

UNIVERSIDAD DE CONCEPCIÓN – CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**MODELACIÓN PREDICTIVA, SIMULACIÓN Y
MINIMIZACIÓN DE LA VARIABLE ESPECÍFICO
DE GENERACIÓN ENERGÉTICA DE CENTRAL
SANTA FE ENERGÍA PERTENECIENTE AL
SISTEMA DE DISTRIBUCIÓN DE VAPOR DE
CMPC PLANTA SANTA FE**

por
Gonzalo Ricardo Escobar Marchant

Profesor Guía:
Dr. Eduardo Javier Salazar Hornig

Concepción, Enero de 2015

Tesis presentada a la
**DIRECCIÓN DE POSTGRADO
UNIVERSIDAD DE CONCEPCIÓN**



Para optar al grado de

MAGÍSTER EN INGENIERÍA INDUSTRIAL

A mis padres Ricardo y Nelly por su incondicional apoyo y al Rey Carmesí por su profundidad musical.



RESUMEN

MODELACIÓN PREDICTIVA, SIMULACIÓN Y MINIMIZACIÓN DE LA VARIABLE ESPECÍFICO DE GENERACIÓN ENERGÉTICA DE CENTRAL SANTA FE ENERGÍA PERTENECIENTE AL SISTEMA DE DISTRIBUCIÓN DE VAPOR DE CMPC PLANTA SANTA FE

Gonzalo Ricardo Escobar Marchant
Enero de 2015

PROFESOR GUIA: Dr. Eduardo Javier Salazar Hornig
PROGRAMA: Magíster en Ingeniería Industrial

El presente estudio emplea un enfoque tipo minería de datos para la recolección y tratamiento de un banco de datos con 39 variables iniciales pertenecientes al Sistema de Distribución de Vapor de CMPC Planta Santa Fe. Dicho banco fue posteriormente reducido a 11 variables que permitieron modelar matemáticamente y luego minimizar la variable de salida Específico de Generación, de tal forma de maximizar la eficiencia de generación eléctrica de Central Santa Fe Energía. Para lo anterior, se utilizó el método de análisis factorial, obteniéndose un modelo lineal, cuya minimización entrega el *output* de las 10 variables de entrada que permiten calcular predictivamente el Específico de Generación. Los resultados se plantearon para dos casos posibles; para el primer caso se obtuvo un Específico de Generación de 4.037 [ton/MWh] y para el segundo caso 3.970 [ton/MWh], con un R^2 global de 78% para la variable calculada respecto a las mediciones.

Palabras Claves: *minería de datos, sistema de distribución de vapor, generación energética, análisis factorial, modelo lineal.*

ABSTRACT

***PREDICTIVE MODELATION, SIMULATION AND MINIMIZATION OF
ENERGETIC GENERATION SPECIFIC OF SANTA FE ENERGY CENTRAL
BELONGING TO CMPC'S SANTA FE STEAM DISTRIBUTION SYSTEM.***

**Gonzalo Ricardo Escobar Marchant
January 2015**

**THESIS SUPERVISOR: Dr. Eduardo Javier Salazar Hornig
PROGRAM: Master in Industrial Engineering**

The current work uses a data mining approach for the recollection and treatment of a database consistent on 39 initial variables that belong to the steam distribution system of CMPC Santa Fe pulp mill. This database was reduced then to 11 variables that allowed to model and minimize the output variable Generation Specific, in such a way as to maximize the electric generation efficiency of Santa Fe Energy Central. Factor Analysis was used to obtain a linear predictive model whose minimization delivers the values for 10 input variables that estimate the Generation Specific. The results were obtained for two possible cases; in the first case Generation Specific was 4,037 [ton/MWh], and in the second case 3,970 [ton/MWh], with a 78% R^2 for the calculated variable compared with measurements.

Keywords: *data mining, steam distribution system, power generation, factor analysis, linear model.*

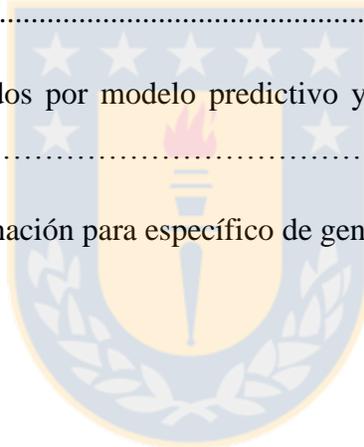
Tabla de Contenidos

1	Introducción.....	1
2.	Descripción del problema.....	3
2.1.	Conformación del Sistema de distribución de vapor.	3
2.1.1.	Planta Santa Fe.....	4
2.1.2.	Santa Fe Energía.	6
2.2	Objetivos de la Investigación.	6
2.3	Hipótesis de la Investigación.	7
3	Minería de datos, modelación predictiva y optimización de procesos industriales.	8
3.1	Proceso KDD. Descubrimiento de Conocimiento en Bases de Datos.	8
3.2	Minería de Datos.	9
3.2.1	Tratamiento Previo de datos.	12
3.2.2	Modelación Descriptiva y Modelación Predictiva.....	13
3.2.2.1	Modelación Predictiva por Regresión.	13
3.2.3	Modelos Estadísticos Basados en Factores.....	13
3.2.4	Software FACTNET Pacific Simulation.	14
3.2.4.1	Método FACTNET.	14
3.2.5	Ejemplos de estudios con modelación basada en software FACTNET.....	15
3.2.6	Modelación y optimización industrial con modelos de minería de datos.....	16
3.2.7	Análisis Factorial.	24
3.2.7.1	Factores comunes y Rotación Ortogonal.	26
3.3	Turbinas y Calderas.....	26
3.3.1	Ciclo de una planta de vapor.....	28
4.	Modelo lineal de optimización.	30
4.1	Definición del Problema.	30
4.1.1	Mercado Energético y Central Santa Fe Energía.....	31
4.2	Análisis del problema.....	33
4.3	Modelamiento del Problema.	39

4.3.1	Banco de datos y Reducción de Dimensionalidad.....	39
4.3.2	Variables que componen el Modelo y Límites Operacionales.	41
4.3.3	Modo de Cálculo.....	45
4.3.4	Variabilidad del Modelo.	46
4.3.5	Algoritmo de Cálculo.....	46
4.3.6	Partición del conjunto de datos.....	46
4.3.7	Modelo Matemático.	47
4.3.7.1	Modelo Matemático Macro.....	47
4.3.7.2	Modelo Matemático Micro.	48
5.	Resultados y Discusión.....	51
5.1	Red Factorial Inicial y Final obtenida.	51
5.2	Variabilidad Explicada.....	54
5.3	Correlación entre variables.	55
5.4	Modelo Matemático.	58
5.5	Minimización del Modelo Obtenido.	60
5.6	Capacidad de Predicción y Ajuste del Modelo Obtenido.	64
5.6.1	Simulación de la variable Específico de Generación.....	64
5.6.2	Coefficiente de Determinación de las Variables Calculadas.	65
5.6.3	Grado de Ajuste del Modelo.....	67
6.	Conclusiones y Recomendaciones.	69
	Bibliografía	71
	Anexos	74
	Anexo 1: Banco de Datos Filtrado	74
	Anexo 2: Programación del Modelo Predictivo.....	81
	Anexo 3: Filtro Factorial y Pre-clasificación	82
	Anexo 4: Set de datos de entrenamiento, número de Factores Comunes.	84

Lista de Tablas

Tabla 4.1. Conjunto de variables iniciales del sistema de distribución de vapor.....	34
Tabla 4.2. Ranking de variables obtenido mediante filtro factorial.....	40
Tabla 4.3. Variables con R^2 menor a 0,02.....	41
Tabla 4.4. Variables del modelo a optimizar y restricciones operacionales.....	42
Tabla 5.1. Correlación entre específico de generación y variables de entrada.....	56
Tabla 5.2. Valores calculados por modelo predictivo y valores del banco de datos, caso 1.....	61
Tabla 5.3. Valores calculados por modelo predictivo y valores del banco de datos, caso 2.....	63
Tabla 5.4. Coef. de determinación para específico de generación y variables de entrada...	66



Lista De Figuras

Figura 3.1. Pasos globales del proceso de descubrimiento de conocimiento en bases de datos.....	8
Figura 3.2. Fases globales del proceso de minería de datos.....	11
Figura 3.3. Ciclo Rankine.....	28
Figura 4.1. Demanda bruta mensual al SIC años 2013-2014. CDEC-SIC.....	32
Figura 4.2. Variables iniciales en el sistema de distribución de vapor.....	35
Figura 4.3. Recolección y minería de datos, modelamiento y optimización.....	36
Figura 4.4. Diagrama de flujo simplificado.....	43
Figura 5.1. Red factorial inicial.....	52
Figura 5.2. Red factorial final.....	54
Figura 5.3. Valores propios para específico de generación Santa Fe Energía.....	55
Figura 5.4. Modelo matemático.....	59
Figura 5.5. Solución para minimización del modelo/caso 1.....	61
Figura 5.6. Solución para minimización del modelo/caso 2.....	63
Figura 5.7. Modelo predictivo vs datos de construcción del modelo.....	64
Figura 5.8. Específico de generación medido vs calculado por modelo predictivo.....	66
Figura 5.9. Gráfico de falta de ajuste.....	67
Figura A2.1. Programación del modelo predictivo en Lingo.....	81
Figura A3.1. Filtro factorial.....	82
Figura A3.2. Pre-Clasificación.....	83

Figura A4.1 Partición de datos.....84

Figura A4.2. Factores comunes.....85



Introducción

La planta CMPC Santa Fe se dedica en primer lugar a la venta de celulosa y en segunda instancia a la venta de energía. Para la generación eléctrica se emplean turbinas de vapor que permiten suministrar la energía demandada para el funcionamiento de los equipos de proceso que constituyen la planta y principalmente para venta al Sistema Interconectado Central. Existen cuatro turbogeneradores asociados cada uno a una caldera; dos asociados a dos calderas recuperadoras y dos asociados a dos calderas de biomasa.

En el presente estudio se orienta en desarrollar un modelo para maximizar la eficiencia de generación energética de Central Santa Fe Energía. De aquí en adelante, cuando se emplee el término generación energética en cuanto a Central Santa Fe Energía se referirá a generación eléctrica neta de dicha central. La central se alimenta principalmente del vapor de alta presión generado por la Caldera de Biomasa N°2 y en menor medida del vapor recibido desde Planta Santa Fe.

La Caldera de Biomasa N°2 y el Turbogenerador N°4 constituyen Santa Fe Energía, entidad que si bien se encuentra dentro de la planta y es parte de ésta, es independiente del rubro celulosa, pues el negocio en este caso no es la venta de celulosa, sino más bien la venta de energía al SIC (Sistema Interconectado Central).

Los valores de venta de la energía varían según la demanda general que exista en el SIC (Sistema Interconectado Central), aumentando en el caso de que el recurso energético sea escaso, como en época estival, o disminuyendo si existe disponibilidad energética, como en época invernal, dado el efecto de las centrales de paso.

Luego, si el precio de venta en [US\$/MWh] es bajo, puede ser conveniente privilegiar el traspaso de vapor de alta presión desde Santa Fe Energía a Santa Fe en el caso de que fuese necesario por requerimientos del proceso productivo, o bien si el precio de venta es alto convendrá optimizar la generación energética de tal forma de maximizar la ganancia del negocio en cuestión.

Cabe destacar que las variables de proceso que dominan la distribución de vapor entre Santa Fe y Santa Fe Energía son altamente dependientes, es decir, un cambio en alguna de las variables modifica directamente el valor operacional del resto en menor o mayor medida. Por lo anterior, resulta conveniente desarrollar un modelo que permita simular la totalidad de las variables que afectan la generación eléctrica, contando el modelo con variables de entrada y de salida que permitan ver el efecto al realizar un cambio en una de las variables sobre el resto sin necesidad de hacerlo en forma real, dado los riesgos o pérdidas que esto conlleva.

La simulación y optimización se llevó a cabo con el software FACTNET 5.0 de Pacific Simulation, software especialmente diseñado para la industria de la celulosa en el tratamiento de un gran número de datos obtenidos a través de históricos de variables de proceso, y que considera los datos a priori como variables altamente correlacionadas.



Descripción del problema

En el presente capítulo se presenta una breve descripción del funcionamiento del Sistema de Distribución de Vapor de CMPC Planta Santa Fe. Para lo anterior es necesario explicar las áreas y sub-áreas en las que, se divide el sistema de distribución, junto con los equipos que inciden en la generación de vapor que sustenta el sistema. Posteriormente se plantean los objetivos y finalmente se presenta la hipótesis de trabajo de esta investigación.

2.1. Conformación del Sistema de distribución de vapor.

El Sistema de Distribución de Vapor consta de dos áreas macro y dos áreas micro. La primera área macro será todo el sistema concerniente a la generación y captación de vapor generado por Planta Santa Fe, mientras que la segunda área macro corresponderá a todo el vapor generado y captado por Santa Fe Energía.

La finalidad última de Planta Santa Fe es la producción y venta de celulosa, mientras que la de Santa Fe Energía es la venta de energía eléctrica al SIC. No obstante, ambos negocios no son independientes, ya que el Sistema de Distribución de Vapor conecta Planta Santa Fe con Santa Fe Energía, mediante el envío -en ambos sentidos- de vapor de alta presión.

Si bien el negocio de Bioenergía Forestal es la venta de energía eléctrica al SIC, mediante la generación del Turbogenerador N°4, este negocio se encuentra restringido por la estabilidad de la planta en el proceso de fabricación de celulosa. Si por determinado motivo Planta Santa Fe tiene problemas en alguna de sus áreas que repercutan en una merma productiva, y dicha merma implica un reajuste de las calderas de Planta Santa Fe, de modo que la generación de vapor se vea disminuida, es muy probable que Santa Fe Energía deba enviar vapor de alta presión a Santa Fe para subsanar la demanda de vapor de ésta, lo que conlleva inevitablemente bajar la generación del Turbogenerador N°4, y por tanto vender menos energía eléctrica al SIC. En otras palabras, la producción de celulosa prima por sobre la generación energética de Santa Fe Energía.

Dicho lo anterior, cabe destacar que Santa Fe Energía compra vapor de alta presión a Santa Fe y viceversa, por lo que en casos de traspaso de vapor desde Santa Fe Energía a Santa Fe por inestabilidad de Planta Santa Fe este precio se paga a un determinado costo, no obstante dicho análisis está fuera del marco de estudio del presente trabajo.

2.1.1. Planta Santa Fe.

Como se dijo anteriormente, el Sistema de Distribución de Vapor está constituido por dos áreas macro y dos áreas micro. Las dos áreas micro pertenecen al área global Santa Fe. Dichas áreas micro son Santa Fe 1 y Santa Fe 2, que corresponden a las dos líneas productivas de la planta.

Ambas líneas productivas generan vapor de alta presión, que a su vez es transformado en vapor de media y baja presión. Los equipos encargados de generar vapor en las plantas de fabricación de celulosa son las calderas. En el caso de Planta Santa Fe se cuenta con Calderas de Recuperación y Calderas de Biomasa.

Las Calderas de Recuperación son las encargadas de quemar el licor negro –subproducto del digester- y recuperar los químicos importantes para el proceso que van contenidos en dicho licor. El proceso de quemado del licor se aprovecha para la generación de vapor, al hacer pasar agua desmineralizada, previamente tratada, por tubos interiores de la caldera que aprovechan el calor generado en el hogar de la caldera y el calor de los gases de combustión, transformándose luego en vapor de alta presión.

Por otro lado, las Calderas de Biomasa queman, como su nombre lo dice, biomasa forestal obtenida de fuentes tanto externas como internas. El quemado se realiza también en el hogar de la caldera, salvo que en este caso se cuenta con un lecho fluidizado o cama en movimiento de arena a alta temperatura encargada de quemar la biomasa forestal. A estas calderas también ingresa agua previamente desmineralizada, pasando por tubos interiores y recibiendo el calor generado tanto en el hogar como por los gases de combustión, obteniéndose por tanto como producto vapor de alta presión.

Una vez explicado lo anterior, es necesario mencionar que Santa Fe Línea 1 posee dos calderas; una caldera recuperadora, denominada Caldera Recuperadora 1, y una caldera de biomasa, denominada Caldera de Biomasa N°1. En el caso de Santa Fe Línea 2, se cuenta con

la Caldera Recuperadora N°2 únicamente. Resumiendo, Planta Santa Fe cuenta con dos calderas recuperadoras y una caldera de biomasa. Distinto es el caso de Santa Fe Energía que se detallará más adelante, que si bien forma parte de la planta globalmente, se considerará como una unidad aparte debido a que su finalidad es la venta energética.

La generación máxima de vapor de alta presión de la Caldera Recuperadora 1 es de 185 [ton/h], mientras que la de la Caldera de Biomasa 1 es de 70 [ton/h]. Para el caso de Santa Fe Línea 2 la generación máxima de la Caldera Recuperadora 2 es de 710 [ton/h].

En general el vapor de alta presión generado por las calderas es almacenado en colectores de alta presión, que no son más que recipientes presurizados. Planta Santa Fe posee dos colectores de alta presión, uno para cada línea. En el caso de Santa Fe Línea 1 este recibe el vapor de alta presión generado por la Caldera Recuperadora N°1 y la Caldera de Biomasa N°1, mientras que Santa Fe Línea 2 el colector de alta presión solo recibe el vapor generado por la Caldera Recuperadora N°2.

La descarga de vapor de alta presión de los colectores se dirige principalmente a los Turbogeneradores. Los Turbogeneradores son turbinas conectadas a generadores que utilizan el paso de vapor de alta presión para generar energía eléctrica. A su vez los turbogeneradores descargan vapor de media y baja presión que es enviado a los colectores de media y baja presión respectivamente. Santa Fe Línea 1 posee un turbogenerador, denominado Turbogenerador N°1 y Santa Fe línea 2 posee dos Turbogeneradores denominados Turbogenerador N°2 y Turbogenerador N°3, por lo que en total Planta Santa Fe cuenta con tres turbogeneradores.

El Turbogenerador N°1 posee una potencia máxima de 20,5 [MW], mientras que el tope del Turbogenerador N°2 es de 44 [MW], a su vez el Turbogenerador N°3 alcanza valores de 33 [MW].

El vapor de los colectores de media y baja presión se utiliza para suplir la demanda de los equipos de proceso de Planta Santa Fe, y en menor medida para calefaccionar la planta.

Santa Fe Línea 1 posee un colector de media presión y un colector de baja presión, al igual que Santa Fe Línea 2. Además de lo anterior Santa Fe Línea 1 posee un acumulador, equipo similar a un colector, pero de menor envergadura, que recibe vapor de media presión tanto

desde Santa Fe Línea 1 como desde Santa Fe Línea 2. Tanto acumulador, como colectores de media y baja presión se encuentran intercomunicados, lo anterior sin mencionar que el acumulador también recibe vapor de media presión de descarga del Turbogenerador N°2 como aporte (*makeup*). Además, cabe destacar que el Acumulador de Alta presión de SF2 envía normalmente vapor a Santa Fe Energía y es capaz de recibir vapor de alta presión desde Santa Fe Energía en casos de contingencia operativa. Es por todo lo anteriormente descrito que se habla de un Sistema de Distribución de Vapor, en que claramente todas las variables en juego se encuentran altamente correlacionadas (ver *Figura 4.2.*).

2.1.2. Santa Fe Energía.

Santa Fe Energía está compuesto por la Caldera de Biomasa 2, con una generación máxima de 250 [ton/h] de vapor de alta presión, y por el Turbogenerador N°4, con una potencia máxima de 90 [MW], siendo esta la más alta generación individual de los 4 turbogeneradores en estudio.

El Turbogenerador N°4 posee tres descargas; vapor de baja presión, del cual una parte se va al Colector de Baja Presión SF2 y el resto se recircula a la Caldera de Biomasa 2 para soplado o deshollinado, vapor de baja-baja presión empleado para aumentar la temperatura del agua de alimentación –para posterior generación de vapor- a la caldera mediante un intercambiador de calor, y a su vez realimentado para generación de vapor después de pasar por un sistema de desionización o intercambio iónico, y por último y en mayor cantidad se descarga condensado proveniente desde el condensador de tal forma de liberar la mayor cantidad de calor latente posible para generación de energía eléctrica.

2.2 Objetivos de la Investigación.

El *Objetivo General* de esta investigación es el desarrollo de un modelo que permita simular el sistema de distribución de vapor y luego maximizar la eficiencia de generación eléctrica de Central Santa Fe Energía, tomando en consideración las restricciones operativas y de proceso concernientes a Santa Fe Energía y la sinergia correspondiente con Santa Fe.

Los *Objetivos Específicos* se indican a continuación:

- Desarrollar un modelo lineal que explique la variable de salida, Específico de Generación de Central Santa Fe Energía, en función del resto de las variables controlables del proceso, que permita su posterior minimización de tal forma de maximizar la eficiencia de generación de la Central Santa Fe Energía.
- Simular las variables de entrada y salida en base al modelo obtenido.
- Comprender de mejor manera la correlación existente entre variables de entrada y salida, y los efectos de la manipulación de estas sobre las restantes, generando nuevo conocimiento sobre el proceso.

2.3 Hipótesis de la Investigación.

La hipótesis planteada en este estudio es que es posible modelar y simular el Sistema de Distribución de Vapor de Planta Santa Fe, de forma tal de obtener valores teóricos para las variables de salida que permitan optimizar dicho modelo, minimizando la variable Específico de Generación.

La innovación de la investigación radica en que en Planta Santa Fe no se han realizado modelaciones o simulaciones matemáticas del proceso en general. Además no se ha maximizado aún la eficiencia de generación eléctrica de Central Santa Fe Energía, como modelo, considerando el sistema de distribución de vapor en su totalidad.

Minería de datos, modelación predictiva y optimización de procesos industriales.

3.1 Proceso KDD. Descubrimiento de Conocimiento en Bases de Datos.

El proceso KDD (*Knowledge Discovery in Databases*) o proceso de extracción de conocimiento desde bases de datos es definido por Fayyad et al (1996) como “el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensibles en los datos”. Frawley et al (1992) lo define de forma muy parecida como la “extracción no trivial de información potencialmente útil, implícita, previamente desconocida, a partir de una base de datos”.

Para el presente estudio el término “patrones” puede ser interpretado tanto como las relaciones a encontrar existentes entre variables de entrada y salida, como un modelo lineal global que explique todo el fenómeno.

En la *Figura 3.1.* se muestran en forma global los pasos que constituyen el proceso de Descubrimiento de Conocimiento en Bases de Datos.



Figura 3.1. Pasos globales del proceso de Descubrimiento de Conocimiento en Bases de Datos.

El proceso KDD es un proceso iterativo e interactivo que involucra numerosos pasos, resumidos de la siguiente forma (Fayyad et al., 1996).

- 1) Aprender el dominio del problema. Incluye conocimiento previo y las metas de la aplicación a desarrollar.

- 2) Crear una base de datos objetivo. Seleccionar una base de datos enfocada en un conjunto de variables o muestras de la base en las que se realizará la exploración, y descubrimiento de nuevo conocimiento.
- 3) Limpieza de datos y pre-procesamiento. Remover el ruido o valores atípicos si es necesario. Además se debe decidir estrategias para datos perdidos, representando información secuenciada en el tiempo.
- 4) Reducción de datos. Consiste en encontrar características útiles para representar los datos, dependiendo de la meta en cuestión, y emplear reducciones dimensionales o métodos de transformación, de tal forma de reducir el número de variables en consideración.
- 5) Decidir la funcionalidad de la exploración de datos. Resumir, clasificar, regresión, simulación, agrupamiento, entre otros.
- 6) Escoger el algoritmo de exploración de datos. Emplear métodos para encontrar modelos o patrones en los datos. Algunos métodos conocidos son las redes neuronales, mínimos cuadrados parciales, regresión por medio de vector de soporte, métodos de lógica difusa, análisis factorial, entre otros.
- 7) Minería de Datos (*Data Mining*). La minería de datos es un paso constituyente del proceso KDD y consiste en encontrar modelos de interés en una representación particular del set de información. Ejemplos pueden ser regresiones, paletas de colores que relacionan variables, reglas de clasificación, árboles de clasificación, análisis lineal, entre otros.
- 8) Interpretación de la información encontrada. Remover patrones redundantes o irrelevantes y transformar en términos entendibles para los usuarios.
- 9) Emplear conocimiento descubierto. Incorporar dicho conocimiento en la performance del sistema.

En el paso número siete se menciona la Minería de Datos, paso representativo del proceso KDD. En las siguientes secciones se describe más a fondo dicho proceso que permitirá modelar y optimizar el modelo planteado en el presente trabajo.

3.2 Minería de Datos.

En el presente capítulo se presentan distintas definiciones de autores y aplicaciones de la minería de datos o *data mining* a nivel industrial.

Platon y Ouloud (2007) plantean que la extracción de datos puede ser definida como la “ciencia de extraer información desde bases de datos o grandes conjuntos de datos en su defecto. La idea es construir modelos empíricos que no se sustentan en la teoría subyacente del proceso o los mecanismos que generan los datos, sino más bien en proveer comprensión y perspicacia acerca de la estructura de los datos analizados y de sus características más importantes, junto con descubrir y extraer patrones (modelos) contenidos en dichos datos”.

Otros autores como Frawley et al (1992) definen la minería de datos como el “proceso de extraer información potencialmente útil, previamente desconocida, desde bancos de datos”. Velázquez et al (2013) describen la minería de datos como el “proceso de extraer información válida, nueva y comprensible de datos masivos con el fin de mejorar y optimizar decisiones de negocios”.

