente. Por lo tanto, se lograría un aumento de la eficiencia global de los *mailings*.

En la actualidad, el conocimiento adquirido en este análisis se está transfiriendo para lograr un mejor ajuste de los requisitos de otros productos distintos a la tarjeta de crédito considerada hasta ahora.

### **4.3 Herramientas de Software Utilizadas para Data Mining**

Las investigaciones recién descritas han sido llevadas a cabo utilizando la herramienta de software "DataEngine" un paquete de software que contiene métodos tales como tecnologías difusas y redes neuronales para el análisis inteligente de datos (MIT 2000). La combinación del pre-procesamiento, el análisis estadístico y los sistemas inteligentes para el diseño de un clasificador, junto con el modelaje del sistema lo convierte en un poderoso instrumento de software utilizable en un rango muy amplio de aplicaciones.

El siguiente gráfico da una visión global de la estructura del DataEngine. Su arquitectura abierta permite al usuario extender este programa de software mediante los llamados bloques de función definida por el usuario. Los modelos desarrollados con DataEngine fácilmente pueden ser integrados a otros paquetes de software y programas de aplicación utilizando la biblioteca DataEngine ADL.

Gráfico 4: Estructura de DataEngine

---

## 5. Conclusiones y Perspectivas Futuras

---

El presente trabajo entrega una breve presentación de enfoques de *data mining* y tecnologías inteligentes, así como sus aplicaciones al *database marketing* en un banco alemán. Los resultados son altamente prometedores, aunque basados en un pequeño conjunto de datos de clientes. El siguiente paso consistirá en aplicar los enfoques recién descritos a un conjunto de datos mayor, con el fin de validar su *performance*. Se pretende integrar los respectivos modelos de *data mining* al sistema de software utilizado para apoyar las actividades de marketing.

Las aplicaciones presentadas en la sección 4 pueden ser transferidas a diversas otras áreas en las que los clientes están siendo analizados, a fin de aumentar las ventas mediante el uso de estrategias optimizadas de marketing: por ejemplo, telecomunicaciones, empresas comerciales y la industria de ventas por catálogo (Bitran, Mondschein 1996). Entre otras aplicaciones parecidas donde el procedimiento de análisis aquí presentado puede ser aplicado, se incluye el área de detección de fraude. En vez de distinguir a los compradores de los no-compradores de la manera presentada en este trabajo, podría ser interesante diferenciar el uso fraudulento del uso correcto, por ejemplo, de tarjetas de crédito o teléfonos móviles. Tales aplicaciones quedan para investigaciones futuras.

---

## 6. Referencias bibliográficas

---

- Adriaans, P. y Zantinge, D. (1996): *Data Mining*. Addison-Wesley, Harlow.
- Anahory, S. y Murray, D. (1997): *Data Warehousing in the Real World*. Addison-Wesley, Harlow
- Angstenberger, J., Weber, R. y Poloni, M. (1998): "Data Warehouse Support to Data Mining: A Database Marketing Perspective". *Journal of Data Warehousing* 3, N° 1, 2-11.
- Behr, D., Kocher, T. y Strackeljan, J. (1997): "Fuzzy Pattern Recognition for Automatic Detection of Different Teeth Substances". *Fuzzy Sets and Systems* 85, 275-286.
- Bezdek, J.C. (1981): *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, Nueva York.
- Bezdek, J.C., Keller, J.M., Krishnapuram, R. y Pal, N.R. (1999): *Fuzzy models and algorithms for pattern recognition and image processing*. Kluwer, Boston Londres Dordrecht
- Bitran, G. R. y Mondschein, S. V. (1996): "Mailing Decisions in the Catalog Sales Industry." *Management Science* 42, N° 9, 1364-1381.
- Famili, A., Shen, W.-M., Weber, R., Simoudis, E. (1997): Data Preprocessing and Intelligent Data Analysis. *Intelligent Data Analysis* Vol. 1, N° 1, 3-23.
- Fayyad, U. M. (1996): "Data Mining and Knowledge Discovery: Making Sense out of Data." *IEEE Expert, Intelligent Systems & Their Applications*, Octubre 1996, 20-25.
- Frawley, W. J., Piatetsky-Shapiro, G. y Matheus, C. J. (1991): "Knowledge Discovery in Databases: An Overview". En: Piatetsky-Shapiro, G. y Frawley, W. J. (eds.): *Knowledge Discovery in Databases*. AAAI Press/MIT Press, Menlo Park, 1-27.
- Furness, P. (1994): "New Pattern Analysis Methods for Database Marketing; Part 1". *The Journal of Database Marketing* 1, N° 3, 220-232.
- Guynes, C. S., Prybutok, V. R. y Myers, B. L. (1996): "Evolving Data Quality Considerations for Client/Server Environments". *Data Quality* 2, No.1, 21-27.
- Hruschka, H. (1986): "Market definition and segmentation using fuzzy clustering methods." *International Journal of Research in Marketing* 3, 117-134.
- Joentgen, A., Mikenina, L., Weber, R. y Zimmermann, H.-J. (1999): Dynamic Fuzzy Data Analysis Based on Similarity Between Functions. *Fuzzy Sets and Systems* 105, 1, 81-90
- Kohonen, T. (1995): *Self-Organizing Maps*. Springer-Verlag, Berlin Heidelberg.
- Meier, W., Weber, R. y Zimmermann, H.-J. (1994): "Fuzzy Data Analysis - Methods and Industrial Applications." *Fuzzy Sets and Systems* 61, 19-28.
- MIT (2000): *DataEngine 3.1 - Manual*. Aachen, Alemania.
- Nash, E. L. (1993): *Database Marketing - The Ultimate Marketing Tool*. McGraw-Hill, Nueva York.
- Rojas, R. (1996): *Neural Networks - A Systematic Introduction*. Springer-Verlag, Berlin Heidelberg.
- Strackeljan, J. y Weber, R. (1999): "Quality Control and Maintenance". En: H.-J. Zimmermann (ed.): *Practical Applications of Fuzzy Technologies, Vol. 7 of The Handbooks of Fuzzy Sets Series*, D. Dubois, H. Prade (series editors), Kluwer Academic Publishers, Boston, Londres, Dordrecht, 161-184

- Weber, R. (1996): "Customer Segmentation for Banks and Insurance Groups with Fuzzy Clustering Techniques". En: J. F. Baldwin (ed.): *Fuzzy Logic*, John Wiley and Sons, Chichester, 187 – 196.
- Weber, R. (1999): Dynamic Data Mining with Functional Fuzzy Clustering. *Actas de Resúmenes Extendidos, III Congreso Chileno de Investigación Operativa - OPTIMA '99*, Arica, Chile, 13-15 de octubre de 1999, 217-224
- Windham, M. P. (1981): "Cluster Validity for Fuzzy Clustering Algorithms". *Fuzzy Sets and Systems* 5, 177-185.
- Zadeh, L. A. (1965): "Fuzzy Sets". *Information and Control* 8, 338-353.
- Zimmermann, H.-J. (1996): *Fuzzy Set Theory - and Its Applications*. 3er edición. Kluwer Academic Publishers, Boston Dordrecht, Londres.

