

**UNIVERSIDAD NACIONAL DE INGENIERIA  
FACULTAD DE INGENIERIA GEOLOGICA,  
MINERA Y METALURGICA**



**CONTROL ESTADÍSTICO MULTIVARIANTE EN  
CIRCUITO DE MOLIENDA EN LA  
CONCENTRADORA CUAJONE**

**TESIS**

PARA OPTAR EL TITULO PROFESIONAL DE:  
**INGENIERO METALURGISTA**

PRESENTADO POR:  
**CRISTHIAN GERONIMO CURO LORO**

LIMA – PERU  
2008

## INDICE

RESUMEN .....	6
INTRODUCCION .....	7

### CAPITULO I MARCO TEORICO

1.1.DEFINICION DE ANALISIS MULTIVARIANTE .....	9
1.2.CONCEPTOS BASICOS DE ANALISIS MULTIVARIANTE .....	10
1.2.1. Valor teorico .....	10
1.2.2. Definicion y clasificacion de las variables .....	11
a).-Tipos de variables .....	12
b).-clasificación por los valores que pueden adoptar.....	133
1.2.3. Escalas de medida .....	13
1.2.4.Variables y escalas de medida.....	14
1.2.5. Clasificación en función del análisis de datos.....	15
1.3. TIPOS DE TECNICAS MULTIVARIANTES.....	16
1.3.1.Verificacion de los supuestos del analisis multivariante.....	16
a. Normalidad .....	16
b. Homocedasticidad .....	18
c. Linealidad.....	18
1.3.2. Tipos de tecnicas multivariantes: .....	21
1.4. ANALISIS GRAFICO Y DATOS ATIPICOS.....	22
1.4.1. Histogramas y diagramas de dispersion .....	23
a. Analisis de la forma de la distribucion: .....	23
b. Análisis de relación entre variables: .....	24
1.4.2. Datos Atípicos .....	25
1.5. DISTRIBUCIONES MULTIVARIANTES .....	28
1.5.1. Distribucion de Wishart .....	28
1.5.2. La $T^2$ de Hotelling .....	31
1.6. CONTROL ESTADISTICO DE PROCESOS CONCEPTOS FUNDAMENTALES .....	34
1.6.1. Gráficos De Control.....	36
1.6.1.1.Cartas de control para $\bar{X}$ y R.....	39
1.6.1.2.Cartas de control para $\bar{X}$ y S.....	39
1.6.2. CAPACIDAD DEL PROCESO .....	47
1.6.2.1. analisis de capacidad del proceso.....	47
Utilizando un Histograma.....	48
Graficación de probabilidades .....	49
1.6.2.2. Índices de capacidad del proceso .....	53
a. Uso e interpretación $C_p$ .....	53
b. Índice de capacidad del proceso para un proceso fuera del centro.....	53
1.6.2.3. Estudios de capacidad de instrumentos y sistemas de medición .....	56

**CAPITULO II  
MARCO CONCEPTUAL**

2.1.CONFIGURACION ACTUAL DE LA PLANTA CONCENTRADORA .....	56
2.2. PROCEDIMIENTOS DE OPERACION.....	61
2.2.1. Lazos de control de molienda .....	62
a.- <b>Control de la tasa de alimentación al molino de bolas 1C</b> .....	<b>62</b>
Control automático.....	63
Modo cascada .....	63
Modo automático .....	64
Control manual .....	65
b.-Control del nivel del tanque de pulpa del polvo de recolección de polvo de recuperación de mineral fino .....	67
Control automático .....	68
Control manual .....	69
c.-Control del flujo de pulpa de polvo a los molinos de bolas 1C A 2E .....	70
Control automático .....	70
Control manual .....	71
2.3. HIPOTESIS: INCREMENTAR TONELAJE EN CIRCUITOS DE MOLIENDA .....	72
Molino 1C – Prueba N° 1:.....	72
Molino 1C – Prueba N° 2:.....	75
Molino 1D – Prueba N° 3:.....	78
Molino 1D – Prueba N° 4:.....	79
Molino 1D – Prueba N° 5:.....	80
2.4. JUSTIFICACION .....	83