Li et al (2008) plantean que la minería de datos es reconocida por investigadores como un tópico clave de búsqueda en sistemas de bases de datos y en el aprendizaje por automatización, siendo usado en campos como comercio, finanzas, telecomunicaciones y manejo empresarial en procesos de toma de decisiones. Además indican que algunas aplicaciones de la minería de datos son “mejorar la habilidad para la toma de decisiones, encontrar patrones anormales, y predecir futuras tendencias basadas en los datos archivados”. Además dividen el proceso en cuatro pasos:

- 1) Definición de la pregunta. En base a necesidades reales, conocimiento relacionado con el objetivo. El fenómeno debe ser idealmente bien comprendido antes de la extracción de datos y la meta debe ser definida y concreta.
- 2) Selección de datos y preprocesamiento. Preparación y transformación del set de datos, paso importante en el proceso, cuesta alrededor del 60% del tiempo total. En este caso existen dos tareas clave; elegir los atributos de entrada y salida adecuadamente e identificar datos implicados.
- 3) Minería de datos. Algunos métodos típicos de la minería de datos son el Análisis por Clasificación (*Classification Analysis*), Análisis de Agregación (*Clustering Analysis*), Análisis por Asociación (*Association Analysis*), Análisis Secuencial (*Sequence Analysis*), Secuencia de Tiempo (*Time Sequence*), Análisis de Valores Atípicos (*Outlier Analysis*),

entre otros. Los resultados del método se listan como conocimiento descriptivo y predictivo.

- 4) Interpretación de datos y validación. Consiste en entender el significado de la minería de datos y sus resultados, además de su validación. El conocimiento y las reglas alcanzadas deben ser enfocadas de un modo comprensible, la visualización debe ser en formatos fáciles de entender como imágenes o gráficos

Velázquez et al (2013) describen también el proceso en los siguientes cuatro pasos, análogo a lo presentado por Li et al (2008), pero con un planteamiento complementario:



Figura 3.2. Fases globales del proceso de Minería de Datos.

La *Figura 3.2* resume las fases globales del proceso de Minería de datos:

- 1) Comprensión. Se requiere tener conocimiento previo del fenómeno a modelar, de tal forma de poder construir un set de variables iniciales que permita comenzar el proceso de generación de nueva información.
- 2) Procesamiento. Se refiere a la remoción de valores atípicos, establecer cotas inferiores y superiores para las variables en juego, complementar datos perdidos y reducción de dimensionalidad del problema en cuestión.
- 3) Modelamiento. El modelamiento puede referirse tanto a cuantificar o graficar relaciones directas entre variables dependientes, ya sea de forma gráfica o numérica, o bien a la obtención de un modelo matemático que describa el comportamiento de variables de salida en función de variables de entrada, previamente seleccionadas. Para el presente caso el modelamiento puede incluir tanto representaciones gráficas como numéricas.
- 4) Despliegue. El despliegue no es más que la obtención de nuevo conocimiento que ayude a comprender de mejor manera el proceso u optimizarlo. Lo anterior depende directamente de si la finalidad del modelo es simular/modelar u optimizar un proceso, o todas las opciones anteriores.

Las técnicas de extracción de datos pueden ser empleadas para identificar patrones en los datos, correlaciones entre distintas variables de proceso y clasificar grupos de variables, entre otros. La finalidad es descubrir nuevas relaciones entre variables contenidas en bancos de datos en última instancia y en consecuencia obtener conocimiento previamente desconocido del proceso.

En el sector industrial de procesos, la extracción de datos puede ser usada para obtener conocimiento útil almacenado en grandes volúmenes de datos históricos, conocimiento que puede ser empleado después para mejorar la operación, calidad del producto y consumo de energía. En el presente estudio la finalidad será mejorar la operación y consumo de energía, la calidad del producto no podrá ser mejorada porque es venta estándar de energía.

No obstante lo anterior, existen relativamente pocas aplicaciones en la industria en cuanto a minería de datos, debido principalmente a la falta de herramientas, experiencia y personal para analizar adecuadamente los datos (Platon and Ouloud Amazouz, 2007). Sin este análisis los datos históricos disponibles de proceso, que representan un enorme repositorio de experiencias de la operación de la planta (obtenidos a menudo con gastos significativos), se convierten en un recurso infrutilizado.

3.2.1 Tratamiento Previo de datos.

Antes de descubrir modelos o patrones y optimizarlos a través de la minería de datos es necesario realizar un pretratamiento de los datos. Los principales pasos son descritos a continuación por Platon y Ouloud-Amazouz (2007):

- Limpieza de datos. Principalmente relacionado con completar valores perdidos, suavizar datos ruidosos, identificar y remover valores atípicos y resolver inconsistencias.
- Transformación de datos. Relacionada con convertir los datos a formas adecuadas para la exploración de datos.
- Integración de datos. Relacionada con combinar distintas fuentes de datos.
- Reducción de datos. Relacionado con obtener un set reducido, sin pérdida significativa de información del conjunto bruto de datos.
- Conciliación de datos. Relacionada con la integridad y precisión de registros individuales de datos.

3.2.2 Modelación Descriptiva y Modelación Predictiva.

La modelación descriptiva representa de manera conveniente las principales características de los datos, resumiendo la información contenida en el set de datos, por otro lado el modelo predictivo permite predecir el valor de algunas variables a partir de otras.

En el presente estudio se emplea un modelo predictivo. El modelo cumple el objetivo específico de predecir el valor o características de algunas variables sobre la base de valores o características de otras variables de la base de datos. Inicialmente se emplea un set de datos disponibles denominados “datos de entrenamiento” para la construcción del modelo.

El set de entrenamiento contiene tanto las variables por predecir o variables de salida (output), como las variables para la predicción o variables de entrada (input) del modelo.

Al aprender las relaciones entre las variables de entrada y salida el modelo estima un mapeo o función que puede predecir el valor de salida dado el vector de entrada. Cuando se le presentan datos nuevos de entrada nunca antes vistos el modelo puede predecir los valores “futuros” (presentes) de salida correspondientes a esta nueva entrada.

3.2.2.1 Modelación Predictiva por Regresión.

Algunos métodos empleados para modelos de predicción regresiva son:

- Modelos lineales y ajuste de mínimos cuadrados. La variable de salida es predicha como una combinación lineal de las variables de entrada.
- Redes neuronales. Principalmente redes neuronales realimentadas. Útil para casos de modelos no lineales.
- Mínimos cuadrados parciales.

En el caso de la generación del modelo lineal las restricciones empleadas para acotar el modelo corresponden a los valores de operación del proceso, es decir, los valores mínimos y máximos que alcanza una variable en la operación continua de la planta.

3.2.3 Modelos Estadísticos Basados en Factores.

Los modelos estadísticos basados en factores más empleados son el modelo de Mínimos Cuadrados Parciales (*Partial Least Squares, PLS*) y el Análisis de Componentes Principales

(*Principal Components Analysis, PCA*). Usualmente son usados para transformar un número de variables de proceso relacionadas y no relacionadas en un grupo menor de variables más fuertemente relacionadas. Estas técnicas son usualmente llamadas metodologías de reducción de datos, debido a que identifican factores que tienen una dimensión mucho menor que los datos originales, y que aún pueden describir las principales tendencias que manifestaba el set de datos original. Para el caso en particular del Análisis de Componentes Principales, el método provee un mejor entendimiento del proceso, indicando correlaciones entre variables de entrada y variables de entrada/salida.

En el presente estudio la modelación estadística se llevará a cabo con el software FACTNET Pacific Simulation. Platon y Ouloud-Amazouz (2007) lo definen como un “sistema de análisis estadístico basado en la técnica de Análisis de Factores (análisis factorial) que ofrece capacidad de análisis de datos y además permite crear modelos predictivos”. El software realiza la salvedad de analizar los datos asumiendo que estos se encuentran fuertemente correlacionados, cuantificando las relaciones entre las variables en estudio.

3.2.4 Software FACTNET Pacific Simulation.

El manual de Factnet Using *Pacific Simulation to Understand your Multivariate Process Data* (Pacific Simulation, 2005) describe a FACTNET como un “sistema de análisis estadístico que evita la mayor debilidad que poseen los análisis de regresión normales. Los análisis de regresión son basados en el principio de variables independientes y dependientes”.

Se plantea que “en el mundo real hay muy pocas variables independientes. La mayoría de las variables están interrelacionadas. FACTNET analiza los datos empleando el supuesto de que los datos están interrelacionados, y eso evita las dificultades inherentes de los análisis regresivos”.

3.2.4.1 Método FACTNET.

Las redes de factores o redes factoriales consisten en agrupaciones de variables de entrada, factores comunes y variable(s) de salida en un sistema. Los factores comunes son los encargados de establecer el grado de dependencia entre variables de entrada y salida. El primer paso en el análisis de redes de factores (*factor network analysis*) se basa en el sistema tradicional de análisis multivariable denominado Análisis Factorial.

Una de las principales motivaciones del análisis factorial es determinar cuántas propiedades independientes significativas existen en la base de datos. Lo hace al examinar el grado de correlación existente entre las variables, mediante los ya mencionados factores comunes. El número de factores comunes es siempre menor que el número de variables analizadas y si existe un alto grado de correlación entre las variables el número de factores comunes requeridos será pequeño.

3.2.5 Ejemplos de estudios con modelación basada en software FACTNET.

Quintana (2004) modela matemáticamente las emisiones de gases TRS (Sulfuros Totales Reducidos) de una caldera recuperadora de una fábrica de pulpa kraft. Lo anterior lo realiza empleando factores comunes para definir la variable en estudio, en este caso la concentración de TRS, como una función lineal de las n variables de entrada del modelo, semejante a lo que se realizará en el presente trabajo referente al Específico de Generación de Central Santa Fe Energía.

La función posee la siguiente forma:

$$[TRS] = C_0 + C_1 \cdot Variable\ 1 + \dots + C_n \cdot Variable\ n ; j = 1, \dots, n \quad (1)$$

Donde C_j corresponde a las ganancias del modelo y se calcula por método de extracción de factores. Junto con la modelación de la emisión de TRS se encuentra que las variables más influyentes en el proceso, en orden descendente de influencia en cuanto a generación son la T° promedio de los gases a la salida del banco generador, T° del aire terciario y flujo de licor negro.

Por otro lado Contreras (2010) optimiza las propiedades físico-mecánicas del papel sackraft, aumentando su resistencia en dirección transversal, desde 4,0 a 4,5 [kN/m], manteniendo los valores de porosidad sobre 210 unidades Sheffield y los de elongación transversal menores a 9,5 %. Para lograr lo anterior también trabaja con los factores comunes, obteniendo ecuaciones de ponderación y ecuaciones de ganancia, que simulan el proceso y manifiestan las correlaciones existentes entre variables de entrada y de salida.

Posteriormente incorpora dichas ecuaciones al *Optimizer* de FACTNET, software que en base a las ecuaciones de ganancia obtenidas y los límites operacionales de cada variable optimiza el modelo de emisiones de TRS descrito previamente (1), indicando cuales son los rangos de trabajo de las variables de entrada (manipuladas en sala de control) para poder obtener los resultados requeridos en las variables de salida, que en este caso son las propiedades físico-mecánicas del papel. Para dicho trabajo emplea 53 variables de entrada y 11 variables de salida, representadas por 16 factores comunes. Una vez empleado *Optimizer* este indica los nuevos valores a los cuales deben ser llevadas las variables de entrada más influyentes para cumplir con los objetivos propuestos.

3.2.6 Modelación y optimización industrial con modelos de minería de datos.

En la presente sección se muestran algunos estudios y sus resultados empleando distintas metodologías de minería de datos, la finalidad es explicar un marco global de modelación y optimización que rige todo proceso de minería de datos independiente de las herramientas empleadas para alcanzar los resultados.

Kusiak y Zhe Song (2006) emplean un enfoque de minería de datos para optimizar la eficiencia de combustión de una caldera a carbón. Mediante pruebas virtuales validan los resultados obtenidos de la optimización, ya que las pruebas en vivo implican altos costos y consumo de tiempo.

Para el caso de la optimización de la eficiencia de combustión, y en general para la modelación de procesos industriales, se plantea que los enfoques pueden ser agrupados en tres categorías. La primera categoría incluye modelos analíticos basados en termodinámica y química. La segunda categoría incluye métodos computacionales como redes neuronales, lógica difusa y computación evolutiva, entre los métodos recién descritos el más empleado en la industria energética son las redes neuronales. La tercera categoría incluye sistemas híbridos que combinan las dos categorías recientemente descritas.

Por otro lado la extracción de información a partir de la minería de datos permite crear modelos de decisión para el personal operativo o ingenieros de las industrias de proceso. Monedero et al (2012) generan un modelo de decisión empleado para optimizar la eficiencia energética de una planta petroquímica. El modelo se realiza luego de un proceso de minería de

datos de parámetros registrados en el pasado. El sistema diseñado involucra un algoritmo combinado, por un lado emplea la eficiencia energética de los puntos de operación empleados en el pasado, mientras que por otro lado un módulo de dos redes neuronales para obtener nuevos puntos de operación interpolados, en definitiva un modelo predictivo. Los autores obtienen un aumento de la eficiencia energética promedio de la planta de un 7%, a partir de la optimización de una función de carácter lineal, compuesta de 11 variables controlables o variables de entrada y 5 variables no controlables o variables de salida.

Velázquez et al (2013) describen el desarrollo de un sistema de gestión energética para una planta reformadora de nafta empleando minería de datos, logrando identificar las variables claves operacionales en el consumo energético, obteniendo un ahorro de un 4 a un 7% del total de la energía generada. El proceso consta de tres reactores adiabáticos y 3 columnas de destilación; un debutanizador, un debencenizador y un depentenizador. El consumo energético de la planta en 2009 era de 233,1 GWh, con un 67% de consumo en la zona de reactores y un 33% en la zona de destilación.

El análisis para llevar a cabo la minería de datos es realizado por el software SPSS Modeler Clementine, con un set de entrenamiento de 12149 datos registrados y un set de validación de 9402, siendo filtrados 120 valores atípicos, debido a operación inusual de la planta, luego se realiza un segundo filtrado de acuerdo a requerimientos de proceso, eliminándose 132 valores adicionales. Posterior al filtrado se realiza la modelación, obteniéndose una función lineal compuesta por 14 variables.

Motlaghi et al (2008) diseñan un sistema experto para una columna de crudo de una refinería de Abadan de tal forma de predecir valores desconocidos del flujo del producto y temperatura requerida en la alimentación. Los autores definen un sistema experto como un “sistema computacional que emplea conocimiento para obtener altos niveles de rendimiento en solucionar problemas de un área específica, permitiendo a las organizaciones respaldar importantes tomas de decisiones y mejorar la productividad”. Se clasifican dentro de lo que se conoce como inteligencia artificial.

El sistema desarrollado por los autores minimiza el error de salida del modelo y maximiza la tasa de producción de petróleo manipulando los parámetros de control. Las variables de entrada son parámetros operacionales del crudo como temperatura de entrada, mientras que las

variables de salida son definidas como características del producto. La base de datos se procesa con una Red Neuronal, a través del programa *MATLAB* y su complemento *Neural Networks Toolbox*, para organizar y coleccionar los datos relacionados al proceso y también para predecir los valores desconocidos de salida con la ayuda de los valores de entrada. La optimización del modelo se realiza por medio de un algoritmo genético. El 60% de los datos son empleados para entrenamiento del modelo y 40% para testeo del modelo. Se obtiene un error de alrededor del 2% entre dato real y calculado por el modelo.

Edwards et al (1999) describen la aplicación de redes neuronales en la predicción de la curvatura del papel, medida única y fiablemente fuera de línea, después de la manufactura. Se presentan parámetros que caracterizan la bobina de papel como variables de entrada de una red neuronal, y se entrena dicha red para predecir si el nivel de curvatura estará dentro de la especificación necesaria. En paralelo se presentan los mismos datos a otra red tratando de predecir el valor absoluto de la curvatura con un Modelo por Regresión Neural (mixto) en el cual los parámetros pueden ser alterados para reducir la curvatura de ser necesario, presentando una luz roja/verde indicando aceptabilidad del papel. La curvatura del papel es controlada usando variaciones en las temperatura de la zona de secado y el nivel de humedad.

En el ámbito del rubro energético Kusiak et al (2010) emplean un enfoque basado en el proceso de Minería de Datos, para maximizar la generación energética de una turbina de viento. Los autores identifican un mapeo funcional entre la energía como variable de salida y las variables controlables y no controlables de la turbina, basado en los datos de una granja industrial de viento. Una granja industrial de viento consiste en un conjunto de turbinas de viento ubicadas contiguamente que conectadas a distintos generadores proveen energía eléctrica.

En cuanto a la optimización, Kusiak et al emplea un algoritmo evolutivo para determinar ajustes de control que maximizan la generación energética (variable de salida) basados en el modelo identificado mediante redes neuronales y mediante arboles de regresión y clasificación (*regression and classification trees*). Como resultado, se demuestra que el ángulo de paso de las palas puede aumentar la generación energética al ser optimizado. Los autores además plantean que la ventaja de un modelo basado en datos es que identifica el poder o capacidad actual de generación energética de los datos de proceso, punto a tomar en cuenta debido a que

la gran parte de los equipos sufre desgaste con los años de operación, y su punto de operación óptima se sitúa en valores distintos a los de diseño.

Para la recolección de datos se emplea sistema SCADA (Control de Supervisión y Adquisición de Datos). El sistema colecta datos de más de 120 parámetros y los recolecta en intervalos de 10 [min] como valores promedios, el rango de estas turbinas varía entre 0-1000 [kW]. Los datos se dividen entre dos subconjuntos, uno de entrenamiento y otro para validación del modelo obtenido, al igual que se hará con el presente estudio.

Li et al (2008) aseguran que con el desarrollo de la industria eléctrica una gran cantidad de datos en tiempo real son enviados a bases de datos por sistemas de adquisición de datos, acumulándose grandes cantidades, por lo que existe abundante conocimiento pasivo acumulado. Plantean además que “la aplicación de minería de datos en la industria eléctrica puede guiar a la operación óptima basada en un histórico de datos y mejorar la eficiencia económica de una planta de energía”, resaltando que por lo general los datos eléctricos poseen alta correlación. Los autores obtienen los resultados de la optimización a través de métodos de asociación difusa (*fuzzy association rules*) y logran concluir que en el 90% del rango la presión de la corriente principal es extremadamente más baja que el valor obtenido de la optimización operacional, principal factor que incide en disminuir la eficiencia económica. Otras variables en juego son la temperatura de agua de alimentación a la caldera, temperatura del gas de combustión, coeficiente de exceso de aire y vacío del condensador.

Continuando con el ámbito energético, Wang et al (2012) emplean datos almacenados en planta para modelar el consumo energético de una caldera de 600 MW y optimizar los parámetros de operación para minimizar el consumo de carbón obteniendo la más alta eficiencia energética. Las restricciones empleadas fueron la carga de potencia de la turbina a vapor con condensador (600 MW) y la calidad del carbón. El método de vector de soporte para regresión (*support vector regression*) y algoritmo genético son usados para modelar el comportamiento del consumo de carbón y determinar el valor óptimo de los parámetros de operación. Algunos parámetros controlables considerados fueron la temperatura del vapor principal, presión del vapor principal, temperatura del vapor recalentado, presión del vapor recalentado, todas obviamente sujetas a las restricciones operacionales.

Entre los resultados se encuentra que con un 100% de capacidad de la unidad de generación el consumo se reduce en 2,87 [g carbón/kWh], con 75% se reduce en 3,93 [g carbón/kWh], con 50% se reduce en 5,80 [g carbón/kWh]. Los valores anteriores implican que en cada caso se necesitan menos masa de carbón para generar la misma cantidad de energía.

Como se mencionó anteriormente los modelos construidos a partir de minería de datos pueden ser tanto para modelación descriptiva como predictiva. Kélouni y Agbossou (2004) construyen un modelo predictivo no lineal basado en una Red Neuronal para la estimación de la generación energética de una turbina, con una desviación o error menor a un 1% con respecto a la predicción de generación energética.

El modelo no lineal propuesto usa la velocidad promedio del viento, la desviación estándar de los datos y la generación energética del pasado como datos de entrada o inputs del modelo. La velocidad del viento se mide con un anemómetro. Se emplea la velocidad promedio del viento cada 10 minutos, junto a su desviación estándar, con datos de Septiembre del 2000 para entrenamiento, mientras que los de octubre del 2000 se emplean para validar el modelo. La configuración de la red neuronal se nutre de 8 variables de entrada, 5 neuronas en la capa oculta y 1 neurona en la capa de salida.

Li et al (2007) emplean minería de datos para optimizar una planta de generación energética, mediante un algoritmo de asociación difusa (*fuzzy association rules*), de tal forma de encontrar valores objetivos óptimos para las variables que guían la operación de la planta, y permitan aumentar la eficiencia de generación energética de una caldera a carbón, disminuyendo a su vez la emisión de contaminantes. Se utilizan datos locales de un equipo de 300 [MW]. Las pérdidas energéticas son causadas por la desviación de los principales parámetros desde su valor óptimo, por lo que estas pueden ser calculadas una vez que se cuenta con los valores óptimos de operación. A su vez estas sirven como base para posteriores análisis económicos y reforma operacional. Li et al (2007) mencionan que “la determinación adecuada del vapor óptimo de operación es importante para la mejora de la eficiencia económica y técnica de operación en las plantas de energía”.

Además, manifiestan que el valor de diseño se obtiene principalmente de experimentos y cálculos térmicos, no obstante las unidades de proceso poseen variaciones de funcionamiento con el tiempo de servicio, por lo que los métodos más tradicionales por lo general no permiten

reflejar el funcionamiento actual de los equipos correctamente. Como resultado de lo anterior los estados de funcionamiento de las unidades de proceso a menudo no son los óptimos, por lo que la minería de datos surge como una herramienta útil a la hora de analizar el estado actual de estas. Dicho lo anterior los autores refuerzan el concepto al decir que “es urgente analizar las bases de datos y encontrar conocimiento valioso oculto, junto con encontrar mejores manera para optimizar el sistema y el proceso”.

En las plantas energéticas por lo general la determinación de valores óptimos se lleva a cabo a partir de los valores de diseño del proveedor, alternativa que cuenta con los siguientes contras

- a) Los proveedores no consideran los factores del sistema al determinar los valores de diseño
- b) Desviaciones durante el proceso del diseño del equipo e instalación

Luego, los cambios en condiciones externas o incremento en el tiempo de operación son factores que el valor de diseño no considera. En base al planteamiento anterior los autores consideran el factor tiempo, por lo que emplean 3 meses de datos relacionados al consumo de carbón. Los parámetros empleados por el método de asociación difusa (*fuzzy association rules*) son la presión de vapor principal, temperatura del vapor principal, temperatura del vapor recalentado, temperatura del gas de salida, razón de aire en exceso, entre otros. Como resultado logran aumentar en promedio un 0,28% la eficiencia energética de la caldera, además se reduce el consumo de carbón en 1,6 [g/KWh], con la consiguiente disminución de la contaminación y costos en combustible.

Zheng y Yu (2008) mencionan que los sistemas de combustión de carbón son fuente significativa de emisiones de NO_x , por lo que mientras la utilización de combustibles fósiles se incrementa, el monitoreo de NO_x se vuelve indispensable. Con el objetivo de minimizar las emisiones de una caldera de 300 [MW] Zheng y Yu emplean una Red Neuronal de una pasada y un modelo de regresión generalizada de red neuronal (*generalized regression neural network*), de tal forma de generar un modelo no lineal entre los parámetros de la caldera y las emisiones de NO_x . El modelo, como el resto de los modelos anteriores, emplea datos de entrenamiento y datos para validación, y si bien plantean que la finalidad es reducir las emisiones de NO_x solo se llega a la fase de simulación de la variable.

Como resultado del estudio se obtiene que la selección de parámetros con métodos GRNN (Regresión Generalizada de Red Neuronal), predice con mayor exactitud las emisiones que el

ampliamente usado método de red de retropropagación (iterativo) y que el modelo lineal de regresión múltiple. Cabe destacar que las redes de retropropagación son empleadas usualmente para modelamiento y control de las emisiones de NO_x de calderas a carbón. GRNN es un procedimiento de una sola etapa, que ayuda en el desarrollo rápido de modelos no lineales de entrada-salida. Por otra parte los autores plantean que los PEMS (Sistemas Predictivos de Monitoreo de Emisiones) son más costosos que los CEMS (Sistemas Continuos de Monitoreo de Emisiones), por lo que la ventaja del modelo GRNN es que permite predecir los valores de NO_x con menores costos y mayor exactitud que los modelos obtenidos por las redes de retropropagación (BPNN). En el estudio los autores realizaron 670 pruebas cambiando los parámetros de la caldera, con 21 *inputs* o variables de entrada.

En cuanto a la recopilación de datos Zheng y Yu (2008) plantean que “en la actualidad la mayoría de las instalaciones industriales están equipadas con sistemas de control distribuido (DCS) que suministran una gran cantidad de información sobre lo que está sucediendo en la planta, de tal forma de asegurar el correcto funcionamiento de los equipos y la producción”. Planta Santa Fe justamente posee un sistema DCS del cual se extraerá el banco de datos.

Siguiendo con los modelos predictivos, y minería de datos en general, Zhou et al (2012) plantean que la modelación de las emisiones de NO_x es crítica para desarrollar un PEMS o Sistema Predictivo de Monitoreo de Emisiones, además buscan implementar un software que permita la optimización de la combustión de una caldera a carbón, de tal forma de minimizar las emisiones de NO_x . La modelación se lleva a cabo con el método de regresión por medio de vector de soporte (*support vector regression, SVR*), empleando gran cantidad de datos para entrenar el modelo y validarlo. La optimización se realiza con el algoritmo de colonia de hormigas, obteniendo parámetros del modelo SVR con gran exactitud.

Zhou et al (2012) plantean que “a diferencia de los modelos estadísticos tradicionales, la red de retropropagación (BPNN), el modelo de regresión generalizada de red neuronal (GRNN) y la regresión por medio de vector de soporte (SVR) son herramientas de modelamiento tipo cajas negras, esto significa que son modelos guiados por los datos y que dejan a estos hablar por sí mismos, en los que el usuario no conoce todo el detalle de la estructura interna del modelo. En general lo que se realiza es ingresar los datos de entrada de tal forma de obtener una respuesta de salida deseable”.