---

# Control de la Contaminación Atmosférica de las Fundiciones Estatales de Cobre Mediante un Sistema de Apoyo a las Decisiones

---

**René Caldentey\***

**Susana Mondschein\*\***

Departamento de Ingeniería Industrial,  
Universidad de Chile

---

## Resumen

---

*En este artículo desarrollamos un sistema de apoyo para las decisiones de inversión en el control de la polución atmosférica y para las decisiones operacionales en la industria del cobre. Este sistema consiste en (i) un modelo entero no-lineal para optimizar la operación en las fundiciones de cobre, incluyendo decisiones de inversión tanto en capacidad de fundición como en plantas de abatimiento de contaminantes para el cumplimiento de las normas medioambientales, y (ii) un modelo de flujo en redes que describe el comportamiento económico del mercado del ácido sulfúrico, el cual considera el ácido sulfúrico producido en las etapas de abatimiento de contaminantes en el proceso de fundición. Este segundo modelo resuelve el equilibrio entre mercados espacialmente separados para determinar el precio y la distribución de ácido en cada región de oferta y demanda.*

*Estos dos modelos interactúan a través de la información que cada uno recibe del otro. De esta manera, el modelo de la fundición usa el precio del ácido sulfúrico en cada fundición para determinar las decisiones óptimas operacionales y de inversión. A su vez el modelo de mercado de ácido sulfúrico considera la producción total de ácido sulfúrico en las fundiciones como parte de la oferta para encontrar el precio del producto en cada fundición. La solución dada por el sistema de apoyo a las decisiones es la de un equilibrio global obtenido cuando este proceso iterativo entre los dos modelos converge. De esta manera, el precio del ácido*

---

\* E-mail: rcaldent@dii.uchile.cl

\*\* Dirección: Departamento de Ingeniería Industrial, Universidad de Chile, Casilla 2777, Santiago, Chile.  
E-mail: smondschi@dii.uchile.cl

*sulfúrico, el cual es un elemento central en el momento de decisión del cuándo y dónde localizar una planta de ácido sulfúrico, está determinado endógenamente, en vez de exógenamente como en la mayoría de los modelos de este tipo. Experimentos computacionales muestran que las ganancias esperadas de la industria del cobre pueden incrementarse significativamente cuando el problema se resuelve considerando en forma conjunta todas las fundiciones estatales de cobre, en comparación a la solución en la cual cada fundición toma sus decisiones de manera independiente.*

---

## 1 Introducción

---

La industria del cobre es una de las industrias más importantes de Chile, correspondiendo aproximadamente a un 38% de las exportaciones totales del país. En los últimos cien años, la producción de cobre ha aumentado significativamente, pasando de 21 toneladas al año en 1897 a 828.300 en 1975 y 3.115.800 toneladas al año en 1996.

No obstante este gran crecimiento, es sólo a partir de 1980 que el gobierno chileno comienza a estudiar los efectos medioambientales de la producción de cobre. Esto llevó a que en 1991 se establecieran nuevas regulaciones medioambientales para las emisiones de anhídrido sulfuroso y material particulado, dos de los contaminantes más importantes emitidos a la atmósfera durante el proceso que transforma el concentrado a cobre refinado. Una nueva regulación medioambiental para las emisiones de arsénico está en vigencia desde 1999. Actualmente, uno de los desafíos más importantes que se le presenta a la industria minera estatal es como resolver el problema de la contaminación en las fundiciones de cobre.

Existen siete fundiciones en el país, dos privadas y cinco estatales. Las estatales son Chuquicamata, Caletones, Potrerillos, Ventanas, y Paipote, las que reciben concentrado para ser transformado en cobre refinado desde 24 plantas concentradoras, siete de las cuales son de propiedad estatal. Actualmente, ninguna de estas fundiciones estatales satisface plenamente las regulaciones medioambientales, aunque ya han presentado planes de limpieza, *planes de limpieza de base*, que eventualmente asegurarán el cumplimiento de las nuevas regulaciones medioambientales. Estos planes ya están siendo implementados y han completado algunas de las etapas con éxito. Cada fundición desarrolló su propio plan de limpieza individualmente, sin considerar los efectos de sus decisiones sobre las otras fundiciones. El costo total esperado de estos planes es de aproximadamente US\$1 billón, presupuesto que está asociado principalmente a la construcción de plantas de ácido sulfúrico y precipitadores electrostáticos – dos de las tecnologías más usadas para el abatimiento de contaminantes emitidos a la atmósfera durante el proceso de fundición. Las plantas de ácido sulfúrico transforman el anhídrido sulfuroso en ácido sulfúrico, haciendo que éste sea uno de los subproductos más importante del proceso de fundición de cobre.



Las grandes inversiones requeridas por las fundiciones estatales para llevar a cabo sus *planes de limpieza de base*, llevan a la Comisión Chilena del Cobre (Cochilco) en 1994 a desarrollar un modelo preliminar para optimizar las decisiones de inversión en plantas de abatimiento de contaminantes considerando las cinco fundiciones estatales como un sistema integrado. Existen cuatro características principales que hacen que la solución que considera a las cinco fundiciones estatales como un sistema único sea mejor a la obtenida considerando las fundiciones independientemente: (i) economías de escala asociadas a la construcción de plantas de ácido sulfúrico, precipitados electrostáticos y plantas de fundición, (ii) la redistribución de los concentrados entre las fundiciones con el fin de minimizar la contaminación (el contenido de azufre y arsénico de los concentrados varía según el lugar de origen, y por ende el grado de contaminación que emite al ser procesado), (iii) la decisión conjunta sobre la distribución de ácido sulfúrico hacia las zonas de demanda y (iv) el efecto de la producción total de ácido sulfúrico en los precios de éste.

La función objetivo de este modelo de programación entera mixta es maximizar la ganancia total esperada del negocio de producción de cobre del sector público, descontada a lo largo del horizonte de planificación, considerando restricciones técnicas, medio ambientales y de mercado. Una descripción completa del modelo puede ser encontrada en Mondschein y Schilkrut (1997). Dadas las importantes diferencias entre las soluciones entregadas por este modelo preliminar y aquellas propuestas en los *planes de limpieza de base*, Cochilco le pidió a los autores de este artículo desarrollar una versión refinada de este modelo que incluyera otras características importantes no consideradas en el primer sistema de apoyo a las decisiones.

En este artículo desarrollamos un nuevo modelo matemático, que incluye tres importantes características que no habían sido consideradas en el modelo preliminar: (i) un modelo económico para los precios del ácido sulfúrico en función de la oferta del producto, (ii) restricciones a la emisión de arsénico y (iii) la consideración explícita de la heterogeneidad de la composición mineralógica del concentrado, en términos de azufre y arsénico, la cual depende principalmente del origen del concentrado. Estas nuevas consideraciones incrementan significativamente la complejidad del modelo, el cual pasa a ser un modelo entero no-lineal que describe la operación de las fundiciones de cobre, incluyendo las decisiones de inversión en plantas de abatimiento de contaminantes y un modelo de flujo en redes que describe el comportamiento económico del mercado de ácido sulfúrico. El modelo del mercado del ácido sulfúrico resuelve el equilibrio entre mercados espacialmente distantes, para determinar el precio y la distribución de ácido en las regiones de demanda y oferta. Estos dos modelos interactúan a través de la información que cada uno recibe del otro.

El resto del artículo se organiza de la siguiente manera: en la sección 2 entregamos una breve descripción del proceso de producción de cobre y del mercado de ácido sulfúrico, en la 3 sección se presentan los modelos matemáticos que describen la operación de las fundiciones y el mercado del ácido sulfúrico. En la sección 4 describimos la resolución y resultados más importantes del modelo. Finalmente en la sección 5 presentamos las conclusiones.

---

## 2 Descripción del problema

---

### **2.1. Proceso de Producción del cobre**

---

El mineral extraído de las minas puede ser clasificado como sulfuro u óxido dependiendo de sus características mineralógicas. Los sulfuros son procesados en plantas concentradoras para producir concentrado de cobre. Aproximadamente el 99% de los minerales de sulfuro son procesados en plantas localizadas cerca de las minas de donde fueron extraídos. Solamente una pequeña fracción es extraída por pequeños empresarios y vendida directamente a las plantas concentradoras. Aunque la composición mineralógica de los concentrados depende directamente de su origen, ésta corresponde aproximadamente a un tercio de cobre, un tercio de azufre y un tercio de hierro. Estos concentrados pueden ser exportados como tal o tratados en fundiciones para producir cobre refinado. El proceso de fundición produce gases con contenido de anhídrido sulfuroso, material particulado y arsénico, para los cuales se han establecido nuevas regulaciones ambientales. El abatimiento de estos contaminantes corresponde al foco central de nuestro estudio.

El método comúnmente utilizado para obtener cobre refinado a partir de los óxidos es el proceso de lixiviación, el cual utiliza ácido sulfúrico como su principal insumo. Este proceso no emite anhídrido sulfuroso hacia la atmósfera.