**CAPITULO III  
VALIDACIÓN DEL METODO**

3.1. TEST ANALITICOS ANALISIS DE FACTORES DE VARIABILIDAD - VFA .....	85
3.1.1. Analisis factores de variabilidad.....	85
Cálculo del primer componente principal (VF1) .....	86
Cálculo del segundo componente principal (VF2).....	87
3.1.2. Índice dE Hotelling $T^2$ .....	92
3.1.3. Índice error cuadratico de prediccion - SPE.....	94
3.2.TEST ANALITICOS PROYECCION DE ESTRUCTURAS LATENTES - PLS .....	98
3.2.1.Cálculo del primer vector latente (VL1).....	100
3.2.2.Cálculo del segundo vector latente (VL2).....	102
3.3. MODELOS PREDICTIVOS.....	113
3.3.1.Determinacion de los límites de control.....	115
3.3.2.Determinacion de la capacidad del proceso: .....	116
3.3.3.Modelo predictivo %malla +65 como control de calidad:.....	117
3.3.3.1.Empleando el software Outocal: .....	119
3.3.3.2.Empleando el EXCEL:.....	122
3.3.3.3.Empleando el SCAN:.....	123

3.3.4. Modelo predictivo tonelaje:..... 128

**CAPITULO IV**

**DISCUSION - ANALISIS DE LOS RESULTADOS ..... 133**

**CAPITULO V**

**CONCLUSIONES y RECOMENDACIONES**

5.1.CONCLUSIONES ..... 135

5.2.RECOMENDACIONES..... 137

BIBLIOGRAFIA ..... 138

GLOSARIO..... 139

ANEXOS ..... 141

ANEXO A: FACTORES PARA CONSTRUIR CARTAS DE CONTROL..... 142

ANEXO B: ARQUITECTURA DEL SISTEMA DE INFORMACION DE PLANTA - PI ..... 143

ANEXO C: VARIABLES DE CONTROL EN LA ETAPA DE MOLIENDA..... 144

ANEXO D: MOLINO DE BOLAS 1C- PI PROCESS BOOK ..... 146

ANEXO E: FLOW SHEET CONCENTRADORA CUAJONE ..... 147

## INDICE DE TABLAS

TABLA N° 1 : REFERENCIA CRUZADA DE LA ALIMENTACIÓN AL MOLINO.....	63
TABLA N° 2 : VÁLVULAS DE CONTROL DEL NIVEL DEL TANQUE DE PULPA.....	69
TABLA N° 3 : REFERENCIA CRUZADA DEL CONTROL DE FLUJO DE PULPA.....	70
TABLA N° 4 : VÁLVULAS DE CONTROL DEL FLUJO DE PULPA.....	72
TABLA N° 5 : ESTADISTICA DESCRIPTIVA DEL MOLINO 1C .....	74
TABLA N° 6: MOLINO 1C PRUEBA N° 1.....	75
TABLA N° 7: MOLINO 1C PRUEBA N° 2.....	76
TABLA N° 8: VARIACIÓN DE LAS AGUAS .....	76
TABLA N° 9: RESULTADOS DE LA PRUEBA N°2.....	77
TABLA N° 10: TABLA DE LOS COEFICIENTES DE REGRESIÓN – MOLINO 1D.....	78
TABLA N° 11: MOLINO 1D PRUEBA N°3.....	79
TABLA N° 12: TABLA DE LOS COEFICIENTES DE REGRESIÓN – MOLINO 1D.....	80
TABLA N° 13: MOLINO 1D PRUEBA N° 4.....	80
TABLA N° 14: TABLA DE LOS COEFICIENTES DE REGRESIÓN – MOLINO 1D.....	81
TABLA N° 15: VARIACIÓN DE LAS AGUAS PRUEBA N°4 .....	81
TABLA N° 16: RESULTADOS DE LA PRUEBA N° 4.....	81
TABLA N° 17: TAGS DEL MOLINO 1C.....	113
TABLA N° 18: ESTADISTICA DEL MOLINO 1C.....	114
TABLA N° 18: DETERMINACIÓN DE LA CAPACIDAD DEL PROCESO .....	116
TABLA N° 19: Cp y Cpk .....	116
TABLA N° 20: MODELO % MALLA + 65- OUTOCAL .....	120
TABLA N° 21: RESULTADOS DEL MODELO % MALLA + 65 - OUTOCAL .....	121
TABLA N° 22: MODELO %SÓLIDOS O/F - OUTOCAL.....	121
TABLA N° 23: RESULTADOS ESTADÍSTICOS .....	121
TABLA N° 24: MODELO %MALLA +65 - EXCEL .....	122
TABLA N° 26: MODELO %SÓLIDOS O/F- EXCEL .....	123
TABLA N° 27: RESULTADO LMODELO %SÓLIDOS O/F- EXCEL.....	123
TABLA N° 29: RESULTADOS DEL INCREMENTO DEL TONELAJE .....	133
TABLA N° 30: BENEFICIO ECÓNOMICO DEL INCREMENTO DEL TONELAJE .....	134