Las emisiones se modelan a partir de 21 parámetros operacionales como variables de entrada obteniéndose para el set de entrenamiento un error absoluto máximo equivalente a 48 [ppm] y un error relativo máximo de 18,56%, mientras que para el set de validación se obtuvo un error absoluto máximo de 41 [ppm] y un error relativo máximo de 16%. En el 92% de los casos del set de validación se obtiene un error relativo menor a 5%.

En un estudio anterior Zhou et al (2010) emplean un enfoque de inteligencia computacional para modelar y minimizar emisiones de NO_x derivadas de la combustión de una caldera a carbón. Se emplea la optimización por enjambre de partículas (*particle swarm optimization*), regulando 4 niveles de aire primario y 6 niveles de velocidad de aire secundario.

Zhou et al (2010) logra reducir las emisiones de NO_x desde 399,7 hasta 269,3 [ppm], superando los resultados obtenidos a partir del algoritmo genético y un poco mejor que los resultados obtenidos por el algoritmo de colonia de hormigas. Otra implicancia positiva es que posee un costo computacional menor a 25 segundos, mucho menor al tiempo empleado por el algoritmo de colonia de hormigas, siendo más aplicable a modelos en línea y en tiempo real. Para este caso, al igual que en los ejemplos anteriormente descritos, se emplea un set de entrenamiento y un set de validación o de prueba de datos.

Wu et al (2009) emplean un método de regresión por medio de vector de soporte para establecer modelos matemáticos que simulen las emisiones de NO_x y quema de carbón en una caldera de 300 [MW], de tal forma de generar ahorro energético y reducción de emisiones.

Para la optimización multi-objetivo se emplea un algoritmo genético celular. El enfoque provisto entrega una herramienta efectiva para la optimización, validada experimentalmente. Las ventajas del Método de Regresión por medio de Vector de Soporte radican en que permite obtener un óptimo global, posee una estructura simple y buenas propiedades generalizadas, aunque debe ser empleado para muestras de tamaño moderado.

Las muestras de NO_x se toman a la salida de la caldera, antes del calentador de aire. Las concentraciones son valores promedio después de horas de operación estable. Como muestra se emplean 119 casos de testeo para evaluar efectos de carga de la caldera y velocidades de aire primario, secundario y terciario, además se considera la calidad del carbón a quemar constante y se realiza seguimiento del carbón no quemado como cenizas.

Entre los resultados para baja carga de la caldera el mínimo de emisiones de NO_x fue de 324,9 [ppm], lográndose una reducción de 8,6%, mientras que el carbón remanente en las cenizas fue de 1,9% con una reducción de 9,5%. Para alta carga de la caldera el mínimo de emisiones fue de 337,7 [ppm], con una reducción de 4,3%, mientras que el carbón remanente en las cenizas fue de un 1,3% con una reducción de un 38,1%.

También en el área de eficiencia energética y reducción de las emisiones Zhen et al (2008) plantean que la optimización de la combustión ha demostrado recientemente su potencial para reducir las emisiones de NO_x en calderas a carbón de alta capacidad. Los autores emplean el método de regresión por medio de vector de soporte para modelar la relación entre las emisiones de NO_x y los parámetros operativos de una caldera a carbón de 300 [MW]. El método de regresión por medio de vector de soporte muestra resultados más acordes a valores obtenidos en las pruebas experimentales que la red neuronal artificial. Lo anterior con la caldera operando a distintas cargas.

El modelamiento presenta un error de 1,58% con un factor de correlación de 0,94 respectivamente. La optimización del modelo se lleva a cabo con algoritmo de colonia de hormigas, reduciendo las emisiones en 65 [ppm] correspondientes en un 18,69%. El modelo por su parte también emplea datos de entrenamiento y datos de validación. Las emisiones se reducen según la predicción del algoritmo de colonia de hormigas desde 346 hasta 283 [ppm], mientras que en la realidad se reduce a 281 [ppm].

3.2.7 Análisis Factorial.

El análisis factorial fue introducido por Spearman (1904), inicialmente empleado en áreas como psicología y ciencias sociales, en un intento de identificar los factores relacionados con la inteligencia humana y vincularlos de alguna manera al origen étnico.

Hair (2010) define el análisis factorial como un “método estadístico multivariado cuyo propósito principal es definir una estructura subyacente en una matriz de datos”. En términos generales aborda el problema de analizar una estructura de correlaciones de un gran número de variables, que define un conjunto de dimensiones latente, denominado factores comunes. Su finalidad es describir los datos en un número mucho más pequeño que las variables originales

individuales en el banco de datos. Johnson et al (2014) dan una explicación acabada de toda la notación matricial y explicación teórica del modelo.

Jeong et al (2012) emplean el análisis factorial para la primera etapa de preprocesamiento, como una “técnica de compresión de datos con orden intrínseco”. Plantean que el análisis factorial sirve para determinar cuántos grupos de variables existen dentro del total, es decir, para explicar la correlación entre grupos de variables, relacionadas a su vez a factores comunes con mayor o menor fuerza.

Souza et al (2013) presentan un método de análisis factorial aplicado para pre-procesar los datos, como parte del concepto de minería de datos. El análisis factorial se emplea para reconocer patrones y clasificar datos, en este caso de problemas de índole energético-ingenieril analizando las variables de entrada de tal forma de obtener conocimiento y comprensión del proceso. En el estudio mencionado los autores extraen 6 factores comunes desde la base de datos representando el 71,22% de la variabilidad de los datos originales.

El Análisis Factorial se emplea para filtrado y procesamiento previo de la información con el fin de obtener una reducción del volumen de datos, problema común, que se traduce en acotar la dimensionalidad del problema en estudio. Este proceso se lleva a cabo cuando el número de variables en estudio es muy elevado. La idea es obtener un nuevo vector que sea combinación lineal de las variables originales. Además el Análisis Factorial permite agrupar las variables y factores de correlación con las direcciones de mayor variabilidad, desestimando las variables que no son significativas para la variabilidad del modelo.

El Análisis Factorial resulta eficaz en:

- Reducción del volumen de datos.
- Agrupación de las variables originales sobre los factores comunes.
- Mejor rendimiento computacional debido a reducción de volumen de datos.
- Análisis de la correlación entre las variables y su importancia en la variabilidad de los conjuntos de datos.

La reducción de dimensionalidad se rige por 2 criterios:

- Conocimiento informal del especialista utilizando la experiencia en el problema.

- Conocimiento a través de criterios estadísticos conocidos como técnicas de selección de variables.

Para reducir la dimensión del vector original se emplean técnicas de análisis multivariado, en específico el método PCA (análisis de componentes principales). El Análisis Factorial utiliza la técnica de componentes principales PCA para estimar los pesos de los factores y permite la visualización de los grupos de variables originales que tengan una cierta asociación y estén en la dirección de mayor variabilidad.

Los factores son aquellas combinaciones lineales de las variables originales, es decir, cada factor estará en función de todas las variables de entrada y/o variables de entrada salida según sea el modelo de cálculo elegido para representar el problema.

3.2.7.1 Factores comunes y Rotación Ortogonal.

La idea del análisis factorial es obtener a partir de la base de datos original un número de factores m , con $m < p$, siendo p la dimensión del vector observado, es decir, el número de variables originales. Los m factores son los encargados de explicar la variabilidad del problema, o por lo menos de explicarla en gran parte, más adelante se hace mención de los valores propios o eigenvalores como medidas para explicar la variabilidad del modelo.

Uno de los métodos empleados para la interpretación de las relaciones entre variables es la Rotación Ortogonal. Esta herramienta sirve para encontrar la estructura más simple de interpretar los factores comunes obtenidos, no obstante no incide en el grado de variabilidad obtenida del modelo, luego el grado de exactitud no variará independiente de la rotación de factores.

3.3 Turbinas y Calderas.

En la presente sección se describen características generales y de funcionamiento de los principales equipos presentes en el sistema de distribución de vapor: las turbinas y calderas.

En Catalog of CHP Technologies (EPA, 2013) la Agencia de Protección Ambiental de Estados Unidos plantea que las turbinas a vapor tienen una alta eficiencia global en general (75% eficiencia general). Para lograr esa alta eficiencia solo es necesario alimentar a las turbinas con vapor de alta presión, vapor que puede ser generado en calderas empleando distintos tipos de

combustible, lo que otorga a los sistemas de cogeneración diversas posibilidades en cuanto a diseño del proceso y utilización de materias primas para generación de vapor de alta presión por parte de las calderas; ya sea biomasa, lodos industriales, carbón, entre otros. Sumado a las ventajas anteriores cabe destacar que las turbinas poseen una alta vida útil y confiabilidad operacional, además de la habilidad para trabajar en distintos rangos de generación. La única desventaja a considerar es la partida lenta, no obstante en plantas de proceso continuo por lo general las turbinas se encuentran en funcionamiento, salvo para detenciones anuales para mantenimiento.

Cuando se menciona en el párrafo anterior que las turbinas a vapor trabajan en distintos rangos de generación, esto quiere decir que abarcan potencias desde 50 [MW] a 250 [MW]. Para el caso de CMPC Santa Fe, y más específicamente de central Santa Fe Energía, el Turbogenerador N°4 cuenta con una potencia máxima de 90 [MW].

La sinergia existente entre las calderas y turbinas se explica a continuación. La energía producida en la caldera es transferida a la turbina a través de vapor de alta presión, que a su vez impulsa la turbina y el generador. Esta separación de funciones permite a las turbinas operar con vapor de alta presión obtenido de calderas que trabajan con una alta variedad de combustibles incluyendo gas natural, desechos sólidos, carbón, madera, desechos de madera y productos relacionados con la agricultura.

A pesar de que las turbinas son competitivas en cuanto a precios con respecto a otros motores primarios, los costos de un sistema CHP completo (Sistema Combinado de Energía y Generación) es relativamente alto. Lo anterior es debido a que las turbinas son típicamente asociadas a bajas razones de generación eléctrica/energía y tienen altos costos de capital asociados con los sistemas de manejo de combustible y vapor, junto con la naturaleza de las instalaciones a las que pertenecen. Por lo que una aplicación ideal de sistemas CHP incluye industrias de mediana-larga escala o instituciones con recursos de altas cargas térmicas en que los desechos sólidos o combustibles se encuentren disponibles para el uso de las calderas, descripción que calza perfectamente con la generación energética que se da en el caso de Santa Fe Energía.

Cabe destacar que las turbinas a vapor son capaces de operar sobre un amplio rango de presiones de vapor. Están diseñadas especialmente para entregar los requerimientos en

términos de aplicaciones CHP mediante el uso de disminución de presión o extracción de vapor a la presión y temperatura necesitada. En general en términos de eficiencia estas alcanzan valores cercanos al 80%, aún cuando dependan del combustible empleado (EPA, 2013).

3.3.1 Ciclo de una planta de vapor.

Woodruff et al (2004) mencionan que el ciclo empleado más simple y de mayor valor práctico es el ciclo Rankine, originado para explicar el funcionamiento de las máquinas de vapor. El ciclo del vapor es importante debido a que conecta procesos que permiten convertir energía en trabajo en una base continua. Este sencillo ciclo se base en vapor sobrecalentado suministrado por una caldera a una unidad de generación energética como una turbina que acciona un generador eléctrico, luego el vapor saliente de la turbina ingresa a un condensador, donde en términos globales el vapor condensado es recirculado por la acción de bombas a la caldera. También se le denomina ciclo de condensación. La siguiente *Figura 3.3.* muestra un esquema simplificado del ciclo.

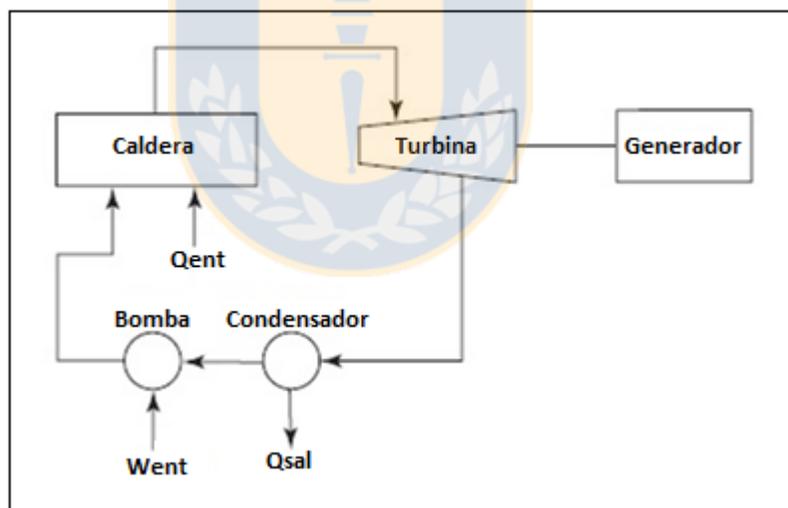


Figura 3.3. Ciclo Rankine.

La *Figura 3.3.* muestra la energía o calor de entrada Q_{ent} suministrada a la caldera. En el caso del presente estudio la energía de entrada se obtiene al quemar principalmente biomasa forestal en el lecho fluidizado interior de la Caldera de Biomasa N°2 de Santa Fe Energía, dicha combustión genera parte del vapor sobrecalentado alimentado al Turbogenerador N°4, el resto del vapor sobrecalentado se obtiene desde Santa Fe. Junto con la caldera y el calor de

entrada también se muestra la turbina conectada a un generador para la producción de electricidad. La electricidad como se mencionó anteriormente se vende al SIC (Sistema Interconectado Central). Q_{sal} representa la energía removida por el condensador, en forma de calor latente, mientras que la bomba suministra energía al condensado en la forma de un incremento de presión para permitirle regresar a la caldera. Para el caso del Turbogenerador N°4 se cuenta con un condensador total, una extracción de baja presión, y otra extracción de baja-baja presión. Por lo general en plantas de gran tamaño se emplea un recalentador (intercambiador de calor) adicional, de tal forma de añadir energía a la sección de baja presión de la turbina, aumentando así la eficiencia global de la planta.



Modelo lineal de optimización.

En el presente capítulo se plantea la definición y contexto del problema en estudio, definiendo brevemente el mercado energético y los fundamentos económicos que rigen la Central Santa Fe Energía. Luego se realiza el Análisis del Problema, para finalizar con el Modelamiento Matemático del mismo.

4.1 Definición del Problema.

Para el presente caso se plantea la simulación y optimización de una nueva variable asociada a la Central Santa Fe Energía, dicha variable será definida como un Específico de Generación Energética, es decir, cuántas toneladas, [ton vapor/h], de vapor es necesario inyectar a la central de tal forma de generar 1 [MWh] en la forma de energía eléctrica. Cabe destacar que resulta importante optimizar dicha generación puesto que la energía es el producto de venta al SIC por parte de la Central Santa Fe Energía. Una vez realizada la optimización será posible obtener nuevos valores de operación para las distintas variables que conforman el Sistema de Distribución de Vapor.

Previo a la optimización debe obtenerse el modelo matemático, que relaciona la variable de salida, en este caso el Específico de Generación, con las restantes variables de entrada del SDV, para lo cual se utiliza el software FACTNET mediante la simulación de dichas variables, a partir de un banco de datos y la aplicación de un proceso de minería de datos, a partir de la técnica de Análisis Factorial.

La Central Santa Fe Energía está constituida por la Caldera de Biomasa N°2 y el Turbogenerador N°4, no obstante también recibe vapor desde Planta Santa Fe, más específicamente de la Línea N°2 de Planta Santa Fe. En general dicho *input* de vapor se ve afectado en caso de haber bajas de generación en las calderas producto de situaciones anormales en la planta, y la generación de la Central debe obligadamente disminuir, puesto que en CMPC Planta Santa Fe prima la producción de celulosa.

Otro factor que afecta a la central es el bloque en que trabaja, llamaremos bloque a un rango de generación según la demanda externa que exista de compra energética. Existen cuatro bloques de generación; Bloque 1 a Bloque 4. Cada bloque posee un costo marginal, siendo el costo marginal del Bloque 1 el más bajo y el del Bloque 4 el más alto, punto que se abordará en la *Sección 4.4.1*. La Central Santa Fe Energía compete con otras centrales de generación eléctrica que proveen de energía al SIC, entre éstas, centrales de paso, centrales a carbón, y semejantes. En el caso de las centrales de paso éstas poseen ventaja en época invernal, puesto que la generación hídrica es mayor debido a los mayores caudales de los ríos, junto con bajos costos variables. Una vez que la temporada invernal finaliza, las centrales de paso disminuyen la generación eléctrica, debido a la menor disponibilidad de suministro hídrico que se da durante la época estival. Santa Fe Energía posee ciertas ventajas en época estival desde el punto de vista de generación, puesto que la biomasa a quemar en la caldera, tanto de procedencia interna como externa, está más seca, por lo que no se pierde energía en la caldera en evaporar agua, y por tanto gran parte de la energía es traspasada al agua que pasa por el interior de los tubos de la caldera para generación de vapor y posterior generación energética a través del paso por el condensador del Turbogenerador N°4. Además no compete tanto con centrales de paso, sino más bien con centrales semejantes como centrales a carbón, o de quema de material orgánico, entre otras.

4.1.1 Mercado Energético y Central Santa Fe Energía.

El mercado de la energía funciona de la siguiente forma: las distintas centrales de generación energética poseen costos fijos y costos variables, y distintos rangos de generación, además de sus costos marginales respectivos. Por otra parte existe una demanda global de energía por parte del país o regiones del país, luego, dicha energía debe ser provista por distintas fuentes de generación. Para el presente caso el órgano encargado de determinar las fuentes de provisión energética es el Centro de Despacho Económico de Carga – SIC (CDEC-SIC), el CDEC-SIC es el “organismo encargado de coordinar y determinar la operación de las instalaciones del SIC, incluyendo centrales generadoras, líneas y subestaciones del sistema de transmisión, y barras de consumo de clientes libres. Entre sus funciones se cuenta velar por la seguridad del servicio en el sistema eléctrico, garantizar la operación más económica del conjunto de las instalaciones del sistema eléctrico y garantizar el derecho a servidumbre sobre los sistemas de transmisión establecidos mediante decreto de concesión eléctrica. El CDEC-

SIC está integrado por las generadoras, transmisoras y clientes libres que operan en el SIC, y son sus integrantes las que lo financian. El marco institucional de los CDEC es el Decreto 291” (<http://www.centralenergia.cl/actores/autoridad-energia-chile/>). El SIC o Sistema Interconectado Central “está integrado por todas aquellas empresas pertenecientes a los segmentos de generación igual o mayor a 200 MW, generación inferior a 200 MW, transmisión troncal, subtransmisión y clientes libres, propietarios de instalaciones interconectadas al Sistema Interconectado Central en conformidad a lo establecido en el Decreto Supremo N° 291/2007 del Ministerio de Economía, Fomento y Reconstrucción (<http://www.cdec-sic.cl/sobre-cdec-sic/>)”. Luego, Los criterios de selección para las centrales son la necesidad energética total a suplir, el grado de generación y los costos variables de dichas centrales, por lo que una central de paso con alta generación y costo variable bajo puede ser una de las primeras en entrar a la lista, cubriendo un porcentaje de la generación, posteriormente entra una de generación con costo variable levemente mayor y generación igual o menor, cubriendo otro porcentaje del remanente del total, y así sucesivamente. Cabe destacar que a medida que van entrando centrales con costos variables más altos el costo del bloque se establece en dicho valor, por lo que las centrales que entraron antes con costos variables menores y posiblemente mayor generación se ven beneficiadas por el aumento de la barrera del costo marginal versus sus costos variables menores. En la *Figura 4.1.* se presenta la demanda de energía total mensual cubierta por el SIC en los años 2013-2014.

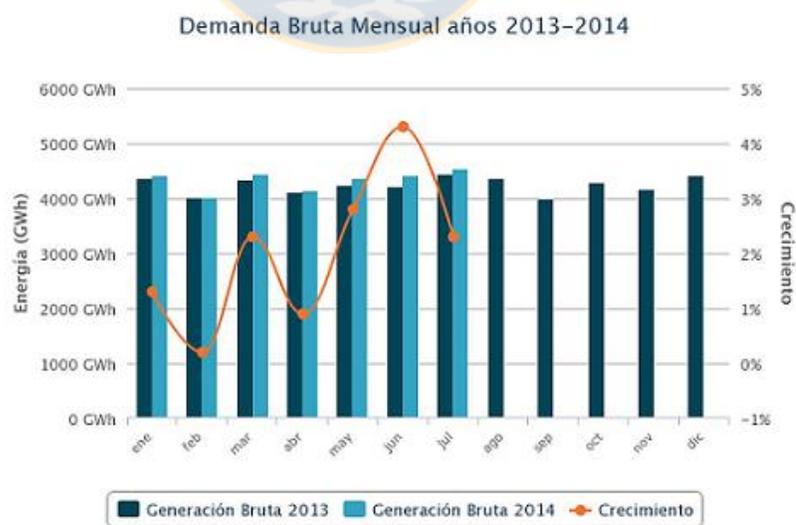


Figura 4.1. Demanda Bruta Mensual al SIC años 2013-2014. CDEC-SIC.

Mankiw (2007) define el costo marginal como el “aumento que experimenta el costo total cuando se produce una unidad más”. Luego, dado que el SIC corresponde a un mercado energético, si es pensado como un comprador su costo marginal será el delta que pagará a las empresas por los montos de generación energética al aumentar entre una y otra, y la cantidad se medirá en [US\$/MWh], intentando siempre cubrir el global de demanda de generación, o consumo energético, de particulares o empresas no generadoras.

Desde el punto de vista de Santa Fe Energía se obtendrá un ingreso por venta energética al SIC igual al costo marginal del bloque multiplicado por la generación en ese momento, y el ingreso final se calcula ponderando los distintos bloques y la generación de la Central Santa Fe Energía.

Por otro lado, los costos variables corresponden a compra de biomasa externa e interna y manejo del agua desmineralizada para generación en la Caldera de Biomasa 2, a lo anterior se le deben sumar los costos fijos (personal interno, externo, mantención, entre otros), luego los ingresos de Santa Fe Energía estarán dados por la diferencia entre la venta de la energía generada al SIC menos los costos totales de generación de la central.

Por las razones recientemente expuestas es que se modelará el Específico de Generación de Central Santa Fe Energía, puesto que si se encuentran los valores óptimos de generación esto permitirá maximizar los ingresos de la Central, generando en términos específicos mayor cantidad de energía en [MWh] con menor consumo de vapor en [ton vapor/h].

4.2 Análisis del problema.

El problema está constituido inicialmente por 39 variables seleccionadas que componen el Sistema de Distribución de Vapor. Dichas variables están correlacionadas debido a la sinergia e interdependencia que posee el sistema, a un nivel macro además unen Santa Fe con Santa Fe Energía, entidades que si bien poseen fines distintos no pueden actuar independientemente, dado que comparten traspasos de vapor de alta y baja presión, luego si Santa Fe funciona deficientemente en la generación y manejo de vapor esto afectará directamente a Santa Fe Energía y viceversa. La idea será determinar los valores operacionales que permitan maximizar la eficiencia de generación energética en función de minimizar el Específico de generación de la Central Santa Fe Energía, es decir, reducir la razón [ton vapor/h/MWh],

considerando una alta generación del Turbogenerador N°4 para obtener una mayor venta energética al SIC.

Para la reducción del set de variables originales se utiliza el software FACTNET, mediante diversos criterios que se explican en la *sección 4.3*. Es importante verificar que a medida que se realiza la reducción de la dimensión del problema la variabilidad explicada por las variables no disminuirá y será tal que permita una alta representatividad del sistema, expresada mediante el modelo matemático a obtener. Dicho modelo simula el Específico de Generación (variable de salida) en función de las variables de entrada. Una vez obtenido el modelo matemático se procede a su optimización, utilizando el software LINGO.

En la *Tabla 4.1*. se muestran las 39 variables iniciales pertenecientes al Sistema de Distribución de Vapor y sus unidades de medida. En la *Figura 4.2*. es posible observar la ubicación de las variables en el Sistema de Distribución de Vapor.

N°	Variable	Unidad de medida
1	Generación TG4	[MWh]
2	Condensado de salida TG4	[ton/h]
3	Vapor BP-BP salida TG4	[ton/h]
4	Vapor BP salida TG4	[ton/h]
5	Presión vapor BP salida TG4	[bar]
6	Temperatura salida vapor BP TG4	[°C]
7	Vapor para soplado CB2	[ton/h]
8	Vapor BP salida TG4 al colector	[ton/h]
9	Vapor AP entrada TG4	[ton/h]
10	T vapor AP entrada TG4	[°C]
11	P vapor AP entrada TG4	[bar]
12	Apertura V/V admisión TG4	%
13	Vapor a SF Energía desde SF	[ton/h]
14	Generación CB2	[ton/h]
15	Generación CR2	[ton/h]
16	Vapor AP entrada TG3	[ton/h]
17	Generación TG3	[ton/h]
18	Vapor MP salida TG3 a Colector MP SF2	[ton/h]
19	Vapor BP salida TG3 a Colector BP SF2	[ton/h]
20	Vapor AP entrada TG2	[ton/h]
21	Generación TG2	[ton/h]
22	Apertura V/V salida TG2 a Acumulador	%
23	Apertura V/V salida TG2 a Colector MP SF2	%
24	Vapor BP salida TG2 a Colector BP SF2	[ton/h]
25	Generación CB1	[ton/h]
26	Vapor CB1 hacia Acumulador	[ton/h]
27	Apertura V/V flujo total a Acumulador	%
28	Flujo vapor salida Acumulador a Colector MP SF2	[ton/h]
29	Apertura V/V bypass vapor Acumulador a Colector BP SF2	%
30	Apertura V/V vapor Acumulador a Colector BP SF1	%
31	Apertura V/V vapor Colector MP SF2 a Colector BP SF1	%
32	Apertura V/V vapor desde Colector BP SF1 a Colector BP SF2	%
33	Generación CR1	[ton/h]
34	Vapor entrada AP TG1	[ton/h]
35	Generación TG1	[ton/h]
36	Vapor MP salida TG1 a Colector MP SF1	[ton/h]
37	Vapor BP salida TG1 a Colector BP SF1	[ton/h]
38	Generación Central SFE Energía	[MWh]
39	Específico de Generación	[ton/h/MWh]

Tabla 4.1. Conjunto de variables iniciales del Sistema de Distribución de Vapor.