### **Plantas Concentradoras**

Previo al tratamiento de los sulfuros en las fundiciones de cobre, ellos deben pasar por un proceso de concentración, debido a que generalmente éstos tienen una muy baja concentración de cobre. Este proceso es generalmente llevado a cabo cerca de las minas: el mineral es triturado, molido y sometido a un proceso de flotación. El concentrado producido en las plantas concentradoras puede ser exportado directamente o procesado en las fundiciones de cobre para producir cobre refinado. Las principales decisiones en esta etapa son cuánto concentrado exportar y cuánto procesar en el país, y qué fundiciones deberían procesar el concentrado que va a ser tratado a nivel doméstico. El volumen de concentrado producido anualmente durante el horizonte de planificación es determinado exógenamente al modelo, y es estimado periódicamente por los expertos de Cochilco.

### **Fundiciones de Cobre**

En esta etapa, el concentrado es sometido a un proceso de fusión en un horno (convertidor El Teniente, horno Flash u otra tecnología) para producir "mata". Posteriormente esta mata es procesada en convertidores (por ejemplo convertidor Pierce-Smith) para producir cobre refinado o blister. En el país existen cinco

fundiciones de cobre estatales con una capacidad total de aproximadamente 4 millones de toneladas de concentrado por año. Las principales decisiones en esta etapa del proceso productivo corresponden a la capacidad de fundición en cada período del horizonte de planificación y la tecnología a utilizar en dicho proceso. Las diferentes tecnologías de fundición difieren principalmente en la inversión y costos de operación y en el volumen y concentración de los gases que ellas emiten en cada etapa del proceso de fundición. Tradicionalmente, el criterio utilizado para elegir la tecnología más adecuada ha sido su eficiencia en términos de la inversión y costos de operación requeridos y su productividad. Sin embargo, con las nuevas regulaciones ambientales, factores adicionales deben ser considerados. Estas regulaciones tienen un impacto directo en el proceso de fundición; por ejemplo, algunas fundiciones deben reducir su producción durante los períodos críticos de contaminación. Así, podría ser conveniente para algunas fundiciones reducir su capacidad instalada o reemplazarla por nuevas tecnologías de fundición. Finalmente, dadas las proyecciones crecientes para la producción de concentrado de cobre para los próximos años, es importante evaluar un aumento en la capacidad de fundición del país.

### **Tratamiento de los Gases Contaminantes**

Los gases producidos en las fundiciones pueden ser emitidos directamente a la atmósfera o tratados en plantas de abatimiento de contaminantes. La tecnología comúnmente utilizada para reducir la cantidad de material particulado corresponde a los precipitadores electrostáticos. Estos usualmente reducen la contaminación en más de un 90%. Las emisiones de anhídrido sulfuroso son comúnmente tratadas en plantas de ácido sulfúrico. A pesar de que el ácido sulfúrico puede ser peligroso, éste tiene un uso en varios sectores productivos, incluyendo la minería y la agricultura. En Chile, el proceso de lixiviación utilizado para tratar los óxidos de cobre es uno de sus principales consumidores de ácido sulfúrico. Las principales decisiones en esta etapa corresponden a la capacidad de los precipitadores electrostáticos y de las plantas de ácido sulfúrico en cada período del horizonte de planificación y el volumen y composición de los gases a ser tratados. Las plantas de ácido son diseñadas para procesar un volumen máximo de gases y gases que contengan una concentración de anhídrido sulfuroso dentro de un rango específico de diseño. Estas dos cantidades –volumen y concentración de los gases que entran a la planta- determinan la producción de ácido sulfúrico. Para una tecnología de fundición dada, los gases contaminantes producidos en cada etapa del proceso de fundición (por ejemplo, horno Flash, convertidor Teniente y convertidor Pierce-Smith) difieren en los volúmenes y concentraciones de los gases contaminantes emitidos. Para determinar la producción de ácido sulfúrico, se debe determinar primero el mix de gases (la combinación de gases de cada etapa del proceso de fundición) que será procesado en la planta de ácido. Esta decisión determina a su vez la cantidad de anhídrido sulfuroso que será emitido a la atmósfera. Los costos de producción del ácido sulfúrico dependen de la concentración de anhídrido sulfuroso de los gases que son procesados en la planta. Esto se debe a que dado un volumen de gases, aquellos gases con un concentración

mayor producen una mayor cantidad de ácido sulfúrico. Esta característica hace que las nuevas tecnologías de fundición, las cuales emiten menores volúmenes de gases altamente concentrados puedan ser atractivas, aún cuando estas inversiones y costos de operación sean más elevados.

## **2.2. Características del mercado del ácido sulfúrico.**

La producción mundial anual de ácido sulfúrico es de aproximadamente 150 millones de toneladas. Esta producción proviene principalmente de dos grandes fuentes: el tostado de azufre (80%) y el tratamiento de gases producidos en fundiciones (20%). Los principales consumidores de ácido sulfúrico están en la industria de fertilizantes, la industria química y en la producción de cobre a través del proceso de lixiviación.

El mercado de ácido sulfúrico es principalmente local, con menos de un 5% de ventas en mercados internacionales. La mayoría del intercambio internacional es generado por los superávits de ácido sulfúrico de las fundiciones, que no pueden ser vendidos directamente en la región. Esto se debe a los altos costos de transporte y de almacenamiento del producto, cuyo origen se encuentra en sus propiedades altamente tóxicas y corrosivas. Estos costos son tan altos que en ciertos casos los productores generan pérdidas al no poder vender el producto a nivel local y tener que comercializarlo en el mercado internacional. Por ejemplo, Japón recientemente vendió el producto a Chile a 50 US\$/ TM; sin embargo, el costo de transporte incurrido por Japón fue de sobre 55 US\$/TM.

La producción chilena de ácido sulfúrico corresponde a aproximadamente 1.6% de la producción mundial total (2.5 millones de toneladas al año), y es utilizado principalmente en los procesos de lixiviación para producir cobre refinado; sólo un porcentaje pequeño se destina a otros procesos industriales. Como resultado de las nuevas regulaciones medioambientales sobre las emisiones de anhídrido sulfuroso al aire, la producción de ácido sulfúrico en las fundiciones de cobre está creciendo significativamente. Este crecimiento de oferta que surge de la producción de cobre se ha convertido en un importante problema para las fundiciones ya que deben vender este producto con sus altos costos de transporte y manejo, preferentemente en mercados regionales.

---

## **3. Descripción del modelo**

---

La formulación matemática consiste en dos modelos que interactúan para buscar una solución de equilibrio general. El primero de éstos, **el modelo de operación de las fundiciones**, optimiza la operación de las fundiciones, incluyendo decisiones de inversión sobre capacidad de fundición y plantas de abatimiento de contaminantes. El segundo modelo, **modelo del mercado del ácido sulfúrico**, encuentra el equilibrio entre mercados espacialmente separados, determinando

el precio y la distribución de ácido en cada región de demanda y oferta. A continuación describimos los dos modelos y la interacción entre ellos.

### **3.1. Modelo de operación de las fundiciones**

El modelo de operación de las fundiciones optimiza las decisiones de inversión en plantas de abatimiento de contaminantes y capacidad de fundición, como también las decisiones operacionales relacionadas con la distribución de concentrados entre las fundiciones estatales, la producción de cobre y ácido sulfúrico y el tratamiento de gases contaminantes. El modelo considera la producción de concentrados como una variable exógena. El modelo tiene una formulación entera no-lineal, donde las fuentes principales de no-linealidad surgen al considerar la heterogeneidad de la composición mineralógica de los concentrados en términos de cobre, azufre y arsénico. La estructura entera del modelo se debe a las decisiones de inversión sobre la capacidad de fundición y plantas de abatimiento de contaminantes. A continuación presentamos una descripción cualitativa del modelo. El modelo matemático detallado puede encontrarse en Caldentey y Mondschein (1999).

#### **3.1.1. Función objetivo**

La función objetivo corresponde a la maximización de la utilidad total descontada de las ventas de concentrado de la plantas concentradoras estatales (concentrado público) como también de la ventas de cobre refinado y ácido sulfúrico, sobre el horizonte de planeación. Los ingresos y costos son los siguientes:

##### 1. Ingresos operacionales

- o Ventas de cobre refinado provenientes de concentrado público.
- o Exportaciones de concentrado público.
- o Servicios de fundición dados a plantas concentradoras privadas.
- o Ventas de ácido sulfúrico.