## RESUMEN

---

La presente tesis trata sobre la aplicación de las técnicas de análisis multivariante en la etapa de molienda con el objetivo de incrementar el tonelaje en la Planta Concentradora – Unidad Cuajone – SPCC, haciendo uso de la tecnología utilizada como es la Aplicación SCAN.

El trabajo consiste en investigar el efecto de todas las variables que existen e interactúan en el proceso de molienda evaluando las variaciones de éstas y el efecto que tienen sobre el tonelaje. Para el estudio se hizo análisis de toda una data extraída usando el Sistema PI, el cual permitía capturar los datos de la historia de todas las variables a examinar para la generación de los modelos predictivos multivariantes.

El logro del incremento del tonelaje, requirió en principio el desarrollo de un Sensor Virtual del %malla +65 como un variable de control de calidad y del %Sólidos en el OverFlow. Los modelos generados fueron el %Carga Circulante, Potencia, %Malla +65, %Sólidos en el OverFlow y el Tonelaje. Dichos modelos permitieron realizar correcciones de las desviaciones en las condiciones de la operación, con los cuales se logró el incremento del tonelaje.

El incremento de tonelaje en la molienda aplicando el control estadístico multivariante se realizó enfocando a los 8 molinos de 16.5' x 20', el cual aumentaría si se generasen los modelos predictivos para los molinos de mayor tamaño de 20' x 33.5'. Los resultados indican que el tonelaje promedio incrementado por día fue de 141.3 TM, lo cual significaría un incremento de 22% de Cobre Total por día.

## INTRODUCCION

---

El objetivo de la tesis es dar a conocer una de las principales técnicas multivariantes como es la técnica de Componentes Principales, del cual esta basado el desarrollo del SCAN. Para ello se utilizo la tecnología con que cuenta la Planta Concentradora Cuajone como es el Sistema PI, equipos de medición de partículas en línea (PSM), los cuales fueron la plataforma para la generación de los modelos predictivos. Con el desarrollo de los modelos predictivos se pudo manejar las principales variables operativas del circuito de molienda mediante un control estadístico multivariante, el cual es una herramienta alternativa que deja de lado el control manual de las variables manipuladas por los operadores.

Para la realización de dichos modelos predictivos se tuvo en cuenta a la variable %malla +65 del Overflow como una variable de control de calidad y que junto a la variable Potencia fueron las que restringieron el límite para poder incrementar el tonelaje de la planta concentradora.

La Planta Concentradora Cuajone que está ubicada a 160Km al noreste del Puerto de Ilo al sur del Perú, y a una altura de 3360m sobre el nivel del mar, consta de:

- Tres circuitos de chancado en tres etapas: primario, secundario y terciario
- Molienda en molinos de bolas Allis Chalmers y Svedala.
- Flotación de Cobre – Moly en celdas OK100 y OK160, Wemco, Dorr – Oliver y celdas columna.
- Recuperación de Moly en Planta de Molibdeno.
- Espesamiento en espesadores Eimco.

- Filtrado de concentrado de cobre en filtros rotatorios Eimco y filtro Larox.
- Recuperación de agua en los Espesadores de relaves N° 1 y 2 de la marca Eimco, en el Espesador N° 3 de Dorr – Oliver, y en el Espesador N° 4 de Hi – Rate Outokumpu.