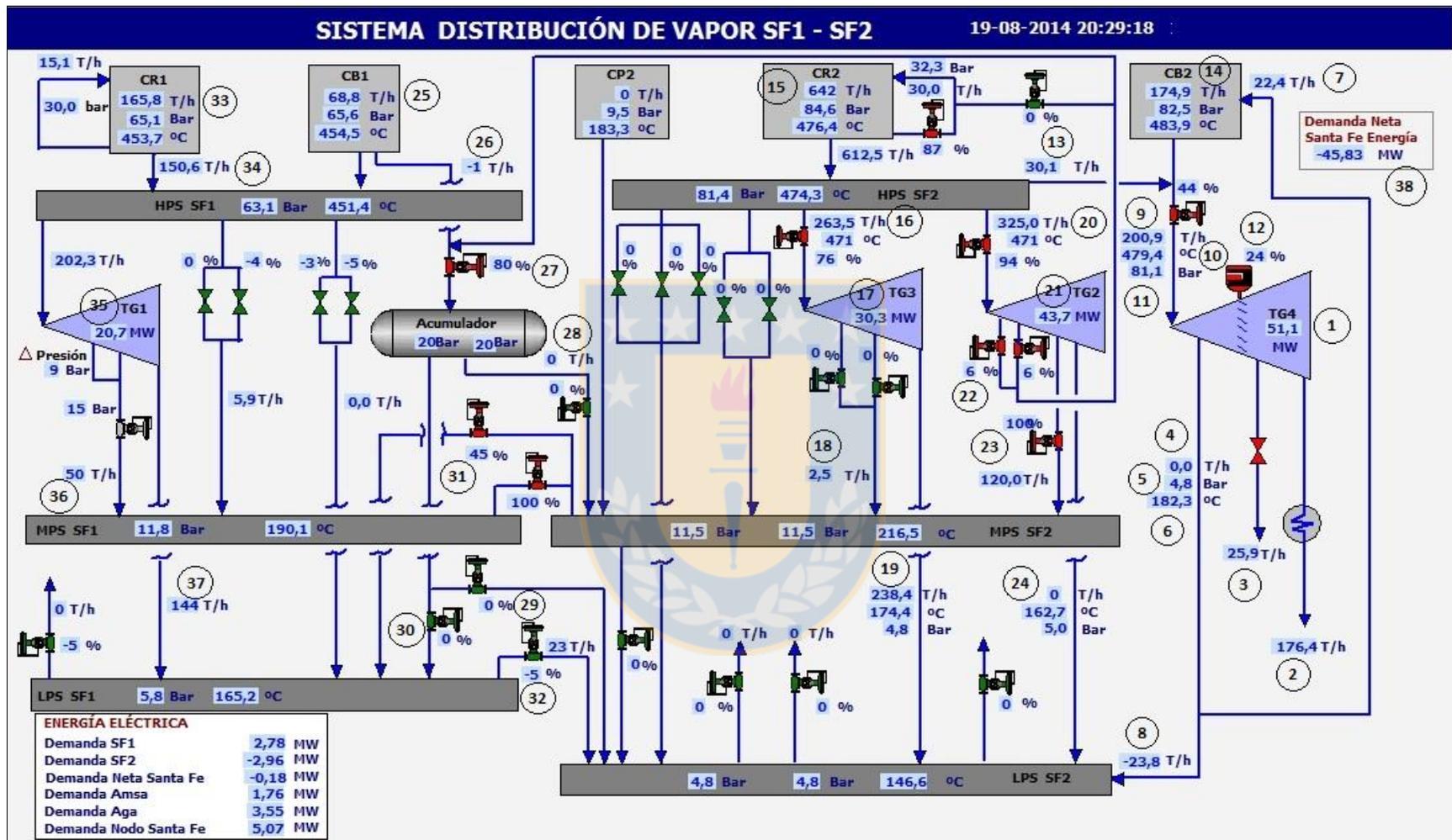


Figura 4.2. Variables iniciales en el Sistema de Distribución de Vapor.

La variable N°39 es el Específico de Generación de Central Santa Fe Energía, definido como

$$E.E.Gen = \frac{Vapor\ AP\ Entrada\ TG\ 4}{Generación\ Central\ Sta\ Fe\ Energía} [=] \left[\frac{ton/h}{MWh} \right] \quad (2)$$

representando las [ton/h] que es necesario inyectar al Turbogenerador N°4 para generar 1 [MWh] por parte de la Central Santa Fe Energía. Lo conveniente es reducir ese valor, operando a la vez con los valores más altos posibles de entrada de vapor de alta presión al turbo, es decir, cuando la Caldera de Biomasa 2 se encuentra generando a tope y se recibe una cantidad considerable de vapor de alta presión desde Santa Fe, como se ha mencionado previamente.

Teniendo en cuenta las variables que participan inicialmente del modelo es necesario definir el marco global del problema propuesto. En la *Figura 4.3.* se presenta un esquema resumen del proceso de recolección de datos, minería de datos, modelamiento del problema y optimización.

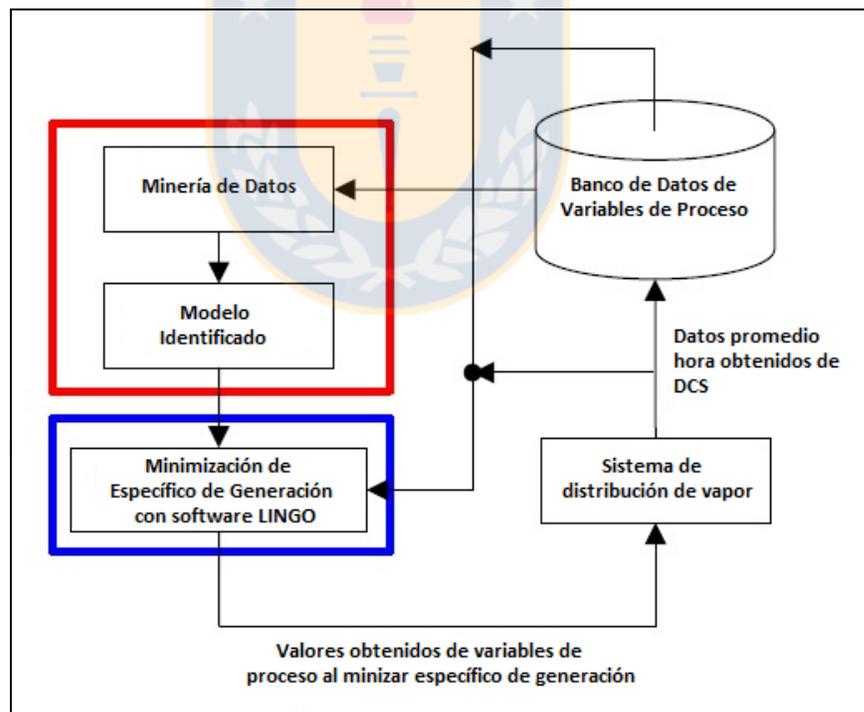


Figura 4.3. Recolección y minería de datos, modelamiento y optimización.

En la *Figura 4.3* se observa un resumen del proceso a llevar a cabo para obtener los valores de las variables de entrada del proceso que permitan minimizar el específico de generación, el orden del proceso es el siguiente:

1) Recolección y generación del Banco de Datos.

La recolección del Banco de Datos se obtiene del Sistema de Control Distribuido o DCS (Distributed Control System), herramienta con la que cuenta Planta Santa Fe, que permite la recolección de un elevado número de datos en forma continua por intervalos de minutos, días, meses, años, o la escala en la que se desee obtener dicha información. A su vez gran parte de los datos del sistema DCS son enviados al sistema PI (OSIsoft, <http://www.osisoft.com>), software que consiste en una interfaz visual que permite visualizar distintas representaciones gráficas, en forma de diagramas de proceso, de diversas áreas del proceso global de fabricación de celulosa, y en el caso de Santa Fe Energía, venta energética. El software PI permite graficar tendencias de las variables de manera separada y conjunta (más de una variable a la vez) en distintos rangos de tiempo, como días, meses o años. Otro factor importante del software PI es que permite la creación continua de pantallas, con un editor denominado Process Book (OSIsoft), por lo que si se desean agregar nuevas variables solo será necesario crearlas con un método establecido, luego si se reconocen nuevas variables claves o indicadores de rendimiento del proceso, será posible agregarlas, en este caso un nuevo indicador de rendimiento será el Específico de Generación de la Central Santa Fe Energía, por lo que es necesario agregarlo más adelante como variable al sistema PI. Lo anterior se realizará para la representación gráfica del Sistema de Distribución de Vapor existente para el software PI en planta.

Otra herramienta valiosa del sistema PI es que contiene una indexación con Excel, a través de dicha interacción es posible descargar históricos de datos de variables. Cabe destacar que el software PI no solo se encuentra indexado al sistema DCS, sino que también a distintas planillas de datos que se llenan en planta con variables de proceso y similares.

Los valores obtenidos de las variables del Sistema de Distribución de Vapor se obtienen como datos promedio hora a hora, considerando todo el mes de enero del año 2014. A partir de 39 variables, que posteriormente se reducen mediante distintos criterios que posee el software FACTNET, detallados en la sección *modelamiento del problema*.

2) Minería de datos.

La minería de datos es llevada a cabo con la ayuda del software FACTNET Pacific Simulation, que a través del Método de Componentes Principales (PCA), incluido como una sub-sección del Análisis Factorial, realiza la reducción de la dimensionalidad del banco de datos, obteniendo así un número menor de variables que representen una variabilidad alta del modelo predictivo. Las variables reducidas a su vez se encuentran interrelacionadas por una red de factores, denominados Factores Comunes, extraídos a partir del método PCA. Dicha red de factores muestra la relación entre las distintas variables de entrada y la variable de salida, en este caso el específico de generación.

Existe cierta dicotomía en la literatura respecto a las definiciones de qué abarca y qué no abarca el análisis factorial y el método de componentes principales, Jolliffe (2005) explica que el “método de componentes principales a menudo es tratado como un caso especial del análisis factorial, y que esta práctica fue continuada por algunos paquetes informáticos (softwares) ampliamente usados, tratando al método PCA como una opción dentro de los programas de Análisis Factorial” y que esto es errado. Precisamente éste es el método que emplea FACTNET, es decir, utiliza indistintamente el método PCA como si se tratase de una parte del análisis factorial, aunque más precisamente emplea un método híbrido que a la larga permite la determinación de la red de factores comunes-variables de proceso y las ecuaciones de ganancia. Más allá de la veracidad o falsedad de dicha premisa, en el modelamiento matemático del problema se explican los pasos teóricos que involucran el proceso del análisis factorial, y dentro de los resultados se comprueba la veracidad de dicho método.

3) Identificación del modelo e interpretación de relación entre variables.

El modelo se rescata a partir de las ecuaciones de ganancia obtenidas del análisis factorial, y más precisamente del método de extracción de factores utilizando el software FACTNET, además de las restricciones operacionales en cuanto a los valores que pueden alcanzar las variables de proceso. Previamente se emplean distintos criterios para ir disminuyendo la dimensión del problema, en cuanto al número de variables finales, y el número de factores comunes, explicados en la sección de modelamiento del problema. Lo importante en dicho caso es que el porcentaje de variabilidad explicada por el modelo obtenido sea lo suficientemente alto de tal forma que permita que el modelo efectivamente “prediga” los

valores de las variables de entrada que permitan minimizar la variable de salida. Además se puede ver las relaciones entre pares de variables, es decir, si una variable afecta positiva o negativamente a la otra, en qué cantidad afecta una variable a la otra, y el grado de correlación o dependencia entre éstas. Lo anterior se evaluará con más detalle en la sección de resultados.

4) Optimización del modelo obtenido.

Una vez obtenido el modelo se procede a realizar la optimización de éste con el software LINGO, para esto se dispone de, las ecuaciones de ganancia que permitirán definir la función objetivo o función que relaciona la variable de salida con las variables de entrada, de las ecuaciones que relacionan las variables de entrada entre sí, y finalmente de los límites operacionales o restricciones de las variables de entrada y variable de salida.

5) Interpretación de valores obtenidos y nuevo conocimiento.

Finalmente se hace un análisis en la sección de resultados de los valores obtenidos para las distintas variables de proceso, grado de veracidad del modelo, relación entre variables y se evalúa la obtención de nuevo conocimiento.

4.3 Modelamiento del Problema.

4.3.1 Banco de datos y Reducción de Dimensionalidad.

Los datos recopilados son los correspondientes al mes de Enero de 2014, la bajada de datos fue realizada para 39 variables originales, pertenecientes al Sistema de Distribución de Vapor, posteriormente filtradas según criterios guía del software FACTNET, mencionados a continuación.

1) Factor Filter/Filtro Factorial.

Este filtro sirve para evaluar las relaciones causa-efecto existentes entre variables establecidas por el modelo la de red de factores. El detalle de cómo emplear esta herramienta se incluye en el Anexo. Lo que entrega es un ranking de las variables en análisis con respecto a cada una de las variables restantes, la variable a evaluar se selecciona previamente. Además de entregar un ranking entrega los valores de las ganancias de las ecuaciones de ganancia de cada variable con respecto a la variable evaluada. Las ecuaciones de ganancia son ecuaciones que permiten obtener una variable en función de las restantes variables, y son detalladas más adelante en el

modelamiento matemático del problema. Como regla general sugerida por la metodología del programa FACTNET (Pacific Simulation, 2005) se considera que relaciones con un ranking menor a 0,05 no son significativas para la variable en estudio, por lo que en el presente trabajo son removidas.

La variable elegida para realizar el Filtro Factorial fue la variable de salida, Específico de Generación, no obstante esto puede realizarse para el resto de las variables indistintamente. Cabe destacar que esto resulta útil para el examinador del sistema, ya que permite conocer el grado de correlación de una variable con respecto a otra, y cómo y en cuánto el aumento o decremento de una variable puede afectar a la variable en análisis. En la *Tabla 4.2.* se muestra el ranking obtenido una vez aplicado el filtro factorial, en orden decreciente, con respecto a la variable de salida Específico de Generación de Central Santa Fe Energía.

Filtro Factorial, Ranking > 0,05		
N°	Variable	Ranking
1	Vapor Entrada AP TG4	64,159
2	Condensado de salida TG4	57,952
3	Vapor BP salida TG4	30,709
4	Vapor BP-BP salida TG4	8,704
5	Generación Central Santa Fe Energía	0,424
6	Generación TG4	0,375
7	Vapor BP TG4 al colector	0,139
8	Temperatura de salida vapor BP TG4	0,118
9	Apertura V/V admisión vapor TG4	0,095
0	Presión vapor AP entrada TG4	0,056

Tabla 4.2. Ranking de variables obtenido mediante el Filtro Factorial.

Las variables que se muestran en *Tabla 4.2.* son hasta el momento con las que se confeccionará el modelo, aunque aún debe considerarse un segundo filtro explicado a continuación.

2) Pre-Clasificación.

La ventana de Pre-clasificación muestra las variables de la red de factores en un orden de preclasificación según su importancia. Cuando se trabaja con un gran número de variables resulta útil remover variables que muestren poca correlación con el resto de las variables, por lo que FACTNET permite a los usuarios pre-clasificar los datos, calculando el valor promedio del coeficiente de determinación, R^2 , para cada variable con respecto a las otras variables. Un criterio sugerido por Pacific Simulation (2005) es remover las variables cuyo valor de R^2 se encuentra bajo 0,01 o 0,02. Para el presente trabajo se eliminan las variables cuyo valor de R^2 es menor a 0,02. En la *Tabla 4.3.* se listan las variables con R^2 menor a 0,02.

N°	Variable
1	Presión de vapor BP salida TG4
2	% Apertura válvula flujo total acumulador
3	Generación CB1
4	% apertura válvula vapor acumulador a colector BP SF1
5	Flujo vapor desde CB1 hacia Acumulador

Tabla 4.3. Variables con R^2 menor a 0,02.

En este caso ninguna de las variables con R^2 menor a 0,02 se encuentra en la lista de variables consideradas para el modelo por el Filtro Factorial, por lo que los análisis con consistentes, y no es necesario eliminar ninguna de las 10 variables de entrada obtenidas por el método.

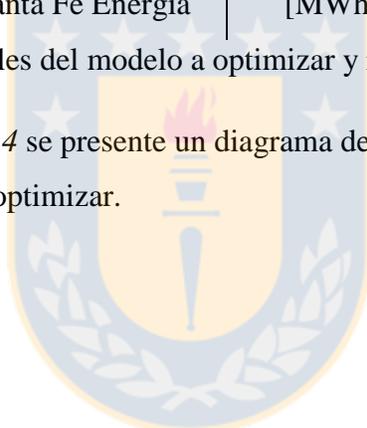
4.3.2 Variables que componen el Modelo y Límites Operacionales.

En la *Tabla 4.4.* se listan las variables que conforman el modelo a optimizar una vez realizado el Filtro Factorial y la Pre-clasificación, además se incorporan los límites o restricciones operacionales a las que se encuentran sometidas.

	Variable	Unidad de medida	Límite Inferior	Límite Superior
y	Específico de Generación	[ton/h/MWh]	3,80	5,50
x ₁	Generación TG4	[MWh]	17,28	89,92
x ₂	Condensado de salida TG4	[ton/h]	96,23	290,30
x ₃	Vapor BP-BP salida TG4	[ton/h]	11,10	41,20
x ₄	Vapor BP salida TG4	[ton/h]	0,00	86,82
x ₅	Temperatura de salida vapor BP TG4	[°C]	156,10	194,30
x ₆	Vapor BP TG4 al colector	[ton/h]	-20,30	60,28
x ₇	Vapor Entrada AP TG4	[ton/h]	108,30	330,90
x ₈	Presión vapor AP entrada TG4	[bar]	79,40	82,70
x ₉	Apertura V/V admisión vapor TG4	%	15,78	99,30
x ₁₀	Generación Central Santa Fe Energía	[MWh]	12,02	83,39

Tabla 4.4. Variables del modelo a optimizar y restricciones operacionales.

A continuación en *Figura 4.4* se presente un diagrama de flujo simplificado de las variables que componen el modelo a optimizar.



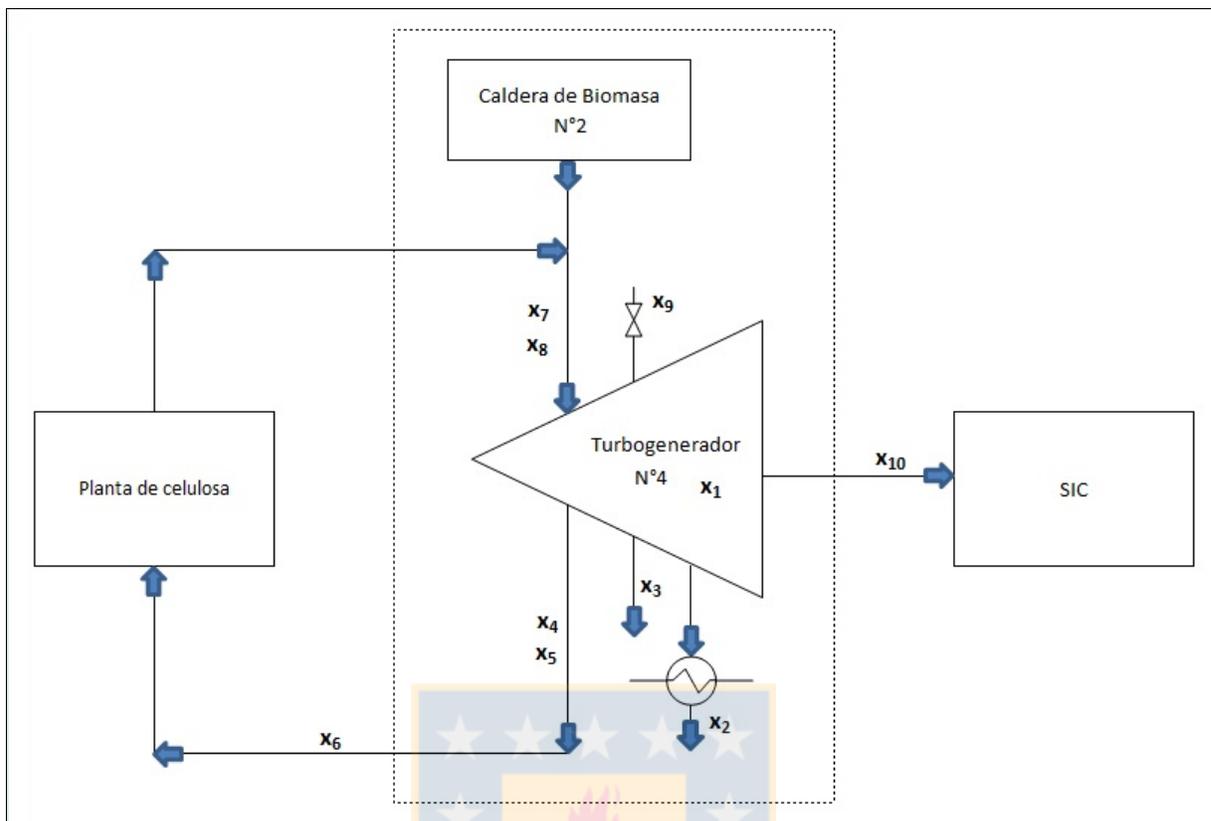


Figura 4.4. Diagrama de flujo simplificado

A continuación se detalla en qué consiste cada variable que compone el modelo a optimizar:

- Específico de generación, y : corresponde a la variable de salida, definida como la razón entre el flujo de vapor de alta presión en [ton/h] que ingresa al turbogenerador N°4 (x_7) y la generación eléctrica de Central Santa Fe Energía (x_{10}). Entre menor sea esta razón más eficiente será la central, puesto que para generar 1 MWh de energía eléctrica por parte de la central se requerirán menos [ton/h] de vapor de alta presión que ingresen al turbogenerador N°4. Lo anterior no implica que se necesiten flujos máscicos totales más bajos de entrada de vapor de alta presión al turbogenerador, sino más bien que el total de entrada de vapor de alta presión que entra al turbogenerador será empleado de forma más efectiva en la generación energética.
- Generación Turbogenerador N°4, x_1 : es la energía eléctrica total generada por el turbogenerador N°4. El vapor de alta presión que entra al turbogenerador entrega su energía mecánica para mover el rotor del turbogenerador, siendo ésta posteriormente transformada en energía eléctrica por el generador asociado a la turbina.

- Condensado de salida TG4, x_2 : corresponde al líquido saturado que sale luego del condensador perteneciente a la unidad que comprende el turbogenerador. Una vez que pasa por el rotor de la turbina, el vapor de alta presión disminuye su entalpía pasando de una presión de 85 [bar] hasta una presión cercana a los 5 [bar], siendo este cambio entálpico el equivalente a la energía mecánica suministrada al rotor de la turbina. Una vez que pasa por el rotor de la turbina, el vapor saturado a 5 [bar] ingresa parcialmente al condensador, saliendo de éste como líquido saturado que luego de un proceso de tratamiento ingresa nuevamente a la caldera de biomasa N°2 para generación de vapor. Se menciona que el vapor saturado ingresa parcialmente al condensador puesto que un remanente se recircula (antes de entrar al condensador) al colector de baja presión, dicho vapor recirculado se conoce como vapor de baja presión (x_6) y otro remanente se emplea como vapor de baja-baja presión (x_3) para precalentar el agua desmineralizada (condensado) que ingresa a la caldera de biomasa N°2.
- Vapor BP-BP salida TG4, x_3 : este flujo másico corresponde a un vapor saturado (5 [bar], baja presión) remanente que no entra al condensador ni es recirculado al colector de baja presión de planta Santa Fe. No obstante, este vapor es empleado para precalentar el agua desmineralizada que se emplea para generación de vapor en la caldera de biomasa N°2.
- Vapor BP salida TG4, x_4 : como se mencionó anteriormente este flujo corresponde al vapor saturado (5 [bar]) total que sale del rotor de la turbina. Una parte de este se recircula al colector de baja presión de planta Santa Fe (x_6) y la otra parte se recircula para ayudar en a aumentar la temperatura del estanque de agua de alimentación de la caldera. El flujo recirculado a la caldera no se considera como variable, ya que se define según sean los requerimientos de vapor hacia el colector de baja presión de planta Santa Fe, restándose por tanto al flujo de vapor de baja presión que sale del turbogenerador lo requerido por planta Santa Fe, y siendo este remanente el recirculado a la caldera de biomasa N°2.
- Temperatura de salida vapor BP TG4, x_5 : es la temperatura del vapor saturado de baja presión (x_4).
- Vapor BP TG4 al colector, x_6 : es el flujo másico de vapor de baja presión (5 [bar]) que vuelve al colector de baja presión de planta Santa Fe, desde Central Santa Fe Energía.
- Vapor Entrada AP TG4, x_7 : corresponde al flujo másico de vapor de alta presión (85 [bar]) que ingresa al turbogenerador N°4. Dicho flujo másico se obtiene por balance de materia al sumar el flujo másico de vapor de alta presión que llega desde planta Santa Fe a Central

Santa Fe Energía con el vapor de alta presión generado por la caldera de biomasa N°2 (en [ton/h]).

- Presión vapor AP entrada TG4, x_8 : como su nombre lo dice corresponde a la presión que tiene el flujo de vapor de alta presión que ingresa al turbogenerador N°4. Mencionar que el vapor que ingresa al turbogenerador N°4 es vapor sobrecalentado, es decir, vapor sobre su temperatura de saturación.
- Apertura V/V admisión vapor TG4, x_9 : corresponde al grado de apertura ([%]) de la válvula que permite el ingreso de vapor de alta presión al turbogenerador N°4.
- Generación Central Santa Fe Energía, x_{10} : corresponde a la generación eléctrica neta de la Central Santa Fe Energía, en [MWh]. Esta generación eléctrica es menor que la que presenta el turbogenerador N°4, ya que la central se autoabastece, es decir, suplente su propia demanda eléctrica interna, el remanente es vendido como energía eléctrica al SIC (Sistema Interconectado Central). La generación del turbogenerador N°4 se conoce también como generación eléctrica bruta de la central.