##### 2. Costos operacionales

- o Costos de transporte del concentrado desde plantas concentradoras estatales hacia las fundiciones de cobre.
- o Costos de producción de cobre refinado. Estos costos dependen de la tecnología de fundición utilizada, de la composición mineralógica del concentrado y de la ubicación de la fundición.
- o Costos de producción de ácido sulfúrico.
- o Costos de operación de los precipitadores electrostáticos.

### 3. Inversiones

- o Costos de instalación de precipitadores electrostáticos y plantas de ácido sulfúrico.
- o Costos de instalación de nuevas capacidades de fundición.
- o Costos de mejora de tecnología. Este ítem incluye inversiones de modernización de plantas de fundición con el objeto de ser más eficiente o contaminar menos. Esta es una alternativa sólo para algunos equipos.

#### 3.1.2 Variables de decisión

Las variables de decisión pueden dividirse entre decisiones de inversión y de operación.

1. Decisiones de inversión: dónde, cuándo y cuánto se debe invertir en plantas de ácido sulfúrico, precipitadores electrostáticos y plantas de fundición. Debido a los costos fijos, hay variables de decisión binarias que se relacionan con el instalar o no una nueva planta ó equipo.
2. Decisiones operacionales
  - o Distribución de concentrados desde las plantas concentradoras hacia las fundiciones estatales. Debido a que la composición mineralógica de los concentrados varía según su origen, la mezcla final de concentrados a ser procesado en una fundición también es una variable de decisión. Las exportaciones de concentrado quedan determinadas al decidir la cantidad de concentrado a ser procesado en el país.
  - o Producción de cobre y ácido sulfúrico en cada fundición estatal.
  - o Volumen de gases contaminantes a ser emitidos directamente al aire. Estas decisiones determinan el volumen de gases a ser tratados en precipitaciones electrostáticas y plantas de ácido sulfúrico y por tanto, las cantidades de azufre, arsénico y material particulado emitidos a la atmósfera.

#### 3.1.3. Restricciones

A continuación describiremos las principales restricciones técnicas, medioambientales y económicas de cada etapa del proceso.

1. Plantas concentradoras
  - o Balance de producción de concentrado.
2. Fundiciones
  - o Restricciones de capacidad de fundición en cada planta.
  - o Actualización de la capacidad de las fundiciones como resultado de nuevas inversiones.

o Balance de producción de cobre y gases contaminantes. Estas restricciones incluyen: (i) conservación de masa para cada sustancia en toda división o combinación de flujos de productos o gases y (ii) conservación de las concentraciones en todas las divisiones de flujos líquidos o gaseosos. Estas últimas restricciones son la fuente de no-linealidad de nuestro modelo.

o Restricciones técnicas sobre la composición de la mezcla de concentrado a ser procesado en una fundición. Debido a las características técnicas, algunas de las fundiciones sólo pueden procesar concentrados con ciertas composiciones específicas.

o Cota para la utilización mínima de las fundiciones. Debido a consideraciones económicas, las fundiciones deben operar sobre un cierto nivel mínimo de producción.

### 3. Plantas de abatimiento de contaminantes

o Balance de producción de azufre, material particulado y arsénico en cada etapa del proceso de fundición.

o Balance entre los gases directamente emitidos al aire sin tratamiento y aquellos tratados en precipitadores electrostáticos. Este balance determina la cantidad de azufre, material particulado y arsénico en cada flujo de gases.

o Balance entre el volumen de gases emitidos a la atmósfera y aquellos tratados en plantas de ácido sulfúrico después del tratamiento en precipitadores electrostáticos. Esto determina la concentración de anhídrido sulfuroso en los gases tratados en las plantas de ácido, el cual debe estar en el rango de diseño de cada planta de ácido.

o Restricciones de capacidad en las plantas de ácido sulfúrico.

o Actualización de la capacidad de producción de las plantas de ácido sulfúrico como resultado de nuevas inversiones.

o Restricciones de capacidad en los precipitadores electrostáticos.

o Actualización de la capacidad de procesamiento de los precipitadores electrostáticos debido a nuevas inversiones.

o Balance de producción de ácido sulfúrico.

### 4. Emisiones de gases contaminantes

o Restricciones para las emisiones máximas de azufre.

o Restricciones para las emisiones máximas de material particulado.

o Restricciones para las emisiones máximas de arsénico.

### 3.2. Modelo de mercado de Ácido Sulfúrico

Como mencionamos en la sección 2.2. los costos de transporte son un factor crucial en el mercado del ácido sulfúrico, así como lo es la distribución geográfica de consumidores y oferentes. El modelo desarrollado en esta sección se basa en la literatura económica que estudia el equilibrio entre mercados espacialmente separados, en donde el precio pagado por los consumidores iguala el precio recibido por los productores más los costos de transporte asociados con el traslado del producto desde su lugar de producción hasta el zona de consumo. Partiendo con el trabajo de Enke (1951), el tema de equilibrios entre mercados geográficamente distantes ha sido ampliamente estudiado en la literatura económica. Nuestro modelo se basa en el trabajo de Paul Samuelson (1952) (Modelo Base), el cual se describe a continuación.

#### 3.2.1. Modelo Base

Consideramos un mercado compuesto por  $n$  zonas (o regiones) de producción y consumo distribuidas sobre una vasta área geográfica. Cada zona  $i$  ( $i = 1, \dots, n$ ) es representada mediante una curva de oferta  $S_i(P_i)$  y una curva de demanda  $D_i(P_i)$ , donde  $P_i$  representa el nivel de precios en la zona « $i$ -ésima». En la ausencia de intercambio comercial entre zonas, el equilibrio competitivo en cada una de ellas viene dado por un nivel de oferta y demanda  $Q_i^c$  y un nivel de precios  $P_i^c$  tales que  $Q_i^c = S_i(P_i^c) = D_i(P_i^c)$ . Ahora bien, al eliminar la barreras comerciales entre regiones se observarán flujos de productos entre ellas como resultado de las diferencias de precio. En particular, si la diferencia de precio entre dos zonas (en la ausencia de intercambio) es mayor que el costo de transporte correspondiente entonces los productores de la región donde el precio es bajo preferirán enviar su producción a la región cuyo precio es alto en lugar de venderla en el mercado local. De esta manera, los precios regionales tenderán a ajustarse y el equilibrio se alcanzará cuando la diferencia de precios entre dos regiones cualesquiera sea menor o a lo sumo igual a los costos de transportes, esto es,  $P_j - P_i \leq T_{ij}$  donde  $T_{ij}$  es el costo unitario de transporte entre las regiones  $i$  y  $j$ .

Samuelson (1952) resolvió el problema de equilibrio espacial mediante el uso de una *función de Beneficio Social*. En particular, Samuelson construye para cada región  $i$ , una función  $BS_i(Q_i)$  que refleja el beneficio adicional generado cuando la zona  $i$  comercializa  $Q_i$  unidades de producto con el resto de las regiones. La suma sobre  $i$  de los  $BS_i(Q_i)$  menos los costos totales de transporte necesarios para mover estos flujos inter-regionales constituye la función de Beneficio Social ( $BS$ ) que al ser maximizada retorna el equilibrio competitivo en el mercado regional. La metodología de Samuelson es otro ejemplo más de un resultado ya standard en economía que dice que, bajo ciertos tecnicismos, todo equilibrio competitivo es equivalente a la solución que un planificador central escogería al optimizar *cierta* función de utilidad. En este caso, dicha función de utilidad es el beneficio social,  $BS$ , definido anteriormente. Gráficamente, el valor de  $BS_i(Q_i)$  cuando la zona  $i$  exporta  $Q_i$  unidades corresponde al área achurada en la siguiente figura:



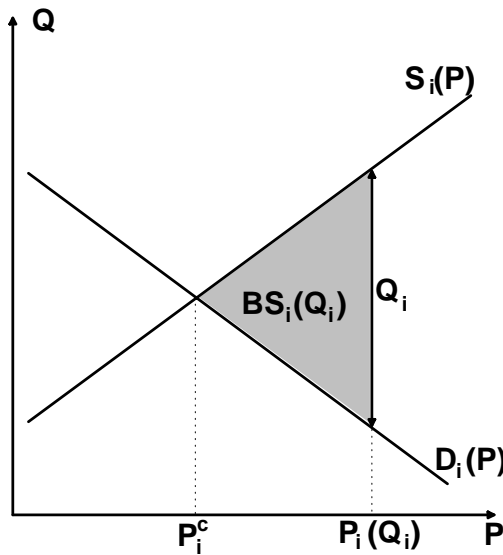


Figura 1 : Beneficio Social

Luego,

$$BS_i(Q_i) = \int_{P_i^c}^{P_i(Q_i)} (S_i(P) - D_i(P))dP, \tag{1}$$

En donde

- $S_i(P)-D_i(P)$  representa el excedente de producción que los productores de la zona  $i$  no pueden colocar en el mercado local cuando el precio en la región es  $P$ .
- $P_i^c$  es el precio de equilibrio, cuando la zona  $i$  no comercializa con otras regiones.
- $P_i(Q_i)$  es el nivel de precios en la región  $i$  para el cual el excedente de producción iguala  $Q_i$  unidades, es decir,  $Q_i = S_i(P_i(Q_i))-D_i(P_i(Q_i))$ .