La concentradora y las instalaciones relacionadas están diseñadas para tratar 87 360 TMD de mineral con un contenido promedio de 0.72 %Cu y 0.029 %MO. La concentradora produce dos tipos de concentrado de mineral:

- Aproximadamente 1,935 TMD de concentrado de Cobre con un promedio de 25 %Cu y 0.127% MoS<sub>2</sub>.
- Aproximadamente 15TMD de concentrado de MO con un promedio de 92 %MoS<sub>2</sub> y menos de 1.2 %Cu.

El análisis de los métodos multivariantes predominará en el futuro y dará por resultado cambios drásticos en el modo que los investigadores piensen sobre los problemas y en como diseñan sus investigaciones. Esos métodos hacen posible plantear preguntas específicas y precisas de considerable complejidad de marcos idóneos, lo que posibilita llevar a cabo investigaciones teóricamente significativas y evaluar los efectos de las variaciones paramétricas ocurridas de forma natural en el contexto en que normalmente ocurren. De esta forma, se pueden preservar las correlaciones naturales entre las múltiples influencias sobre el comportamiento y se pueden estudiar estadísticamente los efectos aislados de esas influencias sin provocar el típico aislamiento de esos individuos o variables.



# CAPITULO I

## MARCO TEORICO

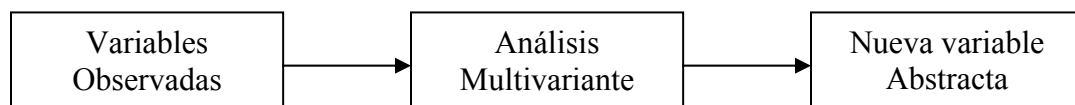
---

### 1.1. DEFINICION DE ANALISIS MULTIVARIANTE

Las técnicas multivariantes son un conjunto de métodos estadísticos que permiten el análisis de forma simultánea de más de dos variables observadas en una investigación. Desde una concepción amplia podemos definir el análisis multivariante como un conjunto de métodos que analizan las relaciones entre un número razonablemente amplio de variables (medidas), tomadas sobre cada elemento de análisis, en una o más muestras simultáneamente.

A través del análisis multivariante lo que hacemos es combinar todas las variables, eliminando la información redundante y se obtiene una nueva variable que no es observable directamente, que representa un concepto abstracto que se puede medir obteniéndose un valor para cada elemento.

Esta situación la podemos resumir en el siguiente esquema



Las técnicas de análisis multivariante se utilizan cada vez más en la investigación comercial por las siguientes razones:

1. Permiten el analizar un gran número de encuestas. Simplificando muchos datos, con la mínima pérdida de información. Consiguiendo hacer más comprensible la información para la mente humana.

2. Permiten analizar toda la información acerca de un determinado fenómeno, considerando simultáneamente todos los factores que intervienen.
3. Permiten trabajar con cualquier tipo de variable.

## **1.2. CONCEPTOS BASICOS DE ANALISIS MULTIVARIANTE**

Aunque el análisis multivariante tiene sus raíces en la estadística univariante y bivariante, la extensión al dominio multivariante introduce conceptos y cuestiones adicionales.

Estos conceptos van desde la necesidad de un entendimiento conceptual del elemento básico del análisis **multivariante – el valor teórico –** a las cuestiones específicas acerca de los tipos de escalas de medida utilizadas y los resultados estadísticos de los tests de significación y los intervalos de confianza. Cada concepto juega un papel importante en la correcta aplicación de cualquier técnica multivariante.

### **1.2.1 VALOR TEORICO**

El propósito del análisis multivariante es medir, explicar y predecir el grado de relación de los valores teóricos (combinaciones ponderadas de variables). Por tanto, el carácter multivariante reside en los múltiples valores teóricos (combinaciones múltiples de variables) y no en el número de variables u observaciones.