4.3.3 Modo de Cálculo.

El modo de cálculo elegido para la modelación del problema es el Modo *Input/Output* (modo de entrada/salida). El programa FACTNET posee dos modos de cálculo, *Filter Mode* (modo de filtro) y el modo de entrada/salida. En el caso del modo de filtro se emplean las variables de entrada y variable(s) de salida para el cálculo de los factores comunes y la red que los relaciona con las variables, mientras que en el caso del modo entrada/salida se emplean solamente las variables de entrada para el cálculo de la(s) variable(s) de salida, por tanto actúa como modelo predictivo.

Pacific Simulation (2005) plantea que el modo de entrada/salida es “comúnmente empleado para desarrollar modelos que realizan predicciones en línea de variables que no pueden ser medidas en línea”, como en el caso de predicciones de propiedades que definen la calidad de la pulpa o del papel. Además se menciona que los datos analizados “usualmente contienen datos de entrada altamente correlacionados” como es el caso del Sistema de Distribución de Vapor.

4.3.4 Variabilidad del Modelo.

La variabilidad del modelo se explica a través de los valores propios o eigenvalores de los factores comunes. Los valores propios se definen como la suma de los cuadrados de los coeficientes de correlación para cada factor común. Por ejemplo, si se tienen 25 variables en total y el valor propio de un factor tiene un valor 5, entonces ese factor común explica el 20% de la variabilidad de las 25 variables ($100 \cdot 5 / 25$). Luego un segundo factor puede obtener un valor de 10, explicando un 40% de la variabilidad del problema. Combinando los dos factores se obtendría entonces un total de representatividad del 60% de la variabilidad. Cabe destacar que uno de los fines a la hora de la creación del modelo es obtener un alto porcentaje de representatividad de la variabilidad del problema.

4.3.5 Algoritmo de Cálculo.

FACTNET cuenta con cuatro algoritmos distintos de cálculo. El elegido para el estudio en cuestión es el método de Extracción de Factores, o Análisis de Componentes Principales (PCA). Este método calcula los coeficientes de saturación o ponderaciones factoriales por extracción factorial, mientras que los coeficientes de ganancia son calculados por regresión. En la *sección 4.3.7.* se explica con más detalle en qué consisten los coeficientes de saturación y los coeficientes de ganancia, junto con las ecuaciones que gobiernan dichos coeficientes.

Los algoritmos restantes son; Extracción de Factores por Pesos, Regresión de Mínimos Cuadrados Parciales, y Regresión de Mínimos Cuadrados Parciales en Modo Entrada/Salida.

4.3.6 Partición del conjunto de datos.

Una vez construido el banco de datos éste se separa en dos partes; una parte se utiliza para analizar los datos y construir un modelo de red factorial, mientras que la parte restante se emplea para construir otro modelo que valide el primero.

Para el presente trabajo se cuenta con un banco de datos comprendido por todo el mes de enero de 2014 con mediciones promedio horarias, correspondientes a 744 mediciones por variable. Para la construcción del modelo final se consideran, luego del proceso de tratamiento de datos y filtrado, un total de 11 variables, de las cuales 10 son variables de entrada, por lo que el total de datos bruto del modelo asciende a 8184 unidades. Como set de entrenamiento se utiliza el 80% de los datos, mientras que el 20% restante se emplea para validación. El 80%

de los datos corresponde a 11 grupos de 596 datos con 6556 mediciones, mientras que el 20% corresponde a 11 grupos de 148 datos con 1628 mediciones.

4.3.7 Modelo Matemático.

Para la modelación matemática se parte de lo macro a lo micro, en este caso el modelo matemático global a optimizar se describe en la *sección 4.3.7.1*.

4.3.7.1 Modelo Matemático Macro.

Una vez realizada la simulación con el programa FACTNET es posible modelar el problema de optimización de la siguiente forma

$$\begin{aligned}
 & \min y \\
 & \text{s.a. } y = f(x_1, \dots, x_n) \\
 & l_i \leq x_i \leq L_i \quad i = 1, \dots, n \\
 & x_i = x_i(x_1, x_2, \dots, x_n) \text{ con } i \in \{1, \dots, n\}
 \end{aligned} \tag{3}$$

La variable de salida del modelo es y , x_i es la i -ésima variable de entrada del modelo ($i = 1, \dots, n$), a y b son el límite inferior y superior de la variable y respectivamente, y l_i y L_i son el límite inferior y superior de la variable x_i respectivamente. El valor de m ($m \leq n$) dependerá de su factibilidad y de la necesidad de incluir ecuaciones de tal forma de mejorar la explicación del modelo con respecto a los datos observados. Luego, el número mínimo de ecuaciones que relacionan las variables de salida será l y el número máximo será n . Las ecuaciones $y = f(x_1, \dots, x_n)$ y $x_i = x_i(x_1, x_2, \dots, x_n)$ representan las ecuaciones de ganancia obtenidas a través del método de extracción de factores, realizado por el programa FACTNET. La notación empleada para modelar las variables es la siguiente:

siguiente:

$y = \text{Específico de Generación}$
 $x_1 = \text{Generación TG 4}$
 $x_2 = \text{Condensado de Salida TG 4}$
 $x_3 = \text{Vapor de BP - BP Salida TG 4}$
 $x_4 = \text{Vapor de BP Salida TG 4}$
 $x_5 = \text{Temperatura de Salida Vapor BP TG 4}$
 $x_6 = \text{Vapor BP TG 4 al Colector}$
 $x_7 = \text{Vapor de Entrada AP TG 4}$
 $x_8 = \text{Presión Vapor AP Entrada TG 4}$
 $x_9 = \text{Apertura V/V Admisión Vapor TG 4}$
 $x_{10} = \text{Generación Central Santa Fe Energía}$

4.3.7.2 Modelo Matemático Micro.

En esta parte de la modelación se obtienen las ecuaciones de ganancia que permiten construir el modelo macro. Las ecuaciones de ganancia se obtienen al combinar las ecuaciones de puntuación y las ecuaciones de saturación, a su vez determinadas por medio del método de extracción de factores empleado por FACTNET, denominado indistintamente como método de componentes principales (PCA). A continuación, a modo de introducción al tema se describe brevemente el concepto de análisis factorial y cuál es su finalidad.

Análisis Factorial.

Jolliffe (2005) plantea que la principal idea subyacente al análisis factorial es que p variables aleatorias observables, x , puedan ser expresadas, excepto por un término de error, como funciones lineales de k ($k < p$) variables aleatorias hipotéticas o factores comunes. Luego si x_1, x_2, \dots, x_p son las variables y f_1, f_2, \dots, f_k son los factores se tiene que

$$\begin{aligned}
 x_1 &= \lambda_{11}f_1 + \lambda_{12}f_2 + \dots + \lambda_{1k}f_k + e_1 \\
 x_2 &= \lambda_{21}f_1 + \lambda_{22}f_2 + \dots + \lambda_{2k}f_k + e_2 \\
 &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\
 x_p &= \lambda_{p1}f_1 + \lambda_{p2}f_2 + \dots + \lambda_{pk}f_k + e_p
 \end{aligned} \tag{4}$$

Para el caso de la ecuación anterior, y como se verá posteriormente, los coeficientes $\lambda_{ij}, i = 1, 2, \dots, p ; j = 1, 2, \dots, k$ corresponden a los denominados coeficientes de saturación, y

$e_j, j = 1, 2, \dots, p$ corresponden a términos de error, denominados en ocasiones como *factores específicos*, puesto que son únicos para una variable x_j , mientras que los factores comunes, f_k , se repiten para distintos x_j .

Ecuaciones de Puntuación.

Las ecuaciones de puntuación relacionan cada factor común, con cada variable considerada para el modelo, generando parte de la red factorial. Dicho grado de correlación se obtiene al aplicar el método de análisis factorial, como se mencionó previamente en la *sección 4.3.5.*, obteniendo así los coeficientes de las ecuaciones, denominados coeficientes de puntuación. Las ecuaciones tienen la siguiente forma:

$$\begin{aligned}
 Factor_1 &= a_{1,0} + a_{1,1} \cdot Variable_1 + a_{1,2} \cdot Variable_2 + \dots + a_{1,n} \cdot Variable_n \\
 Factor_2 &= a_{2,0} + a_{2,1} \cdot Variable_1 + a_{2,2} \cdot Variable_2 + \dots + a_{2,n} \cdot Variable_n \\
 &\vdots \\
 Factor_k &= a_{k,0} + a_{k,1} \cdot Variable_1 + a_{k,2} \cdot Variable_2 + \dots + a_{k,n} \cdot Variable_n
 \end{aligned} \tag{5}$$

dónde:

- Factor_i: i-ésimo factor común.
- a_{i,j}: coeficientes de puntuación.

Cabe destacar que, por definición, cada factor común es cien por ciento independiente del resto de los factores, lo que en lenguaje matemático es equivalente a decir que se definen como ortogonales, es decir, que uno no puede ser obtenido como combinación lineal del resto y viceversa.

Ecuaciones de Saturación.

En forma análoga a cómo lo realizan las ecuaciones de ponderación, las ecuaciones de saturación relacionan cada variable con cada factor común. Para el caso en estudio el método de extracción factorial utilizado por FACNET calcula los coeficientes de saturación o ponderaciones factoriales por medio de regresión. Las ecuaciones adoptan la siguiente forma

$$\begin{aligned}
Variable_1 &= b_{1,0} + b_{1,1} \cdot Factor_1 + b_{1,2} \cdot Factor_2 + \dots + b_{n,k} \cdot Factor_k \\
Variable_2 &= b_{2,0} + b_{2,1} \cdot Factor_1 + b_{2,2} \cdot Factor_2 + \dots + b_{n,k} \cdot Factor_k \\
&\vdots && \vdots && \vdots \\
Variable_n &= b_{n,0} + b_{n,1} \cdot Factor_1 + b_{n,2} \cdot Factor_2 + \dots + b_{n,k} \cdot Factor_k
\end{aligned}
\tag{6}$$

dónde:

- Variable_i: i-ésima variable.
- b_{i,j}: Coeficientes de Saturación o Ponderaciones Factoriales.

Ecuaciones de Ganancia.

Una vez determinados los coeficientes de puntuación y coeficientes de saturación, y por tanto las ecuaciones de puntuación y ecuaciones de saturación, es posible obtener las ecuaciones de ganancia al reemplazar las ecuaciones de puntuación en las ecuaciones de saturación. Las ecuaciones quedan de la siguiente manera:

$$\begin{aligned}
Variable_1 &= c_{1,0} + c_{1,1} \cdot Variable_1 + c_{1,2} \cdot Variable_2 + \dots + c_{n,k} \cdot Factor_k \\
Variable_2 &= c_{2,0} + c_{2,1} \cdot Variable_1 + b_{2,2} \cdot Variable_2 + \dots + c_{n,k} \cdot Variable_k \\
&\vdots && \vdots && \vdots \\
Variable_n &= c_{n,0} + c_{n,1} \cdot Variable_1 + c_{n,2} \cdot Variable_2 + \dots + c_{n,k} \cdot Variable_k
\end{aligned}
\tag{7}$$

dónde:

- C_{i,j}: Ganancias del Modelo.

Una vez obtenidas las ecuaciones de ganancia, y con las restricciones operacionales es posible plantear el modelo macro. La principal ecuación de ganancia en este caso es la de la variable de salida Específico de Generación, correspondiente a la función objetivo $y = f(x_1, \dots, x_n)$ del modelo macro, el resto de las ecuaciones de ganancia corresponden a las de las variables de entrada, ecuaciones que complementan el modelo y son usadas para ir aumentando el grado de certidumbre de éste.

Resultados y Discusión

En la presente sección se presenta y analiza el modelo de simulación de la red factorial obtenida a partir de las variables que componen el Sistema de Distribución de Vapor, y los resultados obtenidos a partir de éste. El modelo se obtiene mediante un proceso de minería de datos, luego se procede a su optimización. Dicho proceso se sustenta en el método de análisis factorial, realizado con el software FACTNET. Para lo anterior se llevó a cabo la recolección de un banco de datos formado por 8184 mediciones, correspondientes a 11 variables con 744 mediciones cada una. Dicho banco de datos se obtuvo luego de reducir un sistema con 39 variables y 744 mediciones mediante los procesos de filtro factorial y pre-clasificación.

A continuación se muestra la red factorial inicial, y red factorial final obtenida; conteniendo las variables de entrada y la variable de salida iniciales y finales respectivamente, junto con los factores comunes obtenidos para el modelo luego del filtrado de variables. Cabe destacar que se incluye el valor de variabilidad explicada por los eigenvalores de cada factor común en cuestión para la variable a predecir Específico de Generación de Central Santa Fe Energía. Posteriormente se incluyen los resultados obtenidos del filtro factorial y coeficientes de correlación. Una vez descrito lo anterior se presenta el modelo matemático, para posteriormente proceder a mostrar el resultado de la minimización de la variable de salida Específico de Generación de Central Santa Fe Energía, obteniendo los valores que permiten dicha minimización para las 10 variables de entrada. Finalmente se incluyen análisis referentes a la exactitud del modelo, correlación entre variable calculada y medida, variabilidad cubierta por el modelo, errores y desviaciones de este, entre otros.

5.1 Red Factorial Inicial y Final obtenida.

Previo a la obtención de las variables finales que configuran el modelo fue necesario disminuir la dimensión del problema, obteniéndose 11 variables finales en total; 10 variables de entrada y una variable de salida. En la *Figura 5.1.* se muestra la red factorial inicial obtenida, sobre la cual se aplicaron los criterios de Filtro Factorial y Pre-clasificación.

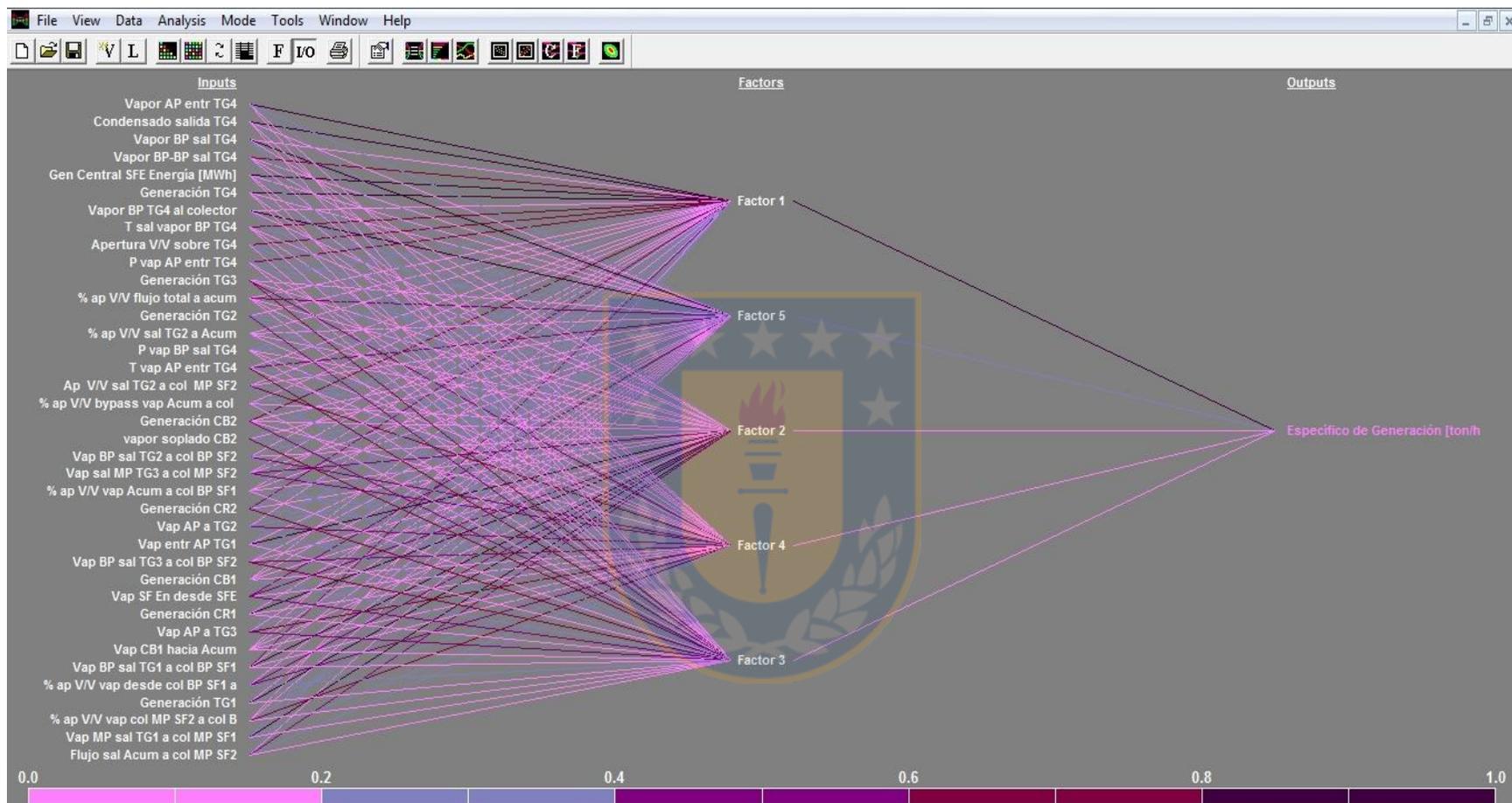


Figura 5.1. Red Factorial inicial.

Como se observa en la *Figura 5.1.*, se tiene una red que incluye 39 variables iniciales; 38 variables de entrada y la restante de salida, más 5 factores comunes. Conceptualmente resulta complejo comprender las interrelaciones entre 39 variables, además no todas las variables descritas dentro del modelo inicial son representativas de cambios sustantivos en el Específico de Generación, por lo que se redujeron las variables iniciales mediante los métodos de filtro factorial y pre-clasificación, descritos previamente en la *sección 4.3.1.* Las variables obtenidas son las siguientes:

$$y = \text{Específico de Generación } (=) [\text{ton} / \text{h} / \text{MWh}]$$

$$x_1 = \text{Generación TG 4 } (=) [\text{MWh}]$$

$$x_2 = \text{Condensado de Salida TG 4 } (=) [\text{ton} / \text{h}]$$

$$x_3 = \text{Vapor de BP - BP Salida TG 4 } (=) [\text{ton} / \text{h}]$$

$$x_4 = \text{Vapor de BP Salida TG 4 } (=) [\text{ton} / \text{h}]$$

$$x_5 = \text{Temperatura de Salida Vapor BP TG 4 } (=) [^{\circ}\text{C}]$$

$$x_6 = \text{Vapor BP TG 4 al Colector } (=) [\text{ton} / \text{h}]$$

$$x_7 = \text{Vapor de Entrada AP TG 4 } (=) [\text{ton} / \text{h}]$$

$$x_8 = \text{Presión Vapor AP Entrada TG 4 } (=) [\text{bar}]$$

$$x_9 = \text{Apertura V/V Admisión Vapor TG 4 } (=) [\%]$$

$$x_{10} = \text{Generación Central Santa Fe Energía } (=) [\text{MWh}]$$

Una vez definidas las variables se muestra la red factorial final obtenida, para la cual se consideraron dos factores comunes, que representan aproximadamente del 76% de la variabilidad del Específico de Generación como se explicará más adelante. En la *Figura 5.2.* se muestra la red factorial obtenida para las variables ya filtradas.

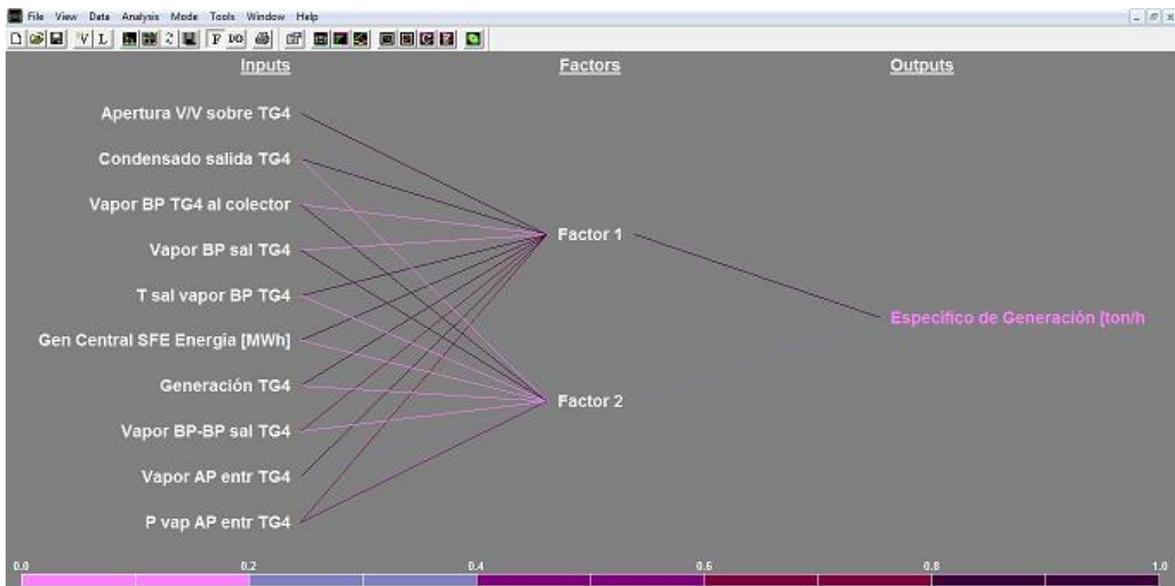


Figura 5.2. Red factorial final.

La Figura 5.2. muestra la red factorial final obtenida, en la cual se incluyen dos factores comunes, la variable de salida Específico de Generación, y las 10 variables de entrada. En la parte inferior pueden observarse distintos colores desde el más claro hasta el más oscuro, que indican el grado de correlación entre variables y factores. Dicho grado de correlación se ve reflejado y cuantificado por las ecuaciones de puntuación, saturación y ganancia.

5.2 Variabilidad Explicada.

La variabilidad explicada para la variable de salida, en este caso el Específico de Generación de Central Santa Fe Energía es de aproximadamente un 76%, esto sin contar la variabilidad explicada para la simulación de todo el resto de las variables. El porcentaje de variabilidad explicada para la variable se manifiesta en el grado de veracidad con que se predice el *output* o salida de la variable para el modelo calculado con respecto a los datos medidos y se construye a partir del banco de datos consistente en el 80% de los datos iniciales; 596 variables con 744 mediciones cada una. El resto de los datos como se mencionó anteriormente son para la validación del modelo.

En la Figura 5.3. se muestra el porcentaje de variabilidad explicado para la variable Específico de Generación por el modelo, el primer factor aporta un 68,8% de la variabilidad, mientras que el segundo factor aporta un 7,5%.

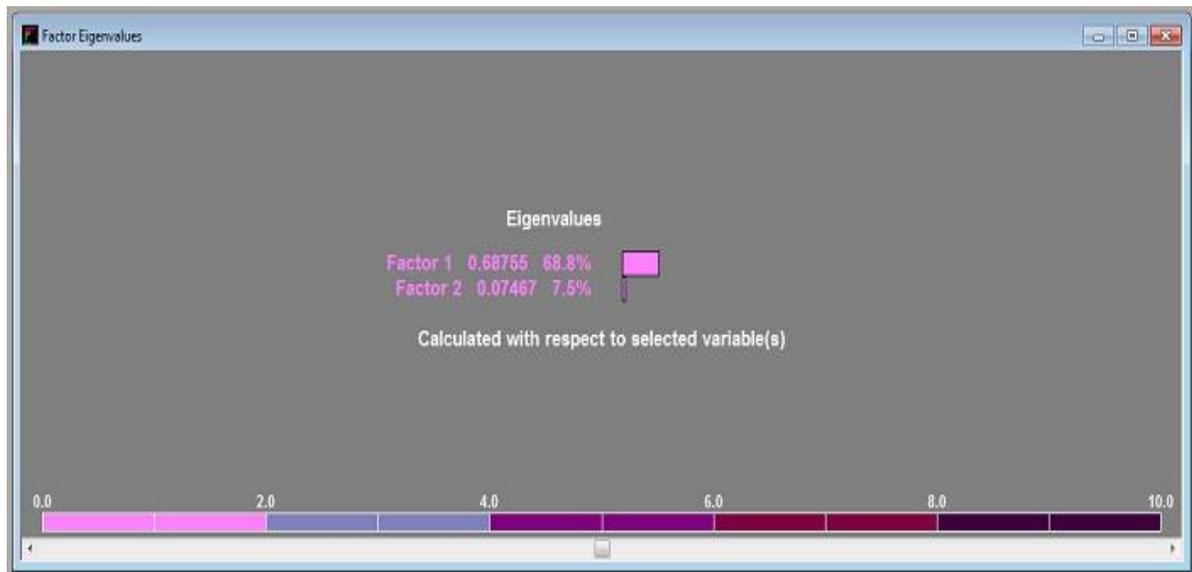


Figura 5.3. Valores propios para específico de generación Santa Fe Energía.

Lo anterior indica que la suma de los cuadrados de los coeficientes de correlación del primer factor es más alta que la misma suma para el segundo factor, por lo que el primer factor relaciona en mayor grado a las variables de entrada con la variable de salida que el segundo factor. Esto se observa en la coloración negra de las líneas que unen el Específico de Generación con el primer factor y este a su vez con las variables de entrada presentes en la Figura 5.2., color más oscuro que el que une el segundo factor con la variable de salida y las de entrada, lo que indica mayor grado de correlación entre factor y variables.