En este punto es importante recalcar que el valor absoluto de  $Q_i$  representa el número de unidades transadas por la zona  $i$  con el resto de las regiones. De esta manera si  $Q_i \leq 0$  la región  $i$  es exportadora mientras que si  $Q_i \geq 0$  la región es importadora. Por último, la formulación matemática del problema se escribe en términos de los flujos entre regiones. Esto es,  $Q_i$  se escribe como  $Q_i = \sum_j Q_{ij}$  donde  $Q_{ij}$  es el flujo desde la región  $i$  hacia la  $j$ . De esta forma los costos totales de transporte vienen dados por :

$$TTC = \sum_{i=1}^n \sum_{j=1}^n (T_{ij} \cdot Q_{ij})^+, \quad (2)$$

y el equilibrio competitivo se encuentra resolviendo:

$$(PS) \quad \max_{Q_{ij}} \left\{ \sum_{i=1}^n BS_i(Q_i) - \sum_{i=1}^n \sum_{j=i}^n (T_{ij} \cdot Q_{ij})^+ \right\} \quad (3)$$

$$s.t. \quad Q_i - \sum_{j \neq i} Q_{ij} = 0 \quad (4)$$

$$Q_{ij} + Q_{ji} = 0 \quad (5)$$

El problema (PS) es en general un problema de programación no-lineal. Por ejemplo, Takayama y Judge (1964) asumiendo curvas lineales de demanda y oferta reformularon (3) como un problema de programación cuadrática. Otro caso especial ocurre cuando las curvas de oferta y demanda son perfectamente inelásticas. En este caso el problema (3) es equivalente a un problema de programación lineal que minimiza el costo total de transporte. (bajo el supuesto comúnmente utilizado que asegure que la producción total iguale la demanda de todas las regiones). Como veremos en lo que sigue, este último caso es apropiado para modelar el mercado Chileno de ácido sulfúrico cuando ciertas características especiales del mercado son combinadas con el algoritmo iterativo desarrollado para resolver el problema.

### 3.2.2 El modelo aplicado al caso chileno

En esta sección aplicamos la formulación de Samuelson al mercado de ácido sulfúrico chileno. Primero definimos las regiones comerciales que constituyen este mercado, luego caracterizamos las curvas de oferta y demanda de estas regiones para finalmente formular y resolver el equilibrio inter regional.

*Definición de las regiones comerciales:*

La primera región a considerar es el mercado internacional. Este es representado por una región única, la cual asumimos como capaz de ofrecer o demandar cualquier déficit o superávit de ácido sulfúrico requerido u ofrecido por Chile. Esta es un

supuesto razonable considerando el tamaño relativo de los mercados internacional y nacional. Nótese que sólo necesitamos saber el precio del producto en los puertos chilenos, ya que el modelo de optimización que presentamos en esta sección considera los costos de transporte correspondientes.

El mercado chileno, por otro lado, se divide en regiones de consumo y producción. Definimos nueve regiones de consumo; cuatro macro-regiones que se asocian con las divisiones político administrativas (I,II,III región y una zona agregada que combina la IV,V,VI, y XIII región) y cinco regiones que corresponden a las fundiciones estatales. Intencionalmente individualizamos las fundiciones como regiones separadas para poder estudiar en detalle los flujos de ácido desde y hacia ellas. Finalmente, los productores nacionales de ácido son solamente diez; cinco de los cuales corresponden a fundiciones estatales y cinco a compañías privadas. La figura 2 muestra la división regional que utilizamos:

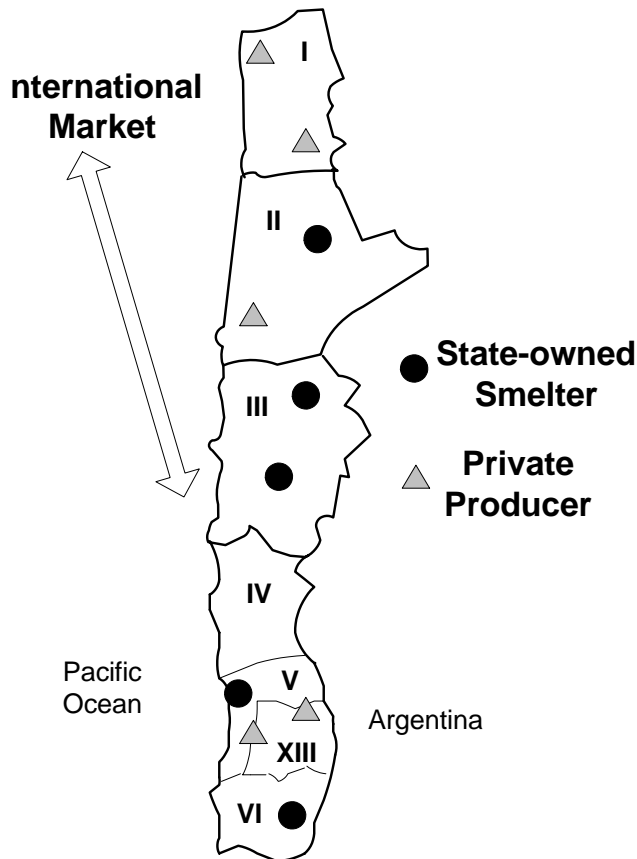


Figura 2: Mercado Chileno de ácido sulfúrico

De esta manera estamos tratando con un total de quince regiones relevantes (algunas únicamente consumidoras, otras solamente productoras, y otras consumidoras y productoras) para el mercado de ácido sulfúrico. El siguiente paso es determinar las funciones de oferta y demanda para cada región.

### *Oferta Regional de ácido sulfúrico*

Las once regiones de producción se dividen en los siguientes tres grupos:

- *I*: Mercado Internacional
- *P*: El conjunto de fundiciones privadas (5 regiones)
- *J*: El conjunto de fundiciones estatales (5 regiones)

Para el mercado internacional asumimos una oferta perfectamente elástica, es decir, en el caso de un déficit nacional, el mercado internacional es capaz de satisfacerlo completamente a un precio que es independiente de la cantidad vendida. De esta manera, si  $QN(t)$  y  $DN(t)$  representan la producción y consumo nacional respectivamente en el año  $t$ , entonces la oferta internacional viene dada por  $SI(t) = (DN(t) - QN(t))_+$ . Además, denotamos por  $PIS(t)$  el precio internacional del ácido sulfúrico en los puertos Chilenos en el año  $t$ , el cual es independiente del consumo nacional. Este supuesto fue discutido con los especialistas de comercio de ácido por parte de las fundiciones y se basa en el tamaño relativo de los mercados domésticos e internacionales.

Por otro lado, para productores privados asumimos una oferta perfectamente inelástica. Es decir, si  $Q_i(t)$  representa la producción de los productores privados en el periodo  $t$ , entonces suponemos que ésta no depende de los niveles de precios observados en las distintas regiones. La razón principal para considerar la producción privada como independiente del precio es que en el rango de precios en el que se vende el producto, las decisiones de producción del sector privado dependen principalmente de la capacidad productiva instalada.