El elemento esencial del análisis multivariante es el valor teórico, una combinación lineal de variables con ponderaciones determinadas empíricamente. El investigador especifica las variables, mientras que las ponderaciones son objeto específico de determinación por parte de la técnica multivariante. Un valor teórico de “n” variables ponderadas ( $X_1$  a  $X_n$ ) puede expresarse matemáticamente así:

$$\text{Valor Teórico} = w_1X_1 + w_2X_2 + w_3X_3 + \dots + w_nX_n \dots\dots\dots(1)$$

Donde  $X_n$  es la variable observada y  $w_n$  es la ponderación determinada por la técnica multivariante.

El resultado es un valor único que representa una combinación de todo el conjunto de variables que mejor se adaptan al objeto del análisis multivariante específico. En regresiones múltiples el valor teórico se determina de tal forma que guarde la mejor correlación con la variable que se está prediciendo.

### **1.2.2 DEFINICION Y CLASIFICACION DE LAS VARIABLES**

En las técnicas del análisis multivariante se entiende por variable alguna magnitud que representa la característica de los elementos objeto de investigación que tratamos de medir.

En una primera clasificación las variables las podemos clasificar en dos grupos, variables independientes (VI) y variables dependientes (VD).

Las variables dependientes son aquellas cuyo comportamiento es explicado o pronosticado por una o más variables independientes. Las variables dependientes también se denominan variables criterio o respuesta, mientras que las variables independientes son las que servirán para explicar el fenómeno estudiado y en ocasiones se denominan como variables explicativas, factores o variables predictoras.

En los estudios no experimentales, la situación de las variables no siempre es clara, definiéndose su papel en el contexto de la investigación. Una misma variable puede adoptar diferentes roles en función de situaciones. Cuando existen diferencias sistemáticas en una variable dependiente (Y) asociada a diferentes niveles de variación de la variable independiente (X) se dice que están relacionadas.

Si todas las variables desempeñan el mismo papel se habla de relaciones de interdependencia. En este caso no hay una variable con las que se intente explicar el comportamiento de otras. En ocasiones, al analizar el modelo la única forma de diferenciar las variables es simplemente por donde están situadas en la ecuación.

#### a).- TIPOS DE VARIABLES

Una variable es una característica o propiedad de un elemento (individuo, objeto, transacción, suceso, etc.), que toma distintos valores para cada elemento. En general se clasifican en dos grandes grupos:

- Variables no métricas o cualitativas
- Variables métricas o cuantitativas

Los diferentes tipos de variable los resumimos a continuación:

Variables **cuantitativas** o **métricas**: Son aquellas en las que los valores tomados por diferentes individuos tienen un significado propio. De hecho, son una medición o cuantificación de una determinada característica, la respuesta a la pregunta: ¿Cuántos? Ejemplos: altura, peso, edad, hijos, ingresos, de un individuo, empleados, oficinas, beneficios, de una empresa, etc.

Variables **cualitativas** o **no métricas (atributos)**: En las que las distintas características de los elementos estudiados son cualidades o categorías alfabéticas. Sin embargo, con el fin de facilitar el tratamiento de los datos, estas categorías se convierten en unos códigos, sin que tenga que existir ningún tipo de relación entre el valor asignado y el significado de la categoría representada. Ejemplos: nacionalidad, sexo, religión, estudios cursados, clase social, calificación (Suspenso, Aprobado, Notable, Sobresaliente), etc.

### **b).- CLASIFICACIÓN POR LOS VALORES QUE PUEDEN ADOPTAR**

De acuerdo con el valor que pueden adoptar las variables estas las podemos clasificar en los siguientes tipos:

- Variable continua. Se trata de una variable cuantitativa que puede adoptar cualquier valor numérico, Para todo par de valores siempre podemos encontrar uno intermedio. Por ejemplo la edad, el consumo de teléfono.
- Variable discreta. Puede adoptar un número finito de valores distintos, entre dos valores consecutivos no se puede encontrar ninguno intermedio. Por ejemplo el número de personas por hogar.
- Variable dicotómica o binaria. Solo puede tomar dos valores, si se definen como 0 y 1 se llama binaria.
- Variables ficticias o Dummy. Se utiliza con variables cualitativas, para poder obtener información a través de operaciones, se convierten en binarias, indicando el valor 1 la presencia de una categoría de la variable y 0 su ausencia. Para realizar la conversión de una variable cualitativa en ficticia se necesitan tantas variables dummy como categorías tiene la variable menos una.