5.3 Correlación entre variables.

Las herramientas filtro factorial y análisis de correlación permiten encontrar todas las relaciones entre las variables de entrada y el Específico de Generación de Central Santa Fe Energía. En el primer caso, y tal como se precisó en sección 4.3.1., el filtro factorial permite listar las variables que más influyen sobre el Específico de Generación. Una vez obtenidas dichas variables es posible realizar un análisis de correlación, herramienta que entrega los coeficientes de correlación existentes para la combinatoria lograda entre todos los pares de variables posibles, para un total de 11 variables, aunque en este caso nos centraremos en las combinaciones (pares) logradas entre el Específico de Generación y las 10 variables de entrada. La Tabla 5.1. muestra los coeficientes de correlación obtenidos, en el mismo orden

de importancia en cuanto a incidencia sobre el Específico de Generación obtenido mediante el Filtro Factorial.

X_i	Variable de Entrada	E.E. Generación
X ₁	Generación TG4	-0,780
X ₂	Condensado de Salida TG4	-0,820
X ₃	Vapor BP-BP Salida TG4	-0,620
X ₄	Vapor BP Salida TG4	0,331
X ₅	Temperatura de Salida Vapor BP TG4	0,659
X ₆	Vapor BP TG4 al Colector	0,363
X ₇	Vapor de Entrada AP TG4	-0,660
X ₈	Presión Vapor AP Entrada TG4	0,507
X ₉	Apertura V/V Admisión Vapor TG4	-0,620
X ₁₀	Generación Central Energética	-0,780

Tabla 5.1. Correlación entre específico de generación y variables de entrada.

Los coeficientes de correlación se derivan a partir de las ecuaciones de ganancia, y permiten relacionar cada una de las variables con las restantes. En general valores entre 0.8-1.0 indican una relación fuerte entre variables, mientras que valores entre 0.0-0.2 indican una relación débil. También es posible conocer si dicha variable afecta negativa o positivamente a las restantes variables, es decir, si el incremento en una unidad de dicha variable generará un aumento en las otras variables o generará un decremento en términos absolutos según el signo del coeficiente de correlación. Lo anterior resulta útil cuando se quiere realizar análisis más detallados referentes a cómo incide un conjunto de variables de entrada sobre la variable de salida en estudio.

Entre las variables incluidas es posible apreciar que el coeficiente que relaciona el Vapor de Entrada de alta presión al Turbogenerador N°4 con el Específico de Generación posee un valor de -0,660, es decir, una ponderación relativamente alta que implica que al aumentar el flujo de entrada de vapor de alta presión el Específico de Generación de Central Santa Fe Energía debe disminuir, lo que implica a su vez que se deben emplear menos [ton/h] de vapor para obtener una generación de 1 [MWh] por parte del Turbogenerador N°4, lo que implica que el Turbogenerador aumenta su eficiencia a medida que aumenta el flujo de entrada de vapor de alta presión a éste.

Por el lado del Condensado de Salida del TG4 se obtiene el coeficiente con valor absoluto más alto, correspondiente a -0.820, lo que tiene sentido puesto que la mayor parte de la generación

energética se obtiene al condensar el vapor de alta presión que entra al Turbogenerador N°4, luego a mayor grado de condensación mayor cantidad de energía generará el Turbogenerador N°4 y más irá disminuyendo el Específico de Generación.

Continuando con la siguiente variable, para el caso del vapor de salida de baja presión se obtiene un coeficiente de 0.331. El sentido del signo implica que al aumentar la extracción de vapor de baja presión debe aumentar el Específico de Generación, aunque la magnitud indique que no se encuentra tan altamente correlacionado con la generación propiamente tal como el Condensado de Salida del TG4. Dicho análisis es cierto, puesto que si se aumenta la extracción de baja presión se dejará de emplear el vapor de alta presión para generación energética, y se empleará como maquila para Santa Fe, luego se disminuye a su vez al condensado de salida y por tanto el Específico aumenta, es decir, se emplean más [ton/h] de vapor para generar 1 [MWh] en la Central Santa Fe Energía.

En el caso del vapor de baja-baja presión de salida del Turbogenerador N°4 el coeficiente es -0.620. Primero es necesario decir que la magnitud es coherente, puesto que dicho vapor se emplea para calentar el agua desmineralizada que se retorna al estanque de agua de alimentación de la Caldera de Biomasa 2, luego dicha agua desmineralizada es empleada para generación de vapor dentro de la caldera producto del intercambio energético que ocurre al quemar la biomasa en el lecho fluidizado interior de esta. En resumen, la energía que entrega el vapor de baja-baja presión es empleada para precalentar el agua que será posteriormente convertida en vapor, si dicha agua entra a mayor temperatura será más fácil evaporarla, por tanto con la misma cantidad de energía aportada por la caldera se generará mayor cantidad de vapor y por tanto mayor aporte al vapor total de entrada de alta presión al Turbogenerador N°4. Si bien es necesario que haya una temperatura mínima de aporte del agua desmineralizada para generación de vapor en la caldera, resulta necesario recalcar que hasta el momento la variable que más aporta a disminuir el Específico de Generación es el Condensado de Salida del Turbogenerador N°4.

Otro punto que es importante recalcar del análisis es que en primera instancia es necesario maximizar el flujo de Condensado de Salida del TG4, mantener valores razonables de Vapor de baja-baja Presión de salida y minimizar dentro de lo posible (en cuanto lo permitan las condiciones operativas de Planta Santa Fe) la extracción de vapor de baja presión. Dichos

valores se obtendrán al optimizar el modelo para la minimización del Específico de Generación.

El análisis para las variables remanentes de entrada es similar, y en definitiva el Análisis de Correlación permite obtener un marco conceptual que permite al profesional, o ingeniero de la planta realizar análisis más profundos respecto a los efectos de distintas variables manipulables sobre la variable a optimizar. En la *sección 5.4.* se procederá a modelar matemáticamente obtenido y a presentar los resultados de la optimización realizada con el software LINGO.

5.4 Modelo Matemático.

El modelo matemático fue obtenido a través del proceso de análisis factorial implementado en el software FACTNET, mediante la herramienta de extracción factorial. El algoritmo de cálculo emplea un método híbrido consistente en un análisis de componentes principales (PCA) para la obtención de las ponderaciones factoriales a partir de las ecuaciones de saturación, y regresión para la obtención de los coeficientes de puntuación pertenecientes a las ecuaciones de puntuación. Una vez conocidas las ecuaciones de puntuación y de saturación se reemplazan las ecuaciones de puntuación en las de saturación y se forman las ecuaciones de ganancia con sus respectivos coeficientes denominados ganancias del modelo. Dicho lo anterior, el modelo matemático desarrollado se muestra en la *Figura 5.4.*

La función objetivo se encuentra compuesta por las variables 10 variables de entrada (x_i), lo que implica que abarca toda la red factorial, con los factores comunes implícitos en la obtención de dicha ecuación en el reemplazo de las ecuaciones de puntuación sobre las de saturación. Las restricciones para cada variable x_i , corresponden a valores o cotas operacionales, mientras que el valor para el específico es el calculado a partir de la división entre el vapor de alta presión de entrada al turbogenerador N°4 y la generación neta de central santa fe energía.

$$\begin{aligned} \text{Min } y = & -0.0083 \cdot x_1 + 44.6062 \cdot x_2 + 44.6013 \cdot x_3 + 44.6093 \cdot x_4 + 0.0036 \cdot x_5 + 0.0029 \cdot x_6 \\ & -44.6041 \cdot x_7 + 0.0400 \cdot x_8 - 0.0007 \cdot x_9 - 0.0171 \cdot x_{10} + 1.6746 \end{aligned}$$

s.a.

$$\begin{aligned} 3.80 & \leq y \leq 5.50 \\ 17.28 & \leq x_1 \leq 89.82 \\ 96.23 & \leq x_2 \leq 290.3 \\ 11.10 & \leq x_3 \leq 41.20 \\ 0 & \leq x_4 \leq 86.82 \\ 156.1 & \leq x_5 \leq 194.3 \\ -20.3 & \leq x_6 \leq 60.28 \\ 108.3 & \leq x_7 \leq 330.9 \\ 79.4 & \leq x_8 \leq 82.7 \\ 15.78 & \leq x_9 \leq 99.30 \\ 12.02 & \leq x_{10} \leq 83.39 \end{aligned}$$

$$\begin{aligned} x_1 = & 0.2818 \cdot x_1 - 1169.6505 \cdot x_2 - 1169.4336 \cdot x_3 - 1169.6556 \cdot x_4 - 0.1304 \cdot x_5 + 0.0124 \cdot x_6 \\ & + 1169.6329 \cdot x_7 - 3.2247 \cdot x_8 + 0.0141 \cdot x_9 + 0.5135 \cdot x_{10} + 296.9430 \\ x_2 = & 0.8585 \cdot x_1 - 4210.2449 \cdot x_2 - 4209.6827 \cdot x_3 - 4210.4536 \cdot x_4 - 0.3825 \cdot x_5 - 0.1753 \cdot x_6 \\ & + 4210.0874 \cdot x_7 - 6.2121 \cdot x_8 + 0.0624 \cdot x_9 + 1.6893 \cdot x_{10} + 640.8699 \\ x_3 = & 0.0940 \cdot x_1 - 411.3170 \cdot x_2 - 411.2479 \cdot x_3 - 411.3251 \cdot x_4 - 0.0430 \cdot x_5 - 0.0029 \cdot x_6 \\ & + 411.3078 \cdot x_7 - 0.9568 \cdot x_8 + 0.0053 \cdot x_9 + 0.1754 \cdot x_{10} + 96.3284 \\ x_4 = & -0.1865 \cdot x_1 + 2109.4844 \cdot x_2 + 2109.5445 \cdot x_3 + 2109.8863 \cdot x_4 + 0.0560 \cdot x_5 + 0.4317 \cdot x_6 \\ & - 2109.2583 \cdot x_7 - 5.3200 \cdot x_8 - 0.0496 \cdot x_9 - 0.5976 \cdot x_{10} + 423.5556 \\ x_5 = & -0.1803 \cdot x_1 + 867.8981 \cdot x_2 + 867.7776 \cdot x_3 + 867.9371 \cdot x_4 + 0.0807 \cdot x_5 + 0.0315 \cdot x_6 \\ & - 867.8676 \cdot x_7 + 1.3950 \cdot x_8 - 0.0126 \cdot x_9 - 0.3516 \cdot x_{10} + 65.8030 \\ x_6 = & -0.1913 \cdot x_1 + 2066.8231 \cdot x_2 + 2066.8700 \cdot x_3 + 2067.2064 \cdot x_4 + 0.0597 \cdot x_5 + 0.4109 \cdot x_6 \\ & - 2066.6068 \cdot x_7 - 4.9156 \cdot x_8 - 0.0480 \cdot x_9 - 0.5943 \cdot x_{10} + 375.8059 \\ x_7 = & 0.4684 \cdot x_1 - 2573.0498 \cdot x_2 - 2572.7852 \cdot x_3 - 2573.2461 \cdot x_4 - 0.2024 \cdot x_5 - 0.1866 \cdot x_6 \\ & + 2572.9195 \cdot x_7 - 1.8469 \cdot x_8 + 0.0424 \cdot x_9 + 0.9749 \cdot x_{10} + 171.4211 \\ x_8 = & 0.2733 \cdot x_1 - 1138.7505 \cdot x_2 - 1138.5409 \cdot x_3 - 1138.7569 \cdot x_4 - 0.1263 \cdot x_5 + 0.0105 \cdot x_6 \\ & + 1138.7327 \cdot x_7 - 3.1012 \cdot x_8 + 0.0138 \cdot x_9 + 0.4988 \cdot x_{10} + 281.6391; \end{aligned}$$

Figura 5.4. Modelo Matemático.

Para el modelo se consideran ocho ecuaciones de ganancia, suficientes para describir correctamente el problema como se aprecia al observar los valores obtenidos en la optimización. La programación y resolución del modelo se llevó a cabo con el *software* LINGO (ver Anexo). El modelo resuelto es de carácter lineal y posee en total 11 variables y 32 restricciones. La linealidad del modelo permite que el cálculo de sus salidas u *outputs* sean rápidos. Lo que en realidad toma la mayor parte del tiempo es la recopilación del banco de

datos, tratamiento del banco de datos, adecuación de la red factorial, y traspasar el código obtenido al software LINGO.

5.5 Minimización del Modelo Obtenido.

Una vez programado el modelo predictivo es posible minimizarlo, esto se realiza con el *software* LINGO, seleccionando la opción *LINGO/Solve*. Se consideran dos casos para la solución del modelo, de tal forma de minimizar el Específico de Generación de Central Santa Fe Energía. El primer caso considera todas las variables de entrada positivas para la solución del modelo, esto implica que Santa Fe recibirá parte del vapor de baja presión que sale desde el Turbogenerador N°4 de Santa Fe Energía, mientras que el remanente se utilizará para el soplado de la Caldera de Biomasa N°2. El segundo caso considerará que Santa Fe Energía recibirá vapor de baja presión desde el colector de baja presión de Santa Fe.

- **Caso N°1.**

El primer caso es la situación más recurrente dentro de la operación de la planta, puesto que por lo general existe demanda en Santa Fe de vapor de baja presión –almacenado en el colector de baja presión de Santa Fe 2 y en el colector de baja presión de Santa Fe 1- por parte de equipos de proceso y procesos en sí. El que exista demanda implicará que el valor que adoptará la variable x_6 , *Vapor BP TG4 al Colector*, será mayor o igual a cero, es decir, que habrá consumo por parte de Santa Fe de vapor de baja presión como mínimo para autoabastecerse, de tal forma que el valor cero significará que no es necesario pedirle vapor como insumo a Santa Fe Energía, mientras que un valor mayor a cero implicará que Santa Fe consume vapor de baja presión desde Santa Fe Energía, lo que desde el punto de vista de Santa Fe Energía no es provechoso, ya que lo conveniente para la Central es emplear una alta cantidad de vapor para condensación y generación energética.

La solución que entrega LINGO para la minimización del Específico de Generación considerando los puntos anteriores y luego de 9 iteraciones se muestra en la *Figura 5.5*.

```

Global optimal solution found.
Objective value:                4.037725
Infeasibilities:                0.000000
Total solver iterations:        9

```

Variable	Value	Reduced Cost
X1	78.81818	0.000000
X2	242.7041	0.000000
X3	32.35786	0.000000
X4	18.27456	0.000000
X5	156.1000	0.000000
X6	0.000000	0.3801730E-02
X7	293.3365	0.000000
X8	80.53971	0.000000
X9	67.91305	0.000000
X10	72.03238	0.000000
Y	4.037725	0.000000

Figura 5.5. Solución para minimización del modelo/Caso 1.

En la *Tabla 5.2.* se muestran valores pertenecientes al Banco de Datos con el que se construyó el modelo, a partir de mediciones reales, y se comparan con los valores obtenidos para las variables de entrada y variable de salida a partir de la optimización del modelo predictivo.

Sistema de Distribución de Vapor											
Fecha	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Y
17-ene-14 01:00:00	87,0684	269,4023	31,4940	20,9486	158,6615	4,6822	321,8449	79,9059	96,2616	79,8659	4,0298
18-ene-14 21:00:00	80,7945	251,7288	29,7323	15,5739	157,5351	-1,8242	297,0351	80,5112	78,9530	73,7215	4,0291
27-ene-14 17:00:00	81,9004	250,2236	29,8429	24,8647	157,5129	10,3501	304,9311	80,1567	78,3112	75,4818	4,0398
28-ene-14 02:00:00	83,1049	251,6281	34,3084	22,0031	157,5661	4,1000	307,9396	80,2872	81,6946	76,2564	4,0382
28-ene-14 10:00:00	80,0501	245,9672	35,1917	14,7052	157,5332	-3,3519	295,8641	80,5324	77,7089	73,5490	4,0227
29-ene-14 21:00:00	71,3797	227,1733	26,8056	9,4905	157,2840	-5,5885	263,4694	80,8267	52,3830	65,2462	4,0381
31-ene-14 14:00:00	78,3505	240,8886	32,7750	15,4860	157,3314	-2,0883	289,1496	80,7072	70,2537	71,6442	4,0359
Valor Promedio	80,3784	248,1445	31,4500	17,5817	157,6320	0,8971	297,1763	80,4182	76,5094	73,6807	4,0334
Valor Calculado por Modelo Predictivo	78,8182	242,7041	32,3579	18,2746	156,1000	0,0000	293,3365	80,5397	67,9131	72,0324	4,0377

Tabla 5.2. Valores calculados por modelo predictivo y valores del banco de datos, caso 1.

El valor obtenido al minimizar la variable de salida Específico de Generación de Central Santa Fe Energía es de 4,0377 [ton/h/MWh], lo que implica que se requerirán 4,0377 toneladas de vapor de alta presión por hora para generar 1 [MW]. Lo anterior se logra definiendo las 10 variables de entrada en los valores calculados por el modelo, como se observa en la *Tabla 5.2.*

Los valores calculados por el modelo predictivo se acercan bastante a los valores operacionales en que se logró un Específico de Generación semejante y más aún al valor promedio de aquellos días, como también se observa en la *Tabla 5.2*. En base a lo anterior, el modelo en efecto logra predecir valores tanto para la variable de salida como para las variables de entrada razonablemente cercanos a lo real, tomando en cuenta el punto extra de que la variable de salida es calculada a partir de las de entrada, efecto que debiese aumentar la desviación del modelo, y que aún siendo parte de éste, no lleva a predicciones erróneas.

Es necesario recordar que el fin de la minimización es hacer más eficiente la generación de Central Santa Fe Energía, lo que implica que los valores calculados por el modelo permitirán trabajar en condiciones cercanas a lo óptimo.

- **Caso N°2.**

El segundo caso no es una situación recurrente dentro de la operación de la planta, no obstante, si se da la ocasión debe ser aprovechada. Este caso implica que no existe demanda por parte de Santa Fe de vapor de baja presión, y que inclusive puede enviar vapor de baja presión desde el colector de baja presión de Santa Fe 2 hacia Santa Fe Energía. Lo anterior implica que para la solución de la minimización en LINGO se admitirán valores negativos para las variables de entrada, más concretamente para la variable x_6 , *Vapor BP TG4 al Colector*. El signo negativo no es más que la definición de la medición de la variable en DCS (Sistema de Control Distribuido); el signo positivo indica dirección de flujo normal de vapor de baja presión desde Santa Fe Energía a Santa Fe, mientras que el signo negativo indica flujo desde Santa Fe hacia Santa Fe Energía.

Un menor consumo de vapor de baja presión por parte de Santa Fe puede darse por una detención o disminución de la velocidad de manufactura de hojas de celulosa por parte de la “máquina” de la línea 2, encargada de la formación de la hoja de celulosa y su posterior corte final, o de otros equipos con consumos considerables como el digester. En dichas situaciones, y de ser posible resulta provechoso obtener vapor de baja presión por parte de Santa Fe hacia Santa Fe Energía. Situación similar puede ocurrir también para Santa Fe 1, y dado que los colectores de baja presión se encuentran interconectados es probable que el vapor de baja presión sobrante en Santa Fe 1 sea enviado a Santa Fe 2 por interconexión entre colectores, y este a su vez debiese ser enviado hacia Santa Fe Energía.

La solución que entrega LINGO para la minimización del Específico de Generación considerando los puntos anteriores y luego de 9 iteraciones se muestra en la *Figura 5.6*.

Global optimal solution found.			
Objective value:		3.972024	
Infeasibilities:		0.000000	
Total solver iterations:		9	
	Variable	Value	Reduced Cost
	X1	75.99528	0.000000
	X2	243.8630	0.000000
	X3	31.73678	0.000000
	X4	0.000000	0.000000
	X5	156.1000	0.000000
	X6	-17.28203	0.000000
	X7	275.5998	0.000000
	X8	80.78347	0.000000
	X9	72.71187	0.000000
	X10	69.36591	0.000000
	Y	3.972024	0.000000

Figura 5.6. Solución para minimización del modelo/Caso 2.

En la *Tabla 5.3*. se muestran valores pertenecientes al Banco de Datos con el que se construyó el modelo, a partir de mediciones reales, y se comparan con los valores obtenidos para las variables de entrada y variable de salida a partir de la optimización del modelo predictivo.

Sistema de Distribución de Vapor											
Fecha	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Y
11-ene-14 06:00:00	85,9864	271,1987	32,1074	11,4576	158,5135	-5,1513	314,7637	80,0772	98,6688	79,0965	3,9795
18-ene-14 19:00:00	78,4948	247,9403	34,0259	3,2303	158,1311	-15,3516	285,1966	80,6920	82,7062	71,6697	3,9793
18-ene-14 22:00:00	80,7362	252,7992	30,9549	9,3368	158,0339	-8,1326	293,0909	80,5849	85,8487	73,8614	3,9681
28-ene-14 09:00:00	81,7876	255,6078	31,7799	12,0700	157,3018	-5,2149	299,4577	80,4672	83,6441	75,3695	3,9732
Valor promedio	81,7513	256,8865	32,2170	9,0237	157,9951	-8,4626	298,1272	80,4553	87,7169	74,9993	3,9750
Valor Calculado por Modelo Predictivo	75,9953	243,8630	31,7368	0,0000	156,1000	-17,2820	275,5998	80,7835	72,7119	69,3659	3,9720

Tabla 5.3. Valores calculados por modelo predictivo y valores del banco de datos, caso 2.

En este caso el valor obtenido para el Específico de Generación es menor, como era de esperar, y corresponde a 3,9720 [ton/h/MWh], es decir, Central Santa Fe Energía requerirá 3,9720 toneladas de vapor de alta presión por hora para generar 1 [MW]. Si bien los valores predichos de las variables de entrada no son tan cercanos a los valores reales de operación como para el Caso N°1, sí lo es el Específico de Generación, luego dado que las

combinaciones posibles en las variables de entrada para predecir la variable de salida son múltiples, no es posible descartar a priori ninguna combinación hasta que esta sea probada, ya que como se ha mencionado previamente la interrelación entre las variables que componen el Sistema de Distribución de Vapor es alta.

5.6 Capacidad de Predicción y Ajuste del Modelo Obtenido.

Se analizan diversos aspectos referentes al grado de exactitud del modelo obtenido, tales como; simulación de la variable Específico de Generación, coeficientes de determinación (R^2) de las variables calculadas con respecto a las variables medidas, y grado de ajuste del modelo.

5.6.1 Simulación de la variable Específico de Generación.

En la *Figura 5.7.* se muestran los valores empleados desde la base de datos para la construcción del modelo junto con los valores predichos por el modelo de red factorial, mostrando gráficamente qué tan bien predice el modelo con respecto a los datos reales.

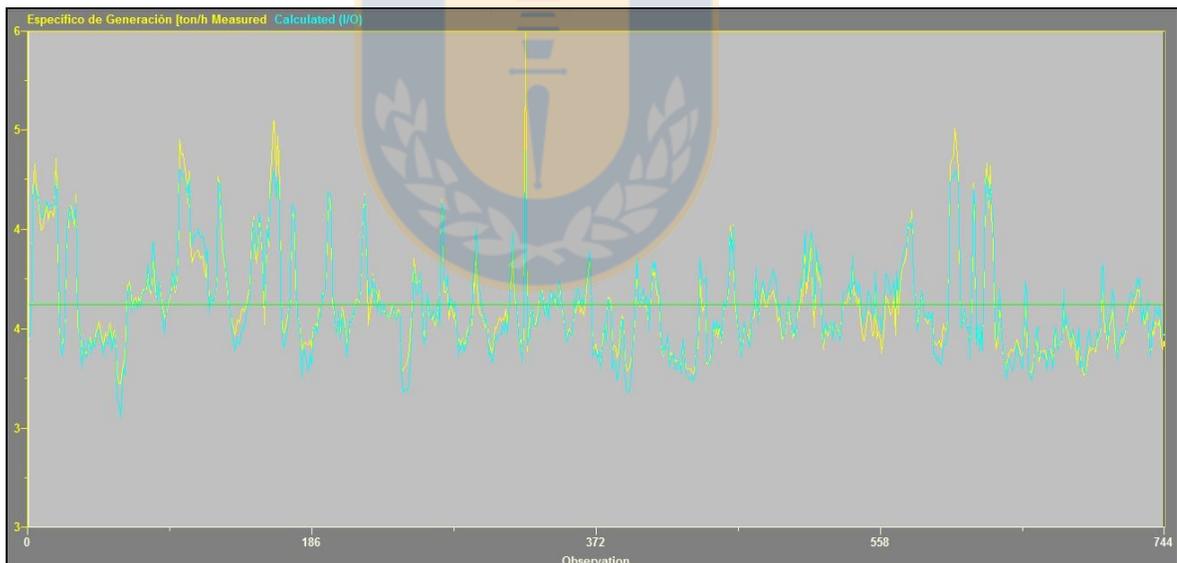


Figura 5.7. Modelo predictivo vs datos de construcción del modelo.

En la *Figura 5.7.* se observa el grado de predicción de la variable Específico de Generación con respecto a los valores graficados del banco de datos considerando todo el mes de enero de 2014. Esta herramienta visual resulta útil cuando se está comprobando el grado de ajuste de un modelo, ya que permite comprobar en forma preliminar si el modelo es aceptable o no.

En el caso de la variable Específico de Generación se observa que el grado de ajuste del modelo es bastante elevado, puesto que las líneas azules, correspondientes al modelo predicho o calculado, se superponen en muchos casos a las líneas amarillas correspondientes a la gráfica del mes de enero para la variable, lo que implica un buen grado de predicción. Por otro lado, se observa que para valores medidos altos de Específico de Generación los valores calculados por el modelo no escapan a valores extremadamente altos como en la realidad, de todas formas dicho punto no es preocupante, ya que lo conveniente desde el punto de vista de los procesos es mantener la variable Específico de Generación en su mínimo valor, ya que permite mayor generación energética con menor consumo de vapor, como se ha mencionado antes.