Finalmente, para las fundiciones estatales, la producción de ácido sulfúrico está principalmente determinada por la producción de cobre, las regulaciones medioambientales y el precio del ácido sulfúrico. La producción de cobre, por un lado, determina el volumen de anhídrido sulfuroso que se produce en los procesos de fundición. De esta manera, si  $V(c)$  es el volumen de dióxido de sulfuro generado cuando se producen  $c$  toneladas de cobre y  $\bar{V}$  representa el volumen máximo de anhídrido sulfuroso que puede ser emitido directamente a la atmósfera según las regulaciones ambientales, entonces la oferta de ácido sulfúrico de una fundición que produce  $c$  toneladas de cobre está dada por:

$$S(c, p) = \alpha \cdot \underbrace{[(V(c) - \bar{V})^+]_{[1]}}_{[1]} + \underbrace{\xi(\bar{V} - (\bar{V} - V(c))^+, p)}_{[2]}, \quad (6)$$

donde  $p$  representa el precio del ácido sulfúrico y  $a$  es el factor de transformación de masa entre anhídrido sulfuroso y ácido sulfúrico. La expresión [1] representa el volumen mínimo de gases a ser tratados en la fundición por regulación ambiental, y éste es independiente del precio. Por otra parte, [2] representa aquella parte de la oferta de ácido sulfúrico que es sensible al precio, esto es  $\bar{V} - (\bar{V} - v(c))^+$  corresponde a la cantidad de anhídrido sulfuroso que la fundición puede tanto desechar directamente al aire o transformar a ácido sulfúrico para la venta. Cuánto va a procesar la fundición de la cantidad existente en [2] dependerá del precio observado de acuerdo a la función de producción  $\xi(V,p)$ . Esta función relaciona los costos de producción de ácido sulfúrico con el precio de venta.

Como describimos al comienzo de esta sección, nuestra metodología resuelve el modelo de operación de fundiciones y el modelo de mercado de ácido sulfúrico separadamente. La interacción entre estos dos modelos está dada por: (i) la producción de ácido sulfúrico de cada fundición estatal, la cual es una output del modelo de fundición y un input al modelo de ácido sulfúrico y (ii) el precio de venta de ácido sulfúrico en cada fundición estatal, el cual es un output del modelo de ácido sulfúrico y un input al modelo de fundición. Por lo tanto, según esta descomposición, la función  $\xi(V,p)$  está implícitamente determinada a través de las iteraciones sucesivas de los dos modelos. La estructura iterativa del algoritmo significa que el modelo de ácido sulfúrico asume curvas de oferta perfectamente inelásticas para cada fundición estatal, i.e., la producción de ácido sulfúrico de cada fundición esta fija **en cada iteración**. Sin embargo, es importante señalar que la producción de ácido sulfúrico en las fundiciones es sensible a los precios en el modelo global; es decir, la producción varía de iteración a iteración.

### *Demanda Regional de Acido Sulfúrico*

En lo que concierne al mercado internacional, asumimos que la demanda está dada por la oferta agregada chilena y que ésta no es sensible al precio. Entonces, la demanda internacional por ácido sulfúrico en el año  $t$ ,  $D_I(t)$ , está dada dado por  $D_I(t) = (Q_N(t) - D_N(t))^+$  en donde  $Q_N(t)$  y  $D_N(t)$  son la producción y el consumo nacional agregado respectivamente en el año  $t$ . Como las ventas chilenas de ácido sulfúrico no alteran los precios internacionales, asumimos que éstos son conocidos y los denotamos como  $\{P_I^D(t) / t\}$ .

La demanda doméstica, por otra parte, está fuertemente ligada a los proyectos de lixiviación del país, los que se utilizan para procesar los minerales de óxidos. En este sentido, la demanda por ácido sulfúrico en Chile es una demanda derivada que surge de la producción de óxidos de cobre. En la práctica, el precio internacional del cobre es el factor decisivo en los proyectos de lixiviación, siendo irrelevante el precio internacional de ácido sulfúrico importado dentro del rango observado. Esta conclusión es consistente con el hecho que la demanda por ácido sulfúrico ha estado creciendo monotónicamente en los últimos años, mientras su precio ha variado erráticamente en el mismo periodo. De acuerdo con esto, asumimos que la demanda chilena por ácido sulfúrico en proyectos mineros es independiente de su precio y el modelo adopta los niveles de consumo en las fundiciones estatales y las cuatro macro regiones como dadas.

Además, como las fundiciones estatales son simultáneamente productoras y consumidoras de ácido sulfúrico, su demanda efectiva es simplemente su déficit de producción, i.e., si  $S_j$  es la producción de ácido sulfúrico y  $D_j$  es el consumo de ácido en la fundición  $j$  entonces su demanda efectiva esta dada por  $(D_j - S_j)^+$ . De manera similar, la oferta efectiva de ácido sulfúrico de las fundiciones estatales está dada por  $(S_j - D_j)^+$ .

Dado que hemos restringido las curvas de ofertas y demandas regionales de ácido al caso inelástico, no podemos directamente determinar los precios de comercialización aplicando el modelo de Samuelson. Sin embargo, podemos resolver el problema de distribución ya que en este caso es equivalente a minimizar los costos totales de transporte como veremos más adelante.

Por otro lado, el precio del ácido sulfúrico en las diferentes regiones está determinado tanto por los niveles de consumo y producción como también por la ubicación geográfica de estas regiones. De esta manera, mientras más alejada esté una región demandante de sus zonas de oferta, más alto será el precio en esta localidad. A continuación presentamos los modelos econométricos que usamos para estimar los precios de comercialización del producto en los distintos puntos del país. Estos modelos son variaciones ad-hoc del tradicional *Modelo Gravitacional* ampliamente usado en el área del comercio internacional.

### *Modelo de Precios Regionales.*

Para cada una de las cuatro macro regiones desarrollaremos un modelo que relaciona el precio promedio en la región con el precio promedio nacional y con el déficit o superávit en aquella región y en el país. El modelo se define por:

$$P_i(t) - P_N(t) = \beta_i \cdot (ED_i(t) - ED_N(t)) \quad \forall i \in Q \quad (7)$$

En donde:

- $P_i(t)$  = precio promedio de compra en la región  $i \in Q$  en el año  $t$ .
- $P_N(t)$  = precio promedio de compra en el país en el año  $t$ .
- $ED_i(t)$  = déficit porcentual de ácido en la región  $i \in Q$  en el año  $t$  (el cociente entre demanda y oferta en la región en el año  $t$ ).
- $ED_N(t)$  = déficit porcentual de ácido en el país.
- $\beta_i$  = factor de aislamiento geográfico para la región  $i \in Q$ .

El modelo anterior fue resuelto usando mínimos cuadrados ordinarios y a partir de información histórica de contratos de compra y venta de ácido por parte de las fundiciones estatales. La intuición detrás de (7) es la siguiente: mientras

más alto el déficit de ácido en una región con respecto al país, más alto será el precio en esta región con respecto al precio medio a nivel nacional ( $\beta_i > 0$ ).

Como cada fundición estatal es ubicada dentro de una de estas cuatro macro-regiones, (7) también refleja los precios de compra y venta que observan las fundiciones.

### *Modelo para precios Nacionales*

En este modelo determinamos el precio promedio nacional en función del precio internacional y el déficit (superávit) de ácido sulfúrico de un año dado.

$$P_N(t) = \delta + \epsilon \cdot P_I(t) + \gamma \cdot (D_N(t) - Q_N(t)) \quad (8)$$

en donde

- $P_N(t)$  = precio promedio nacional año  $t$ .
- $P_I(t)$  = precio promedio internacional año  $t$ .
- $D_N(t)$  = demanda agregada nacional año  $t$ .
- $Q_N(t)$  = oferta agregada nacional año  $t$ .
- $\delta, \epsilon, \gamma$  = parámetros.

De acuerdo con (8), el precio promedio nacional depende directamente del precio internacional a través de factor  $\epsilon \geq 0$ . De esta forma, y como es evidente, alzas o bajas en el precio internacional impactarán directamente los precios locales. Por otro lado, el precio nacional se ve también afectado por el nivel de déficit del país. Esto es, en la medida que la producción local sea incapaz de satisfacer la demanda nacional ( $D_N(t) - Q_N(t) \geq 0$ ), el precio nacional se incrementará como resultado de la importación ( $\gamma \geq 0$ ). Por último, la constante  $\delta$  captura los efectos del aislamiento geográfico entre Chile y el resto del mercado internacional.