## **ESCALAS DE MEDIDA**

Prácticamente todas las investigaciones de mercado recogen los datos en forma de números, interesando al investigador lo que estos números representan, por medio de las correspondientes operaciones de medida.

Medir consiste en asignar números a los sucesos, elementos, objetos, atributos, según unas normas predeterminadas.

Puesto que se utiliza diferentes reglas para la asignación de los números, un mismo número puede dar lugar a diferentes interpretaciones, ello da lugar a la existencia de diferentes escalas de medida.

Por escala de medida entenderemos la correspondencia entre los números asignados a las propiedades de los elementos y la significación de los cálculos matemáticos realizados con los números. Básicamente en el análisis estadístico se utilizan las cuatro escalas siguientes:

Nominal, Ordinal, Intervalo y Ratio o de proporción. Las características de estas escalas se resumen a continuación:

- **Nominal:** los posibles valores de la variable representan diferentes categorías, no existiendo ninguna relación entre el código asignado a una categoría y su significado. Ejemplos: profesión, raza, estado civil, o Caso particular: variables dicotómicas, sólo admiten dos posibles respuestas. Ejemplos: sexo, verdadero/falso, si / no. Se denominan binarias si se codifican 0 / 1.
- **Ordinal:** los códigos o valores de cada categoría mantienen la misma relación de orden que el significado de las categorías. Ejemplos: clase social, escala de preferencia.
- **Intervalo:** los códigos asignados a diferentes respuestas permiten conocer la magnitud de la característica medida, ya que se mantiene una relación de orden y distancia. Ejemplos: temperatura, fechas, cualquier variable redondeada.
- **Ratio o razón:** los códigos representan el propio valor de la característica estudiada, observándose una relación de orden y de distancia y la existencia de un origen. Ejemplos: cifra de ventas, ratio económico financiero.

#### 1.2.4. VARIABLES Y ESCALAS DE MEDIDA

Partiendo de los dos grandes grupos de variables, (cualitativas y métricas), podemos resumir la relación entre las variables y las escalas de medida como sigue:

1. Variables no métricas o cualitativas, vienen medidas en escala nominal u ordinal.
2. Variables métricas o cuantitativas se utilizan las escalas de intervalo o de razón.
3. Variables binarias se utiliza la escala de razón.

### 1.2.5. CLASIFICACIÓN EN FUNCIÓN DEL ANÁLISIS DE DATOS

En ocasiones en el estudio se deben realizar transformaciones de escala y origen, las variables las podemos dividir en:

**Valores o puntuaciones directas**, también llamadas brutas, se obtienen directamente del instrumento de medida y en sus mismas dimensiones. Se suelen representar por letras mayúsculas X, Y, Z, ... teniendo medias  $m_X$ ,  $m_Y$ ,  $m_Z$  ... y las correspondientes desviaciones típicas  $s_X$ ,  $s_Y$ ,  $s_Z$ , ... medidas en la misma escala.

**Valores o puntuaciones centradas en la media o diferenciales**, son el resultado de un cambio en el origen al obtenerse de la resta de la media del valor, se suelen representar con letras minúsculas (x, y, z, ...) y se obtienen a partir de la siguiente operación:

$$x = X - m_x \quad \dots\dots\dots (2)$$

Este tipo de puntuación tiene de media 0 y una desviación típica igual a la de las puntuaciones originales. Se produce un cambio de origen no de escala.

**Valores típicos o estandarizados**, se obtienen restando de cada valor la media y dividiendo por la desviación típica. Se suelen representar por la letra Z y el subíndice de a correspondiente categoría de la variable.

Se obtienen de:

$$Z = \frac{X - m_x}{s_x} \dots\dots\dots (3)$$

Los valores tipificados están libres de escala y siempre tienen media igual a 0 y desviación típica igual a 1.