5.6.2 Coeficiente de Determinación de las Variables Calculadas.

Walpole (2007) denomina a R^2 como el coeficiente de determinación y lo califica como una “medida de la proporción de la variabilidad explicada por el modelo ajustado”, matemáticamente se interpreta como; “entre más cercano a 1 sea el coeficiente de variabilidad expresado al comparar las variables más variación será explicada”.

Para el caso de la variable Específico de Generación la proporción de la variabilidad explicada por el modelo predictivo es de un 78,3%, el margen de error se encuentra distribuido en su mayor parte para valores de Específico más altos, mientras que la mayor exactitud de predicción del modelo se encuentra para valores de Específico más bajos, lo que no incide en la minimización, pues se trabaja en rangos de valores bajos. Cabe recordar que el modelo calcula cada variable de entrada y en base a los cálculos de dichas variables de entrada calcula la variable predictiva Específico de Generación.

En la *Figura 5.8.* se observa la gráfica de la variable Específico de Generación medido vs Específico de Generación calculado por el modelo predictivo.

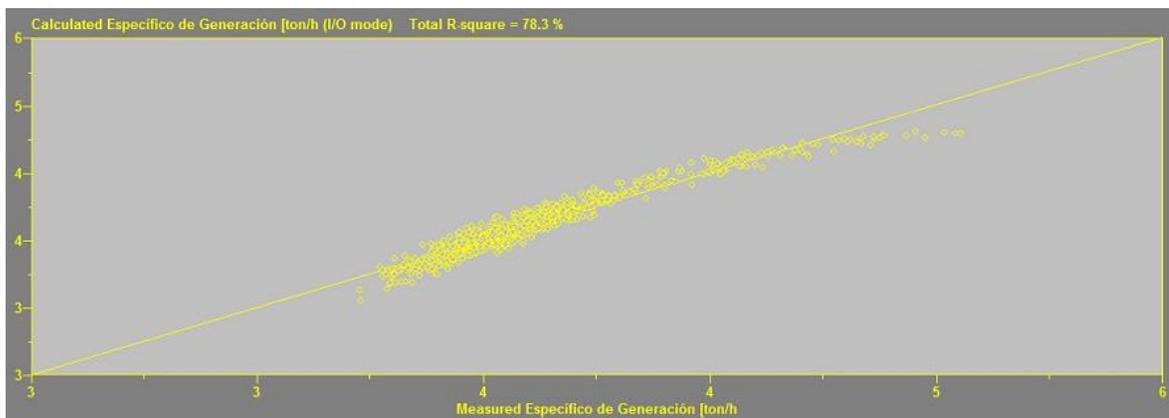


Figura 5.8. Específico de Generación medido vs calculado por modelo predictivo.

En la *Tabla 5.4* se listan los coeficientes de determinación para todas las variables que componen el modelo predictivo.

Coeficientes de Determinación	
Variable	R^2
Específico de Generación	78,3
Generación TG4	99,1
Generación Central de Energía	99,0
Vapor BP TG4 al Colector	98,9
Vapor AP Entrada TG4	98,9
Vapor BP Salida TG4	98,4
Condensado de Salida TG4	96,2
Apertura V/V Admisión Vapor TG4	82,4
Presión Vapor AP Entrada TG4	75,9
Temperatura de Salida Vapor BP TG4	71,5
Vapor BP-BP Salida TG4	62,4

Tabla 5.4. Coef. de determinación para específico de generación y variables de entrada.

En general se observan valores sobre el 70% para los coeficientes de determinación de todas las variables (entrada y salida), con 6 variables sobre el 98% de variabilidad explicada, por lo que el modelo en términos globales y particulares es bueno ajustando los valores medidos (reales).

5.6.3 Grado de Ajuste del Modelo.

El grado de ajuste general del modelo de red factorial que se obtiene se muestra en la *Figura 5.9*. El error (falta de ajuste) representado por FACTNET está dado por “la suma de los cuadrados de los errores del modelo divididos por sus errores estándar”.



Figura 5.9. Gráfico de Falta de Ajuste.

Como se observa en la *Figura N°5.9*, la totalidad de los puntos máximos de falta de ajuste se encuentran bajo el valor 2.0 y la media se observa alrededor de 0,6. El manual *Factnet Using Pacific Simulation to Understand your Multivariate Process Data* (Pacific Simulation, 2005) plantea que la falta de ajuste normalmente varía entre 0.5 y 1.5 , como ocurre en el presente estudio, y que si esta sobrepasa en promedio el valor 3.0 el modelo debe considerarse inválido, situación lejana para este estudio. Luego, basado en los puntos anteriores puede plantearse que el grado global de ajuste del modelo resulta aceptable.

En la presente sección se abordaron los resultados del estudio, referentes a la obtención de una red factorial inicial a partir del banco de datos de las variables para el mes de enero de 2014, acotada a través de las herramientas de Filtro Factorial y Pre-clasificación a una red factorial final que permitió la obtención de un modelo matemático para la variable Específico de Generación. Cabe destacar que la variabilidad explicada para la variable Específico de Generación con dos factores comunes, fue de un 76%. Una vez obtenido el modelo se procedió a la minimización de la variable de salida Específico de Generación. Los valores

obtenidos de la minimización fueron un Específico de Generación de 4,0377 [ton/h/MWh] para el Caso N°1 y 3,9720 [ton/h/MWh] para el Caso N°2, ambos casos se diferenciaron por no recibir vapor de baja presión desde Santa Fe a Santa Fe Energía y sí recibir, respectivamente, y corresponden a condiciones operacionales posibles dentro de la dinámica del proceso de CMPC Planta Santa Fe. Finalmente se analizó la capacidad de predicción y ajuste del modelo obteniendo un coeficiente de determinación, R^2 , de 78.3 % para la variable predictiva Específico de Generación y una falta de ajuste del modelo global menor a 2%, con una media aproximada de 0,6%.



Conclusiones y Recomendaciones.

En el presente estudio se simuló el Sistema de Distribución de Vapor de CMPC Celulosa Planta Santa Fe. La finalidad fue desarrollar un modelo matemático que permitiese expresar la variable de salida Específico de Generación de Central Santa Fe Energía en función de las variables de entrada controlables del proceso, de tal forma de poder minimizar su valor, y así poder maximizar la eficiencia de generación eléctrica de la Central, minimizando las toneladas de vapor hora necesarias para generar 1 [MW] en Central Santa Fe Energía, para tasas elevadas de alimentación de vapor de alta presión al Turbogenerador N°4.

La minimización del Específico de Generación se logró, mediante el modelo matemático, y permitió obtener los *outputs* de todas las variables de entrada que permiten dicha minimización, lo que implica definir nuevos puntos operacionales para dichas variables que permiten operar a altos grados de eficiencia energética para Central Santa Fe Energía, y conocimiento nuevo respecto al comportamiento del Sistema de Distribución de Vapor frente a dicho escenario.

Además del desarrollo de un modelo matemático también se obtuvieron las correlaciones para las distintas variables que componen el modelo, centrando el análisis en las opciones logradas al combinar las 10 variables de entrada con la variable de salida. Las correlaciones permitieron comprender y analizar cómo y en qué grado afectan las variables de entrada a la variable Específico de Generación. Este tipo de análisis reporta un grado de información interesante a los profesionales dedicados al seguimiento del proceso, puesto que permite analizar las relaciones cualitativas y cuantitativas entre distintas variables, otorgando nuevo conocimiento o reafirmación de ideas previas.

Dicho lo anterior, la minimización del modelo, y por tanto de la variable Específico de Generación de Central Santa Fe Energía, se llevó exitosamente a cabo para dos casos: el Caso N°1 considera la extracción de baja presión desde el Turbogenerador N°4 hacia el colector de baja presión de Santa Fe 2, mientras que el Caso N°2 no consideraba extracción y sí consideraba el envío de vapor desde el colector de baja presión desde Santa Fe a Santa Fe

Energía, ambos casos factibles, es decir, posibles de llevar a cabo en la realidad operacional cotidiana de la planta.

El modelo obtenido es de carácter predictivo, lo que implica que las 10 variables de entrada calculadas se emplean para predecir el valor de la variable de salida Específico de Generación con un grado de ajuste obtenido y exactitud de predicción más que aceptables, alta variabilidad explicada por los factores comunes, y un ajuste general alto del modelo, considerando todas las variables.

Como recomendación para mejorar el modelo predictivo se debiese considerar agregar la variable *% de humedad* de la biomasa alimentada a la Caldera de Biomasa 2, ya que esta incide directamente en la capacidad de generación energética de la caldera, debido a que entre mayor humedad menor será el poder calorífico inferior, y por tanto menor será la generación de vapor de alta presión por parte de ésta. Esta variable responde a la estacionalidad, lo que implicaría un grado de representación de la realidad aún más elevado. No obstante, no se incluye en el modelo propuesto en el presente estudio, puesto que no se observarían grandes variaciones del % de humedad de la biomasa alimentada al analizar solo un mes, ya que la humedad que contienen los distintos tipos de especies que conforman el *mix* alimentando a la caldera varían a lo largo de todo el año. Luego, para realizar el análisis anterior debe generarse un nuevo modelo que considere a lo menos un año como base para el banco de datos.

Además se recomienda implementar un sistema en línea que permita alimentar al modelo con un banco de datos mayor, con el fin de que en el largo plazo la capacidad predictiva de éste fuese aumentando. Como última recomendación resultaría útil abordar el Sistema de Distribución de Vapor por subsistemas, es decir, atacar distintas áreas como Santa Fe 1 y Santa Fe 2, y sus respectivas calderas, turbinas, acumulador y colectores, para en última instancia unir dichos modelos y generar un sistema global que interactúe en línea, semejante a un sistema de decisión o sistema experto.

Bibliografía

- Contreras, L., 2010. Optimización de las Propiedades Físico-Mecánicas del Papel Sackraft en la Máquina Papelera N°12 de CMPC Celulosa Planta Laja. ATCP Rev. Celul. Pap. 25.
- Edwards, P.J., Murray, A., Papadopoulos, G., Wallace, A., Barnard, J., Smith, G., 1999. The application of neural networks to the papermaking industry. *Neural Netw. IEEE Trans.* On 10, 1456–1464.
- EPA, 2013. Catalog of CHP Technologies.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM* 39, 27–34.
- Frawley, W.J., Piatetsky-Shapiro, G., Matheus, C.J., 1992. Knowledge discovery in databases: An overview. *AI Mag.* 13, 57.
- Hair, J.F., 2010. *Multivariate Data Analysis*. Prentice Hall.
- Jeong, H., Cho, S., Kim, D., Pyun, H., Ha, D., Han, C., Kang, M., Jeong, M., Lee, S., 2012. A heuristic method of variable selection based on principal component analysis and factor analysis for monitoring in a 300 kW MCFC power plant. *Hydrog. Enriched Methane* 37, 11394–11400.
- Johnson, R.A., Wichern, D.W., Education, P., 2014. *Applied Multivariate Statistical Analysis: Pearson New International Edition*. Pearson Education, Limited.
- Jolliffe, I., 2005. *Principal component analysis*. Wiley Online Library.
- Kélouwani, S., Agbossou, K., 2004. Nonlinear model identification of wind turbine with a neural network. *Energy Convers. IEEE Trans.* On 19, 607–612.
- Kusiak, A., Zhe Song, 2006. Combustion efficiency optimization and virtual testing: a data-mining approach. *Ind. Inform. IEEE Trans.* On 2, 176–184.
- Kusiak, A., Zheng, H., Song, Z., 2010. Power optimization of wind turbines with data mining and evolutionary computation. *Renew. Energy* 35, 695–702.
- Li, J., Liu, J., Gu, J., Niu, C., 2007. Study of operation optimization based on data mining technique in power plants. *Front. Energy Power Eng. China* 1, 457–462.
- Li, J., Wang, S., Niu, C., Liu, J., 2008. Research and application of data mining technique in power plant, in: *Computational Intelligence and Design, 2008. ISCID'08. International Symposium on.* IEEE, pp. 250–253.
- Mankiw, N.G., 2007. *Principios de economía/ Principles Of Economics*. Cengage Learning Latin America.

- Monedero, I., Biscarri, F., León, C., I. Guerrero, J., González, R., Pérez-Lombard, L., 2012. Decision system based on neural networks to optimize the energy efficiency of a petrochemical plant. *Expert Syst. Appl.* 39, 9860–9867.
- Motlaghi, S., Jalali, F., Ahmadabadi, M.N., 2008. An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework. *Expert Syst. Appl.* 35, 1540–1545.
- Pacific Simulation, 2005. Using Pacific Simulation's Factnet to Understand Your Multivariate Process Data.
- Platon, R., Ouloud Amazouz, M., 2007. Application of Data Mining Techniques for Industrial Process Optimization. *CANMET Energy Technol. Centre–Varenes*.
- Quintana, J., 2004. Modelación de Gases TRS de una Caldera Recuperadora de Reactivos. *ATCP Rev. Celul. Pap.*
- Souza, W.E., Aoki, A., Neto, A., 2013. Factor analysis in data mining applied for recognition and classification pattern for smart grid. *Innov. Smart Grid Technol. Lat. Am. ISGT 2013 IEEE PES Conf.* On 1–8.
- Spearman, C., 1904. "General Intelligence," Objectively Determined and Measured. *Am. J. Psychol.* 15, 201–292.
- Velázquez, D., González-Falcón, R., Pérez-Lombard, L., Marina Gallego, L., Monedero, I., Biscarri, F., 2013. Development of an energy management system for a naphtha reforming plant: A data mining approach. *Energy Convers. Manag.* 67, 217–225.
- Walpole, R.E., 2007. *Probabilidad y estadística para ingeniería y ciencias*. Pearson Educación.
- Wang, N., Zhang, Y., Zhang, T., Yang, Y., 2012. Data Mining-based Operation Optimization of Large Coal-fired Power Plants. *AASRI Procedia* 3, 607–612.
- Woodruff, E., Lammers, H., Lammers, T., 2004. *Steam Plant Operation*. Mcgraw-hill.
- Wu, F., Zhou, H., Ren, T., Zheng, L., Cen, K., 2009. Combining support vector regression and cellular genetic algorithm for multi-objective optimization of coal-fired utility boilers. *Fuel* 88, 1864–1870.
- Zheng, L., Yu, M., 2008. Monitoring NO_x emissions from coal fired boilers using generalized regression neural network, in: *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on.* IEEE, pp. 1916–1919.
- Zheng, L., Zhou, H., Wang, C., Cen, K., 2008. Combining Support Vector Regression and Ant Colony Optimization to Reduce NO_x Emissions in Coal-Fired Utility Boilers. *Energy Fuels* 22, 1034–1040.

Zhou, H., Pei Zhao, J., Gang Zheng, L., Lin Wang, C., Fa Cen, K., 2012. Modeling NO_x emissions from coal-fired utility boilers using support vector regression with ant colony optimization. *Eng. Appl. Artif. Intell.* 25, 147–158.

Zhou, H., Zheng, L., Cen, K., 2010. Computational intelligence approach for NO_x emissions minimization in a coal-fired utility boiler. *Energy Convers. Manag.* 51, 580–586.

www.centralenergía.cl/actores/autoridad-energía-chile/.

www.cdec-sic.cl/sobre-cdec-sic/.

www.osisoft.com



Anexos

Anexo 1: Banco de Datos Filtrado

A continuación se incluye el banco de datos filtrado del modelo, por problemas de espacio no se incluye el banco de datos total.*

Fecha	Generación TG4 [MWh]	Condensado salida TG4 [ton/h]	Vapor BP- BP sal TG4 [ton/h]	Vapor BP sal TG4 [ton/h]	T sal vapor BP TG4 [°C]	Vapor BP TG4 al colector [ton/h]	Vapor AP entr TG4 [ton/h]	P vap AP entr TG4 [bar]	Apertura V/V sobre TG4 [%]	Gen Central SFE Energía [MWh]	Específico de Generación [ton/h/MWh]
01-ene-14 00:00:00	76,01	237,81	26,74	20,80	157,31	3,51	285,34	80,29	60,65	69,23	4,12
01-ene-14 01:00:00	76,62	232,71	29,70	29,62	157,76	11,78	292,02	80,22	54,96	69,73	4,19
01-ene-14 02:00:00	68,62	203,43	29,43	39,85	159,74	21,45	272,71	80,29	40,36	61,83	4,41
01-ene-14 03:00:00	48,09	152,16	19,69	41,28	184,05	27,73	213,14	80,88	20,84	42,29	5,04
01-ene-14 04:00:00	45,28	145,71	18,91	40,59	188,06	27,37	205,22	80,89	19,01	39,47	5,20
01-ene-14 05:00:00	47,27	149,75	22,14	38,08	186,56	24,26	209,97	80,77	21,11	41,63	5,04
01-ene-14 06:00:00	43,86	151,73	15,79	25,83	188,10	12,91	193,34	81,02	19,52	38,37	5,04
01-ene-14 07:00:00	45,76	154,42	17,43	27,98	190,41	15,60	199,83	80,96	20,86	40,35	4,95
01-ene-14 08:00:00	47,61	160,59	20,04	22,16	186,95	9,74	202,79	80,92	23,63	42,24	4,80
01-ene-14 09:00:00	48,87	163,21	19,17	25,43	185,17	12,85	207,82	80,80	23,91	43,18	4,81
01-ene-14 10:00:00	47,71	160,42	18,26	25,90	187,25	13,74	204,58	80,89	23,20	42,14	4,86
01-ene-14 11:00:00	46,25	157,64	17,57	24,73	185,60	13,53	199,94	80,80	22,17	40,64	4,92
01-ene-14 12:00:00	45,71	154,79	17,11	27,62	189,79	16,17	199,52	80,80	21,31	40,14	4,97
01-ene-14 13:00:00	45,78	157,70	17,50	21,56	187,28	10,67	196,76	80,81	22,21	40,48	4,86
01-ene-14 14:00:00	45,30	155,68	17,64	22,94	189,13	11,92	196,26	80,84	21,46	39,96	4,91
01-ene-14 15:00:00	46,50	156,77	17,58	27,57	187,87	16,40	201,92	80,73	21,80	41,10	4,91
01-ene-14 16:00:00	46,72	157,00	18,10	27,40	187,95	16,55	202,50	80,73	22,12	41,47	4,88
01-ene-14 17:00:00	46,81	155,71	18,05	30,99	186,17	20,19	204,75	80,62	21,52	41,54	4,93
01-ene-14 18:00:00	43,17	143,09	17,57	37,17	190,73	26,08	197,84	80,80	18,01	37,78	5,24

01-ene-14 19:00:00	46,55	157,99	18,84	22,89	183,00	12,37	199,72	80,69	23,08	41,25	4,84
01-ene-14 20:00:00	61,52	196,97	22,70	23,46	175,99	10,74	243,13	80,40	39,53	55,73	4,36
01-ene-14 21:00:00	78,29	242,51	28,98	24,43	157,95	7,91	295,92	79,96	62,38	71,79	4,12
01-ene-14 22:00:00	82,10	249,71	30,91	29,28	157,91	12,57	309,90	79,72	72,05	75,27	4,12
01-ene-14 23:00:00	77,89	236,62	30,21	30,56	157,77	13,50	297,40	79,92	61,82	71,55	4,16
02-ene-14 00:00:00	66,53	209,58	24,46	27,22	159,67	12,61	261,26	80,16	45,14	60,46	4,32
02-ene-14 01:00:00	45,04	159,12	18,84	13,05	165,11	1,02	191,01	81,88	22,99	39,71	4,81
02-ene-14 02:00:00	45,63	159,26	19,60	15,77	177,03	4,12	194,63	81,85	23,73	40,39	4,82
02-ene-14 03:00:00	43,11	153,90	18,74	15,07	180,40	3,81	187,70	81,88	20,95	37,89	4,95
02-ene-14 04:00:00	42,55	152,70	21,79	9,60	167,51	-3,56	184,09	81,92	21,78	37,27	4,94
02-ene-14 05:00:00	43,59	158,31	17,59	11,26	168,08	-0,27	187,16	81,87	22,06	38,32	4,88
02-ene-14 06:00:00	44,82	160,52	17,52	13,33	174,56	2,30	191,37	81,86	22,50	39,68	4,82
02-ene-14 07:00:00	41,52	153,47	16,31	12,43	177,55	0,90	182,22	81,96	20,31	36,33	5,02
02-ene-14 08:00:00	69,89	219,55	25,22	25,25	163,78	9,31	270,02	81,28	47,13	63,93	4,22
02-ene-14 09:00:00	77,90	238,54	27,79	29,92	157,70	12,46	296,25	80,95	58,69	71,12	4,17
02-ene-14 10:00:00	79,64	242,12	30,60	31,12	157,86	10,23	303,84	80,72	66,01	72,92	4,17
02-ene-14 11:00:00	83,14	259,36	31,61	17,83	157,82	1,00	308,80	80,53	84,62	76,30	4,05
02-ene-14 12:00:00	79,22	244,79	27,98	26,10	157,80	8,65	298,87	80,93	62,80	72,58	4,12
02-ene-14 13:00:00	80,94	253,33	28,54	21,15	157,64	4,32	303,03	80,82	73,64	74,35	4,08
02-ene-14 14:00:00	81,40	253,94	29,30	18,61	157,57	2,02	301,85	80,82	77,59	74,48	4,05
02-ene-14 15:00:00	77,63	241,30	27,89	20,63	157,63	3,65	289,83	81,04	60,86	70,56	4,11
02-ene-14 16:00:00	81,82	248,45	28,86	32,56	157,68	15,37	309,87	80,70	66,25	74,54	4,16
02-ene-14 17:00:00	83,04	254,76	29,39	28,22	157,53	11,82	312,37	80,48	78,07	76,06	4,11
02-ene-14 18:00:00	80,75	244,55	28,67	34,16	157,63	17,38	307,38	80,58	63,64	74,33	4,14
02-ene-14 19:00:00	81,95	251,71	28,88	29,70	157,58	12,92	310,29	80,60	70,84	75,37	4,12
02-ene-14 20:00:00	81,22	244,36	29,95	35,37	157,57	17,94	309,68	80,53	63,54	74,34	4,17
02-ene-14 21:00:00	80,72	240,66	30,22	39,78	157,64	21,47	310,66	80,56	58,83	73,79	4,21
02-ene-14 22:00:00	80,72	237,52	32,49	43,66	157,67	24,67	313,67	80,56	58,64	73,88	4,25

02-ene-14 23:00:00	82,78	248,90	30,37	38,23	157,62	20,88	317,50	80,29	67,18	76,00	4,18
03-ene-14 00:00:00	82,44	249,64	29,86	34,96	157,74	17,33	314,46	80,54	67,82	75,63	4,16
03-ene-14 01:00:00	83,45	255,55	29,79	30,30	157,67	13,90	315,65	80,30	75,22	76,84	4,11
03-ene-14 02:00:00	82,10	246,45	30,05	38,67	157,74	20,90	315,16	80,48	65,19	75,36	4,18
03-ene-14 03:00:00	81,57	245,87	31,72	33,50	157,70	15,73	311,10	80,51	65,78	74,71	4,16
03-ene-14 04:00:00	80,83	243,33	32,90	32,72	157,65	14,44	308,95	80,53	65,29	74,05	4,17
03-ene-14 05:00:00	78,23	229,84	34,36	39,58	157,65	20,05	303,78	80,67	59,11	71,64	4,24
03-ene-14 06:00:00	82,81	243,99	36,97	36,84	157,66	17,26	317,79	80,48	68,79	76,33	4,16
03-ene-14 07:00:00	81,99	245,48	32,77	35,60	157,60	17,37	313,84	80,44	65,34	75,71	4,15
03-ene-14 08:00:00	79,36	234,50	33,82	38,86	157,82	20,14	307,18	80,60	57,23	73,12	4,20
03-ene-14 09:00:00	79,96	247,65	30,20	22,77	158,30	5,58	300,62	80,72	72,50	73,69	4,08
03-ene-14 10:00:00	88,69	281,32	32,00	11,27	158,57	-3,15	324,59	79,77	99,29	82,32	3,94
03-ene-14 11:00:00	85,42	277,18	30,07	0,13	158,45	-13,24	307,38	79,90	99,30	79,38	3,87
03-ene-14 12:00:00	89,83	290,37	32,75	0,04	158,39	-14,31	323,16	79,72	99,30	83,39	3,88
03-ene-14 13:00:00	82,31	262,38	30,64	7,97	157,83	-6,67	300,99	79,84	90,03	76,07	3,96
03-ene-14 14:00:00	80,34	253,23	29,85	11,24	157,82	-4,46	294,32	80,30	84,65	73,83	3,99
03-ene-14 15:00:00	83,23	259,12	31,67	17,34	157,58	1,02	308,14	79,93	87,87	76,46	4,03
03-ene-14 16:00:00	81,35	233,24	35,26	50,59	158,59	31,51	319,09	79,82	61,11	74,48	4,28
03-ene-14 17:00:00	78,22	218,36	30,60	68,77	159,50	49,33	317,74	79,73	42,27	71,20	4,46
03-ene-14 18:00:00	76,35	212,13	30,62	69,35	160,06	50,01	312,10	79,91	40,04	69,51	4,49
03-ene-14 19:00:00	78,91	222,87	32,34	60,91	158,98	41,56	316,12	79,88	45,03	71,98	4,39
03-ene-14 20:00:00	77,46	225,45	30,50	49,60	159,12	30,89	305,55	80,08	45,31	70,51	4,33
03-ene-14 21:00:00	77,13	221,16	31,15	55,08	159,40	35,67	307,40	80,17	44,64	70,11	4,38
03-ene-14 22:00:00	76,63	217,05	31,93	57,95	159,94	37,72	306,94	80,17	42,82	69,80	4,40
03-ene-14 23:00:00	78,05	226,09	29,72	54,02	159,08	34,67	309,83	80,17	46,18	71,38	4,34
04-ene-14 00:00:00	77,69	221,57	30,89	60,12	159,55	40,75	312,58	80,03	42,32	71,16	4,39
04-ene-14 01:00:00	79,50	226,68	30,87	61,58	159,38	42,33	319,13	79,85	45,48	72,97	4,37
04-ene-14 02:00:00	77,59	218,90	31,03	63,68	159,52	43,57	313,61	80,11	41,61	71,03	4,42