### *Formulación Matemática*

Finalmente, presentamos la formulación matemática que determina en forma conjunta la distribución y los flujos de ácido sulfúrico entre las distintas zonas como también los precios promedio de comercialización. Lo primero es observar que al considerar los mercados nacionales con demandas y producciones fijas (perfectamente inelásticos) y al mercado internacional como el buffer que absorbe o suministra cualquier excedente o déficit nacional, la formulación de Samuelson se hace equivalente a la minimización de los costos totales de transporte entre regiones.

$$(PSS) \quad \min_{Q_{ij}(t)} \sum_{t \geq 1} \sum_{i=1}^{15} \sum_{j=i}^{15} T_{ij} \cdot Q_{ij}(t) \quad (9)$$

s.t.

$$\sum_{j=1}^{15} Q_{ij}(t) = Q_i(t) \quad \forall i \text{ s.t. } Q_i(t) \geq 0, \quad \forall t \geq 1 \quad (10)$$

$$\sum_{i=1}^{15} Q_{ij}(t) = -Q_j(t) \quad \forall j \text{ s.t. } Q_j(t) \leq 0, \quad \forall t \geq 1 \quad (11)$$

$$Q_{ij}(t) \geq 0 \quad \forall i, j, \quad \text{and} \quad \forall t \geq 1 \quad (12)$$

Este modelo determina los flujos de ácido sulfúrico entre las quince regiones definidas anteriormente. Entonces, y mediante el uso de los modelos de precio (7) y (8), el precio promedio de venta para cada fundición estatal puede ser calculado de la siguiente forma:

$$\bar{P}_j(t) = \frac{\sum_{i \in \mathcal{Q} \cup \mathcal{J}} [(P_i(t) - T_{ji}) \cdot Q_{ji}(t)] + (P_{\mathcal{I}}^D(t) - T_{j,\mathcal{I}}) \cdot Q_{j,\mathcal{I}}(t)}{\sum_{i \in \mathcal{Q} \cup \mathcal{J}} Q_{ji}(t) + Q_{j,\mathcal{I}}(t)} \quad \forall j \in \mathcal{J} \quad (13)$$

Como podemos ver en la ecuación (13), el precio promedio de venta en la función  $j$  en el periodo  $t$  representa la ganancia promedio de conjunto que la fundición  $j$  recibe por cada tonelada de ácido sulfúrico que vende, y es por ello que el costo de transporte debe ser descontado del precio de venta 1.

### **3.3. Interacción entre los dos modelos**

La interacción entre los dos modelos recién presentados (operación de las fundiciones y mercado del ácido sulfúrico) es la siguiente. El modelo de operación considera fijo el precio del ácido sulfúrico en cada fundición y así determina el nivel de inversión óptima en plantas de abatimiento de contaminantes y capacidad de fundición, como también determina las decisiones operacionales para cada fundición, incluyendo la producción de ácido sulfúrico. Por otra parte, el modelo del mercado del ácido considera fija la producción de ácido sulfúrico de cada fundición y determina el precio de comercialización. Los dos modelos interactúan iterativamente hasta que convergen a una solución de equilibrio.



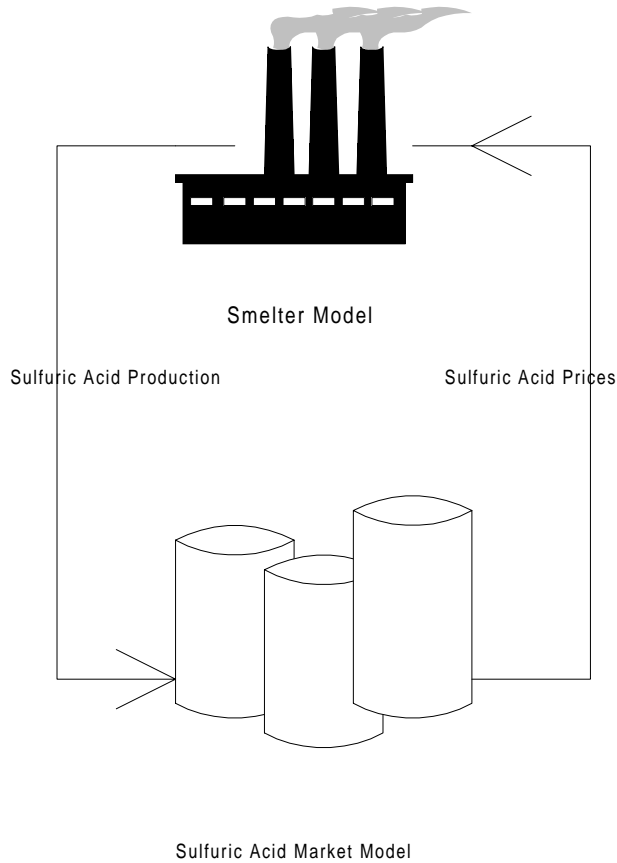


Figura 3: Interacción entre el modelo de operación de fundición y el modelo de mercado de ácido sulfúrico.

El algoritmo para encontrar la solución de equilibrio entre dos modelos es el siguiente:

1) Inicialización

- $P_0$  = vector de precios de ácido sulfúrico iniciales en cada fundición y periodo de tiempo.

- $k = 1$ .

2) Resolución del modelo de fundiciones

- $Q^k(P^{k-1})$  = niveles óptimos de producción de ácido sulfúrico en las fundiciones dado el vector de precio  $P^{k-1}$ .

3) Solución del modelo de mercado de ácido sulfúrico

- $P^k$  = vector de precios de ácido sulfúrico para niveles de producción  $Q^k(P^{k-1})$ .

4) SI  $|| P_k - P_{k-1} || \leq \epsilon$  entonces PARAR, el algoritmo ha encontrado una  $\epsilon$  - solución,

SINO  $k = k+1$  y volver a 2.

El criterio de parada en el paso cuatro se basa en la variación en los precios de ácido sulfúrico calculados en dos iteraciones consecutivas. El algoritmo itera hasta que la diferencia entre estos dos vectores de precios consecutivos es menor o igual a una tolerancia  $\epsilon \geq 0$ . Por ejemplo, si el modelo optimiza la operación de las cinco fundiciones estatales sobre un horizonte de veinte años y escogemos  $\epsilon = 1$ , entonces en promedio el algoritmo termina cuando el precio en cada fundición y cada año difiere en menos de US \$0.01 entre dos iteraciones consecutivas.

Desde un punto de vista económico, el algoritmo busca una solución que corresponde a un equilibrio competitivo. Bajo condiciones apropiadas de las curvas de demanda y oferta (ver Mas-Colell, Whinston and Green (1995)) se puede demostrar que un equilibrio competitivo existe y es único. Lamentablemente estas condiciones no se satisfacen con nuestra formulación, principalmente por la existencia de variables de inversión binarias que determinan los niveles de capacidad productiva en las fundiciones estatales. De todas maneras, desde un punto de vista empírico, la presencia de estas irregularidades no ha sido problema ya que en todos los experimentos computacionales que condujimos no observamos la presencia de equilibrios múltiples, es decir, nuestro problema aparece como muy estable.

La presencia de variables binarias también nos impide garantizar la convergencia del algoritmo. Dependiendo de la sensibilidad de las curvas de oferta al precio del ácido sulfúrico, es posible que el algoritmo itere sin converger, de manera tal que las variables binarias alteran sus valores entre una iteración y la próxima. Afortunadamente, nuestra información muestra que no es probable que esta situación ocurra en el mercado de ácido sulfúrico chileno. Como describimos en la sección 3.2.2 la producción de ácido sulfúrico esta principalmente determinada por el precio del cobre blister, siendo el precio del ácido sulfúrico poco relevante en la banda de precios que impone el mercado internacional. Más aún, al algoritmo ha convergido en todo experimento computacional realizado.

---

## 4. Implementación del modelo

---

Para el modelo de operación de fundiciones hemos formulado un modelo de programación entera no lineal de aproximadamente 18000 variables (600 variables binarias) y 12000 restricciones para un horizonte de planeación de veinte años. Fue corrido en un PC, usando GAMS, DICOPT (maneja el problema maestro), CPLEX (programación entera y lineal) y CONOPT (programación no lineal). El modelo para el mercado de ácido sulfúrico también fue corrido en GAMS y CPLEX,

y tiene aproximadamente 2000 variables y 500 restricciones. El proceso iterativo entre los dos modelos fue escrito en GAMS.

El tiempo de resolución depende fuertemente del tamaño del problema y sus parámetros. Se resolvieron cinco problemas usando información real de la operación de las fundiciones, variando los niveles máximos permitidos de emisión a la atmósfera para los contaminantes bajo análisis. Para un horizonte de planeación de diez años los tiempos de resolución fluctuaron entre 1.150 y 11.920 segundos. En estos casos el tamaño del problema fue aproximadamente 3.500 restricciones y 4.000 variables (180 binarias).

El sistema de apoyo a las decisiones descrito en este artículo se ha convertido en una herramienta importante para el proceso de toma de decisiones en la industria del cobre. Aunque el modelo fue concebido originalmente como una herramienta para determinar una política de inversión óptima para el control de la contaminación, también ha sido usado como un sistema de apoyo para diversas decisiones respecto de las fundiciones de cobre y sistemas que se le relacionan (más abajo se mencionan algunos ejemplos). Este amplio abanico de aplicaciones se logra principalmente por su visión integrada de las decisiones operacionales y las de inversión en las fundiciones de cobre estatales. Esta visión integrada permite la evaluación de un conjunto de soluciones factibles, incluyendo alternativas que son de difícil visualización cuando se optimizan las operaciones de cada fundición independientemente.

En lo que sigue, presentaremos varias aplicaciones del sistema de apoyo a las decisiones. En algunas de ellas el modelo se usa en forma periódica mientras en otras fue usado sólo en una oportunidad (la decisión debía tomarse una sola vez).

1. En el momento en que el modelo fue concebido, había un tema importante en la mesa de discusión: la instalación de una planta de ácido sulfúrico para reducir las emisiones de azufre de la fundación Paipote, (existía la posibilidad de comprar una planta de ácido de segunda mano a una fundición privada). Cochilco y Mideplan debían autorizar una inversión de US \$15 millones de dólares, sin embargo había opiniones contrapuestas respecto de si era o no una decisión rentable. Más aun, el Ministerio de Hacienda y el director de presupuesto fueron mas drásticos y pensaron que la mejor decisión era cerrar Paipote. Finalmente todos los grupos involucrados en la decisión concordaron esperar hasta que el modelo estuviese construido para contar con la información adicional de la decisión sugerida por el modelo. Con la información disponible en ese momento, el modelo sugirió continuar las operaciones de Paipote. Producto de ello Paipote continuó sus operaciones y la inversión fue autorizada.

2. El sistema de apoyo de decisiones ha sido usado para estimar el costo de políticas gubernamentales y el impacto de otros factores exógenos, como los cambios en el precio del cobre y la incorporación de nuevas tecnologías de fundición. También ha sido usado por el gobierno para evaluar el impacto económico de distintos estándares ambientales. Por ejemplo, mientras se preparaba el borrador con los nuevos estándares para las emisiones de arsénico, por encargo de la Conama el modelo se usó para evaluar el impacto económico de distintos niveles de emisión

de cada fundición. Desde un punto de vista medioambiental, mientras más baja la cantidad máxima de emisiones de arsénico a la atmósfera permitidas, mejor. Sin embargo, límites superiores más restrictivos para emisiones máximas implican mayores inversiones en plantas de abatimiento de contaminantes. En algunos casos extremos, la operación de las fundiciones podría no ser rentable y tendrían que cerrar.

El modelo demostró que los estándares de emisión propuestos inicialmente por Conama eran muy restrictivos: de aplicarse, muchas fundiciones tendrían que detener sus operaciones. Considerando el compromiso entre impactos ambientales y económicos, se fijó un estándar menos restrictivo.

3. Hasta donde alcanza nuestro conocimiento, nuestro modelo es el primero en incorporar el comportamiento del mercado de ácido sulfúrico. El precio del ácido sulfúrico, que es un componente central para tomar la decisión de cuándo y cómo ubicar una planta de ácido sulfúrico, es determinado endógenamente, en lugar de ser un dato exógeno como en la mayoría de los modelos de este tipo. Pensamos que esta es una contribución importante: nuestra experiencia ha demostrado que hasta los expertos que comercializan ácido sulfúrico no tienen herramientas formales para integrar el mercado y predecir la evolución de su precio y distribución en el país. En términos de producción y distribución local este tipo de resultado es importante para anticipar requerimientos futuros de infraestructura (carreteras, puertos) para apoyar la comercialización del producto.

4. El sistema de apoyo a las decisiones también constituye un aporte importante al resolver problemas de localización de capacidad: la instalación de una nueva fundición ha estado en discusión por un largo tiempo, y hasta hoy no se han logrado acuerdos. Para poder contribuir a una decisión en este tema, nuestro modelo incluye la opción de instalar una nueva fundición cualquier año durante el horizonte de planeación de veinte años. Los resultados que hemos obtenido sugieren que no sería económicamente beneficioso instalar una nueva fundición en este horizonte.

---

## 5. Conclusiones

---

En este artículo hemos desarrollado un sistema de apoyo a las decisiones de inversión y operación para el control de contaminación en la industria del cobre. El sistema consiste en: (i) un problema de programación entera no lineal para describir las operaciones de las fundiciones estatales y (ii) un modelo de flujo en redes para describir el mercado del ácido sulfúrico. Los dos modelos interactúan a través de la información que cada uno recibe del otro, y así el sistema itera hasta que se logra un equilibrio competitivo. Evidencia empírica usando la información tecnológica, económica y ambiental vigente, muestra que se puede lograr un equilibrio en un tiempo computacional razonable. Más aún, el equilibrio es único: distintos puntos de partida han llevado al mismo equilibrio en todos los experimentos llevados a cabo.

Una de las características más nuevas del modelo es que la distribución y el precio de ácido sulfúrico son determinados endógenamente dentro del modelo, en lugar de considerarlo información exógena como en la mayoría de los modelos anteriores que analizan la industria del cobre. La incorporación del mercado del ácido sulfúrico en nuestra formulación es una mejora sustancial sobre trabajos anteriores, considerando que una de las principales fuentes de inversión durante la próxima década tiene relación con la decisión de cuándo y dónde localizar plantas nuevas de ácido sulfúrico.

El sistema descrito en este artículo entrega la primera herramienta formal y analítica para analizar decisiones de inversión en el control de contaminación en la industria del cobre. Creemos que una de las mayores contribuciones del sistema de apoyo a las decisiones es su presente uso en el análisis de políticas ambientales y técnicas, así como el ser capaz de suministrar a los tomadores de decisiones nuevas alternativas de inversión interesantes.

### **Agradecimientos**

Los autores de este artículo agradecen a Cochilco, Conama, Enami y el Ministerio de Minería por el apoyo financiero y técnico en el desarrollo de este trabajo, a Pablo Andalaft por su aporte en la implementación computacional de este trabajo y a Eduardo Engel por sus valiosos comentarios en el desarrollo del modelo del mercado del ácido sulfúrico.

---

## Bibliografía

---

1. Caldentey R. y Mondschein (1999), "Decision Support System for Pollution Control in the Copper Industry, Including a Model for the Sulfuric Acid Market", Documento de Trabajo, Mayo 1999.
2. Enke S. (1951), "Equilibrium Among Spatially Separated Markets: Solution by Electric Analogue," *Econometrica*, Vol 19, pp. 40-47.
3. Mas-Colell A., Whinston D. y Green J. (1995), "Microeconomic Theory," *Oxford University Press*.
4. Mondschein S. y Schilkrot A. (1997), "Optimal Investment Policies for Pollution Control in the Copper Industry," *Interfaces*, Vol 27, No 6, November-December, pp. 69-87.
5. Samuelson P. (1952), "Spatial Price Equilibrium and Linear Programming," *The American Economic Review*, Vol 42, pp. 283-303.
6. Takayama T. y Judge G. (1964), "Equilibrium Among Spatially Separated Markets: A Reformulation," *Econometrica*, Vol 32, No 4, pp. 510-523.
7. *The Copper Manual*, (1976). Chilean Copper Commission and Chilean National Copper Corporation.