04-ene-14 03:00:00	77,77	216,82	32,31	67,23	163,97	46,26	316,36	80,06	41,00	71,16	4,45
04-ene-14 04:00:00	76,83	215,92	31,19	64,94	161,98	45,26	312,05	79,99	42,01	70,19	4,45
04-ene-14 05:00:00	73,39	208,31	29,80	60,34	161,15	40,20	298,45	80,27	38,69	67,14	4,45
04-ene-14 06:00:00	70,09	196,59	28,44	64,18	166,36	44,21	289,21	80,51	34,45	63,76	4,54
04-ene-14 07:00:00	74,30	210,03	29,91	61,66	163,81	41,36	301,61	80,35	38,94	67,81	4,45
04-ene-14 08:00:00	75,25	214,25	29,30	60,66	160,21	41,17	304,22	80,27	40,06	68,86	4,42
04-ene-14 09:00:00	60,56	180,88	24,17	47,28	169,66	30,40	252,33	81,24	28,49	54,70	4,61
04-ene-14 10:00:00	60,46	180,70	24,66	45,90	169,92	29,21	251,26	81,42	29,45	54,58	4,60
04-ene-14 11:00:00	70,72	206,80	26,85	51,90	160,46	34,04	285,55	81,10	37,91	64,36	4,44
04-ene-14 12:00:00	76,86	221,66	31,53	54,07	158,95	34,58	307,27	80,80	45,05	70,22	4,38
04-ene-14 13:00:00	73,01	207,51	30,64	59,96	159,42	40,34	298,12	80,89	38,50	66,46	4,49
04-ene-14 14:00:00	77,01	219,20	31,12	61,54	159,12	41,84	311,86	80,59	42,62	70,03	4,45
04-ene-14 15:00:00	78,46	228,83	31,06	51,76	158,42	32,79	311,65	80,58	49,93	71,48	4,36
04-ene-14 16:00:00	79,86	233,95	33,17	46,88	157,90	27,86	313,99	80,49	55,56	72,92	4,31
04-ene-14 17:00:00	79,68	233,00	38,55	39,55	157,95	19,68	311,10	80,63	59,51	72,88	4,27
04-ene-14 18:00:00	77,95	232,31	27,17	47,44	157,98	29,90	306,92	80,63	51,16	71,43	4,30
04-ene-14 19:00:00	74,73	220,63	27,90	47,94	158,20	30,11	296,47	80,88	45,13	68,16	4,35
04-ene-14 20:00:00	75,33	219,11	29,24	52,94	158,31	33,79	301,29	80,85	44,03	68,55	4,40
04-ene-14 21:00:00	74,24	216,17	29,53	51,17	158,94	31,70	296,87	80,95	41,59	67,48	4,40
04-ene-14 22:00:00	71,11	199,52	31,32	61,16	161,32	39,68	292,00	81,11	36,16	64,27	4,54
04-ene-14 23:00:00	71,77	206,85	29,54	53,87	160,38	33,57	290,26	81,00	38,61	64,93	4,47
05-ene-14 00:00:00	75,25	215,20	29,40	58,01	158,42	38,40	302,62	80,89	42,92	68,44	4,42
05-ene-14 01:00:00	70,20	200,58	28,55	57,87	159,18	37,82	286,99	80,97	35,70	63,65	4,51
05-ene-14 02:00:00	51,05	163,27	20,57	35,24	174,54	20,13	219,07	81,65	23,75	45,46	4,82
05-ene-14 03:00:00	42,80	139,48	18,36	40,72	191,70	25,21	198,55	81,83	16,57	37,14	5,35
05-ene-14 04:00:00	43,77	141,96	19,06	39,11	191,67	24,06	200,13	81,83	17,99	38,08	5,26
05-ene-14 05:00:00	43,15	141,95	18,68	36,94	191,40	22,56	197,58	81,83	17,59	37,55	5,26
05-ene-14 06:00:00	45,66	145,66	20,11	41,34	189,02	25,88	207,11	81,79	19,03	40,04	5,17

05-ene-14 07:00:00	46,52	146,52	20,47	43,85	188,33	28,47	210,84	81,74	19,21	40,99	5,14
05-ene-14 08:00:00	48,66	155,57	19,56	38,55	185,11	23,93	213,68	81,71	21,83	43,04	4,96
05-ene-14 09:00:00	47,25	147,33	19,67	46,89	189,05	30,56	213,88	81,76	20,02	41,49	5,16
05-ene-14 10:00:00	57,20	172,96	23,36	43,93	177,04	26,65	240,25	81,57	27,97	50,86	4,72
05-ene-14 11:00:00	61,55	183,67	23,58	46,76	168,08	30,73	254,00	81,28	29,84	55,12	4,61
05-ene-14 12:00:00	61,80	182,89	21,86	51,68	171,71	35,83	256,43	81,33	29,90	55,20	4,65
05-ene-14 13:00:00	62,15	179,95	24,04	56,12	170,06	40,10	260,10	81,11	29,90	55,74	4,67
05-ene-14 14:00:00	60,57	179,90	22,87	49,59	172,77	34,78	252,36	81,13	29,75	54,01	4,67
05-ene-14 15:00:00	59,08	173,93	23,44	51,71	174,81	36,49	249,08	81,20	28,01	53,18	4,68
05-ene-14 16:00:00	58,66	177,92	23,70	41,58	175,50	27,35	243,19	81,24	30,05	52,56	4,63
05-ene-14 17:00:00	59,61	180,88	22,20	43,61	172,85	29,19	246,69	81,28	29,99	53,10	4,65
05-ene-14 18:00:00	58,16	177,68	21,91	42,42	172,65	28,46	242,02	81,30	28,88	52,10	4,65
05-ene-14 19:00:00	60,04	184,45	21,44	40,78	169,39	27,48	246,67	81,18	31,21	54,14	4,56
05-ene-14 20:00:00	62,08	188,23	24,23	40,00	162,43	26,21	252,47	81,11	33,60	55,93	4,51
05-ene-14 21:00:00	65,01	179,93	31,69	62,62	165,31	43,78	274,24	81,07	31,31	58,58	4,68
05-ene-14 22:00:00	77,99	221,71	35,54	52,19	158,20	32,40	309,44	80,51	48,89	71,00	4,36
05-ene-14 23:00:00	75,92	217,85	28,94	57,59	158,57	38,95	304,38	80,56	43,35	69,22	4,40
06-ene-14 00:00:00	78,12	223,27	30,26	57,74	158,93	38,69	311,27	80,62	45,81	71,26	4,37
06-ene-14 01:00:00	79,34	226,29	29,24	61,07	158,35	42,58	316,60	80,38	46,68	72,30	4,38
06-ene-14 02:00:00	76,36	218,04	29,52	58,19	158,29	39,03	305,75	80,73	43,84	69,38	4,41
06-ene-14 03:00:00	72,56	204,73	28,15	64,99	159,21	45,14	297,88	80,77	36,64	66,01	4,51
06-ene-14 04:00:00	50,82	147,70	21,50	63,50	179,66	47,52	232,71	81,47	19,48	45,39	5,13
06-ene-14 05:00:00	54,20	154,04	20,11	70,64	177,96	54,71	244,79	81,30	19,79	48,52	5,04
06-ene-14 06:00:00	62,40	172,86	22,24	75,79	171,77	57,49	270,89	81,18	25,70	56,29	4,81
06-ene-14 07:00:00	65,71	183,21	23,05	72,33	168,11	54,02	278,59	81,13	28,68	59,53	4,68
06-ene-14 08:00:00	72,42	199,25	26,10	77,43	160,66	57,23	302,78	80,43	34,55	65,78	4,60
06-ene-14 09:00:00	73,60	205,46	25,97	73,11	160,02	53,73	304,54	80,27	36,91	66,76	4,56
06-ene-14 10:00:00	73,17	207,24	26,22	63,74	158,21	44,12	297,21	80,44	38,79	66,42	4,47

06-ene-14 11:00:00	74,40	216,92	26,33	52,41	158,03	34,06	295,66	80,45	43,16	67,44	4,38
06-ene-14 12:00:00	76,90	228,86	28,17	42,71	157,87	24,74	299,74	80,39	52,04	69,93	4,29
06-ene-14 13:00:00	80,61	238,23	33,80	40,13	157,73	21,50	312,16	80,09	60,89	73,67	4,24
06-ene-14 14:00:00	80,90	242,88	31,74	37,06	157,70	18,98	311,67	80,16	64,36	73,81	4,22
06-ene-14 15:00:00	81,77	249,63	31,63	29,81	157,72	12,00	311,07	80,14	73,07	74,68	4,17
06-ene-14 16:00:00	81,30	243,47	31,25	39,46	157,65	21,68	314,18	80,12	66,08	74,03	4,24
06-ene-14 17:00:00	83,91	247,12	32,99	46,10	157,72	28,05	326,22	79,65	68,32	76,48	4,27
06-ene-14 18:00:00	84,81	250,17	32,61	46,86	157,74	28,91	329,64	79,67	71,29	77,73	4,24
06-ene-14 19:00:00	84,24	244,49	33,22	53,20	157,75	34,66	330,91	79,62	62,93	76,85	4,31
06-ene-14 20:00:00	83,27	239,51	32,41	56,83	157,83	38,15	328,75	79,71	57,94	75,95	4,33
06-ene-14 21:00:00	82,40	237,31	33,67	52,65	157,85	33,75	323,62	79,68	57,15	75,16	4,31
06-ene-14 22:00:00	80,57	230,01	32,91	55,97	157,94	36,76	318,89	79,87	52,05	73,58	4,33
06-ene-14 23:00:00	80,02	225,31	31,83	63,45	157,90	43,95	320,59	79,80	48,31	73,10	4,39
07-ene-14 00:00:00	79,00	218,17	33,43	68,57	157,96	48,57	320,17	79,79	44,78	72,29	4,43
07-ene-14 01:00:00	79,64	219,86	32,55	71,02	158,04	50,89	323,43	79,80	44,67	72,73	4,45
07-ene-14 02:00:00	66,95	184,95	28,67	71,98	164,61	49,72	285,60	80,13	33,85	60,09	4,75
07-ene-14 03:00:00	55,98	162,47	25,17	52,60	177,20	28,30	240,24	81,06	25,83	49,24	4,88
07-ene-14 04:00:00	56,80	172,62	25,18	38,30	174,10	15,40	236,10	81,08	29,24	50,24	4,70
07-ene-14 05:00:00	58,31	179,44	25,11	32,60	172,19	10,72	237,15	81,09	31,01	51,57	4,60
07-ene-14 06:00:00	55,76	167,34	25,07	41,16	179,11	17,98	233,57	81,11	27,14	49,17	4,75
07-ene-14 07:00:00	53,98	158,38	24,43	50,24	180,34	26,63	233,06	81,12	23,79	47,73	4,88
07-ene-14 08:00:00	57,89	163,15	27,42	56,94	175,27	33,07	247,50	81,00	25,70	51,53	4,80
07-ene-14 09:00:00	57,62	172,06	27,20	39,60	170,11	16,91	238,86	81,05	30,04	51,44	4,64
07-ene-14 10:00:00	62,40	201,35	26,92	9,32	157,98	-11,85	237,59	81,06	40,93	56,13	4,23
07-ene-14 11:00:00	56,31	176,62	26,08	23,94	163,99	2,21	226,63	81,17	31,21	50,01	4,53
07-ene-14 12:00:00	55,07	159,14	27,86	48,54	169,77	24,47	235,54	81,09	25,72	48,77	4,83
07-ene-14 13:00:00	58,02	168,34	27,95	44,94	173,40	21,16	241,22	81,06	28,68	51,34	4,70
07-ene-14 14:00:00	54,61	149,01	26,54	67,02	174,26	41,86	242,57	81,07	21,65	47,95	5,06

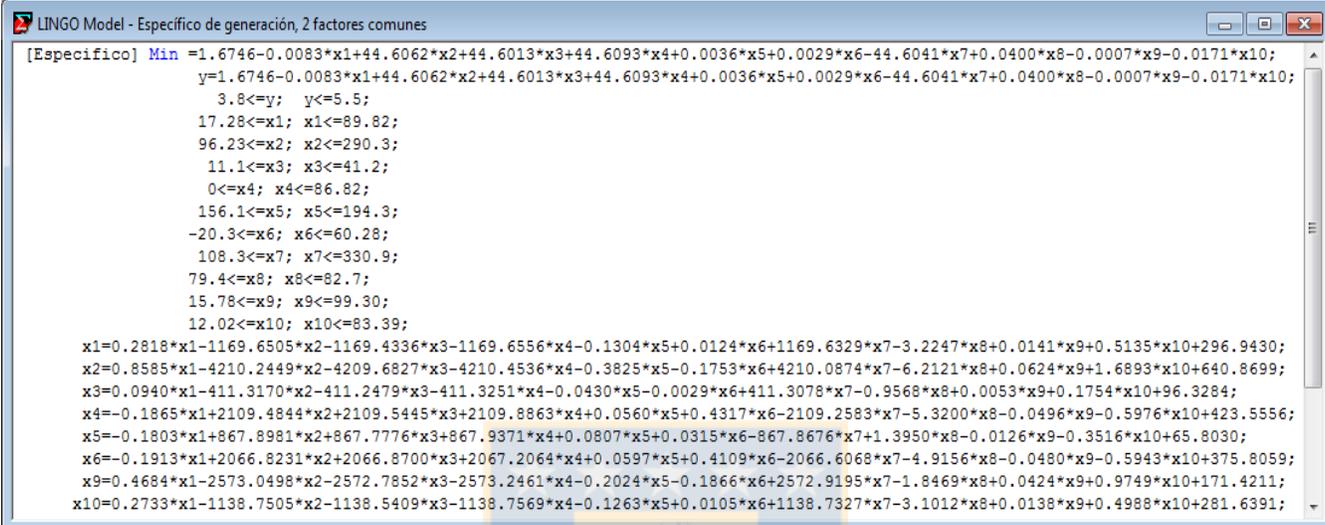
07-ene-14 15:00:00	53,53	142,14	25,33	77,83	170,00	52,45	245,29	80,97	19,30	46,93	5,23
07-ene-14 16:00:00	51,67	134,68	23,56	84,55	168,03	58,75	242,79	81,05	17,23	44,55	5,45
07-ene-14 17:00:00	50,98	132,12	24,21	83,53	166,64	57,17	239,85	81,12	17,44	43,90	5,46
07-ene-14 18:00:00	54,89	145,87	25,81	75,15	165,52	49,77	246,83	80,99	21,71	48,12	5,13
07-ene-14 19:00:00	52,17	133,42	25,08	86,82	163,28	60,28	245,33	81,01	17,98	45,68	5,37
07-ene-14 20:00:00	58,70	150,77	28,27	84,47	159,20	58,46	263,52	80,96	24,17	52,05	5,06
07-ene-14 21:00:00	73,69	209,12	31,09	56,24	156,25	36,07	296,45	80,30	48,59	67,03	4,42
07-ene-14 22:00:00	80,09	242,20	30,45	34,28	156,13	16,82	306,93	80,07	73,27	73,12	4,20
07-ene-14 23:00:00	83,47	248,04	30,82	40,60	156,29	23,10	319,46	79,85	79,91	76,41	4,18

*Para consultas de banco de datos original preguntar en Universidad de Concepción, Departamento de Ingeniería Industrial, incluido en cd de copia de entrega final.



Anexo 2: Programación del Modelo Predictivo.

La programación del modelo predictivo fue realizada mediante el software LINGO 5.0. El código ingresado al software quedó de la siguiente forma.



```
LINGO Model - Especifico de generación, 2 factores comunes
[Especifico] Min =1.6746-0.0083*x1+44.6062*x2+44.6013*x3+44.6093*x4+0.0036*x5+0.0029*x6-44.6041*x7+0.0400*x8-0.0007*x9-0.0171*x10;
y=1.6746-0.0083*x1+44.6062*x2+44.6013*x3+44.6093*x4+0.0036*x5+0.0029*x6-44.6041*x7+0.0400*x8-0.0007*x9-0.0171*x10;
3.8<=y; y<=5.5;
17.28<=x1; x1<=89.82;
96.23<=x2; x2<=290.3;
11.1<=x3; x3<=41.2;
0<=x4; x4<=86.82;
156.1<=x5; x5<=194.3;
-20.3<=x6; x6<=60.28;
108.3<=x7; x7<=330.9;
79.4<=x8; x8<=82.7;
15.78<=x9; x9<=99.30;
12.02<=x10; x10<=83.39;
x1=0.2818*x1-1169.6505*x2-1169.4336*x3-1169.6556*x4-0.1304*x5+0.0124*x6+1169.6329*x7-3.2247*x8+0.0141*x9+0.5135*x10+296.9430;
x2=0.8585*x1-4210.2449*x2-4209.6827*x3-4210.4536*x4-0.3825*x5-0.1753*x6+4210.0874*x7-6.2121*x8+0.0624*x9+1.6893*x10+640.8699;
x3=0.0940*x1-411.3170*x2-411.2479*x3-411.3251*x4-0.0430*x5-0.0029*x6+411.3078*x7-0.9568*x8+0.0053*x9+0.1754*x10+96.3284;
x4=-0.1865*x1+2109.4844*x2+2109.5445*x3+2109.8863*x4+0.0560*x5+0.4317*x6-2109.2583*x7-5.3200*x8-0.0496*x9-0.5976*x10+423.5556;
x5=-0.1803*x1+867.8981*x2+867.7776*x3+867.9371*x4+0.0807*x5+0.0315*x6-867.8676*x7+1.3950*x8-0.0126*x9-0.3516*x10+65.8030;
x6=-0.1913*x1+2066.8231*x2+2066.8700*x3+2067.2064*x4+0.0597*x5+0.4109*x6-2066.6068*x7-4.9156*x8-0.0480*x9-0.5943*x10+375.8059;
x9=0.4684*x1-2573.0498*x2-2572.7852*x3-2573.2461*x4-0.2024*x5-0.1866*x6+2572.9195*x7-1.8469*x8+0.0424*x9+0.9749*x10+171.4211;
x10=0.2733*x1-1138.7505*x2-1138.5409*x3-1138.7569*x4-0.1263*x5+0.0105*x6+1138.7327*x7-3.1012*x8+0.0138*x9+0.4988*x10+281.6391;
```

Figura A2.1. Programación del modelo predictivo en LINGO.



Anexo 3: Filtro Factorial y Pre-clasificación

El Filtro Factorial y la Pre-clasificación son herramientas que entrega FACTNET para disminuir la dimensión de modelo de red factorial inicial, sin pérdida de representatividad, dejando las variables que más inciden en el modelo y que mayor grado de correlación poseen. A continuación se presentan las gráficas de estas en el *software* FACTNET.

Filtro Factorial.

Para emplear la herramienta Factor Filter se debe seguir la siguiente secuencia en FACTNET, *Tools->Models->Factor Filter*. Luego el software despliega la siguiente sección.

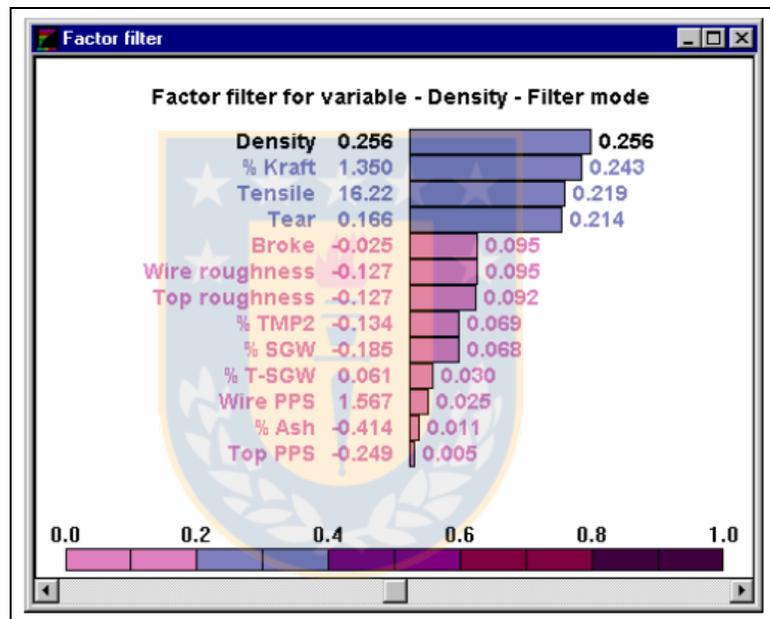


Figura A3.1. Filtro Factorial.

El sentido del Filtro Factorial es mostrar la correlación de una variable seleccionada respecto a las remanentes. Los valores a la izquierda representan las Ganancias del Modelo correspondientes a las Ecuaciones de Ganancia, mientras que los valores a la derecha son el *ranking* de variables, coincidentes con la longitud de la barra correspondiente.

Pre-clasificación.

La pre-clasificación calcula el coeficiente de correlación promedio, \bar{R}^2 para cada una de las variables respecto a las otras, de tal forma de mostrar el grado de correlación existente entre variables. El uso de la herramienta fue descrito previamente por lo que no se entrará en detalles. Para emplear el método se debe seguir la siguiente secuencia *Tools->Models->Pre-sorting* en el programa FACTNET. A continuación en la *Figura A3.2.* se muestra la ventana que emerge.

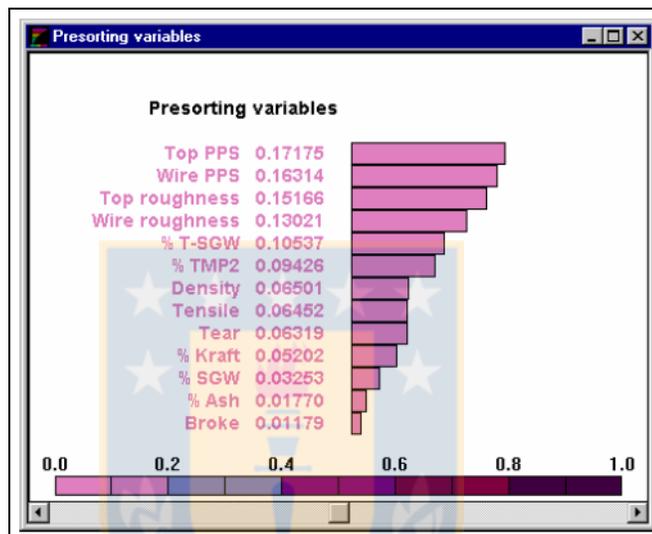


Figura A3.2. Pre-clasificación.

Los números representan el coeficiente de correlación promedio de cada variable R^2

Anexo 4: Set de datos de entrenamiento, número de Factores Comunes.

Partición de Datos.

Al abrir acceder al menú, *Analysis->Settings*, se abre la siguiente ventana.

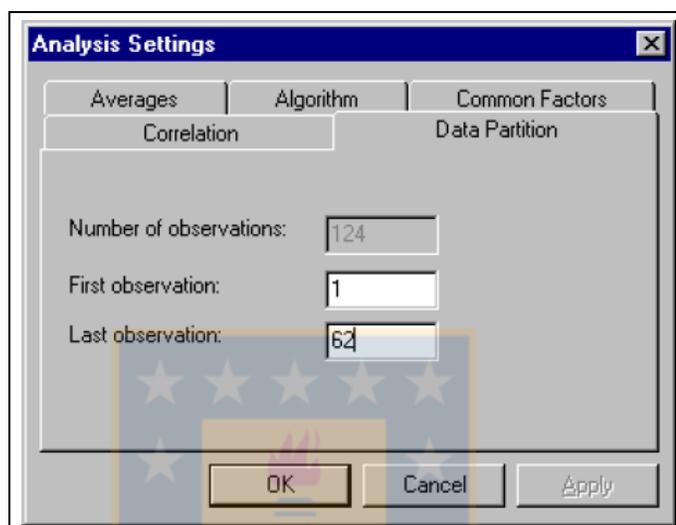


Figura A4.1. Partición de Datos.

En la ventana de Partición de Datos se elige el set de entrenamiento para el modelo y el set de validación del modelo. La primera observación se introduce en la ventana observación inicial llenando el recuadro indicado, mientras que la segunda observación se obtiene llenando el recuadro que alude a esta también. En el caso del presente estudio la primera observación sería 1, mientras que la última sería 596, correspondiente al 80% de la información inicial como se mencionó previamente.

Número de Factores Comunes.

El número de factores comunes se elige de acuerdo al grado de variabilidad que representan sus eigenvalores como se mencionó previamente. En la misma ventana presentada en la *Figura N°8.4.* previa se debe seleccionar la pestaña *Factores Comunes* y se obtiene la siguiente sub-ventana representada por la *Figura A4.2.*

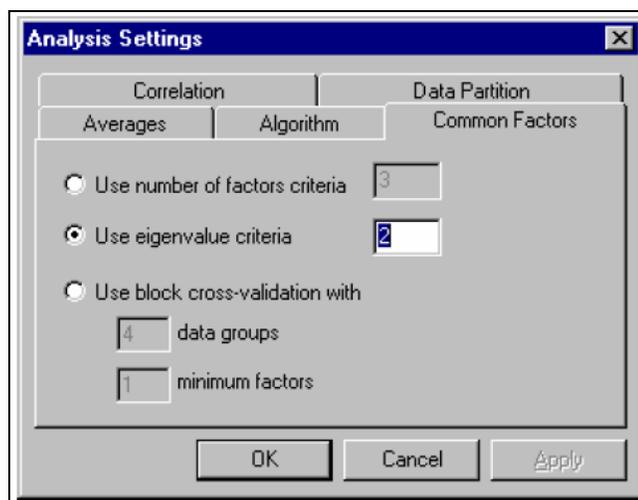


Figura A4.2. Factores Comunes.

Teniendo en cuenta el grado de variabilidad explicado se debe ir reduciendo el número de factores estos de forma tal de representar la red factorial final. La selección del número de factores comunes debe llevarse a cabo llenando el campo *Número de factores*.