### 1.3 TIPOS DE TÉCNICAS MULTIVARIANTES

El empleo de las técnicas de análisis multivariante implica la necesidad de comprobar algunos supuestos estadísticos. Esta necesidad aumenta como consecuencia de dos características de este tipo de análisis.

1. La complejidad de las relaciones, debido al uso habitual de una gran cantidad de variables, hace que las distorsiones y los sesgos potenciales sean más potentes cuando se incumplan los supuestos.
2. La complejidad de los análisis y de los resultados pueden enmascarar los “signos” de las violaciones de los supuestos que son aparentes en los más sencillos análisis multivariantes.

El análisis multivariante requiere que los supuestos subyacentes a las técnicas estadísticas sean contrastados dos veces: en primer lugar para las variables aisladas, semejante a la prueba de los supuestos del análisis univariante, y en segundo lugar para el valor teórico del modelo multivariante, que actúa colectivamente sobre las variables a analizar y por tanto debe cumplir los mismos supuestos que las variables individuales.

#### 1.3.1. VERIFICACION DE LOS SUPUESTOS DEL ANALISIS MULTIVARIANTE

##### a. NORMALIDAD

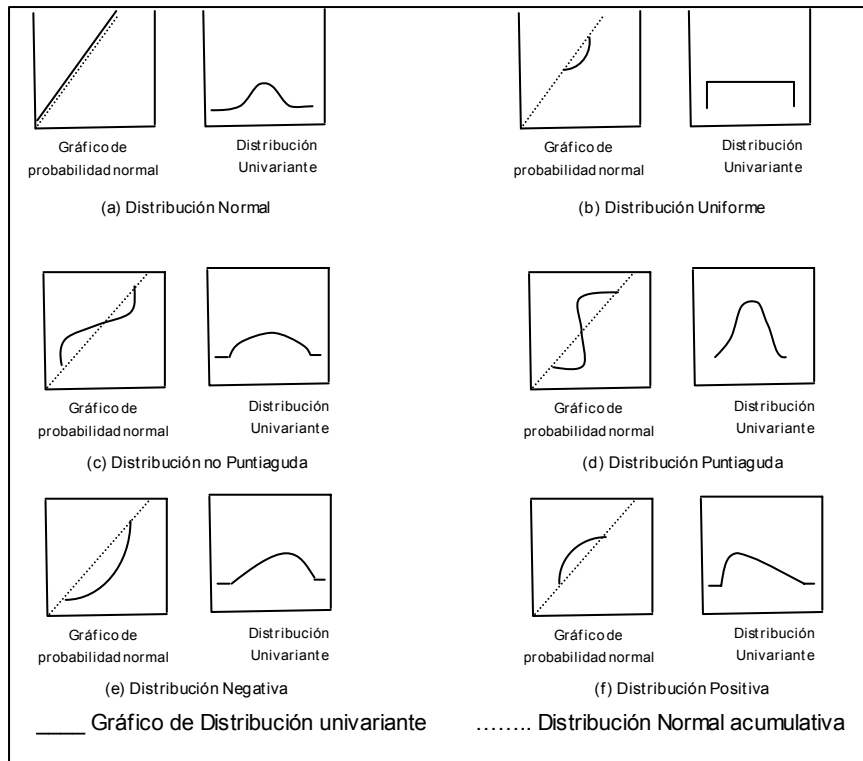
El supuesto fundamental del análisis multivariante es la normalidad de los datos, en referencia al perfil de la distribución de los datos para una única variable métrica y su correspondencia con una distribución



normal, punto de referencia de los métodos estadísticos. Si la variación respecto de la distribución normal es suficientemente amplia, todos los test estadísticos resultantes no son válidos, dado que se requiere la normalidad para el uso de los estadísticos t (test de Student) y de F (test de Fisher).

La normalidad univariante para una variable es fácil de contrastar, siendo posible varias medidas correctoras. La normalidad multivariante (la combinación de dos o más variables) implica que las variables individuales son normales en un sentido univariante y que sus combinaciones también sean normales.

Por tanto, si una variable es normal multivariante, es también normal univariante. Sin embargo, lo contrario no es necesariamente cierto (dos o más variables normales univariantes no son necesariamente normal multivariante).

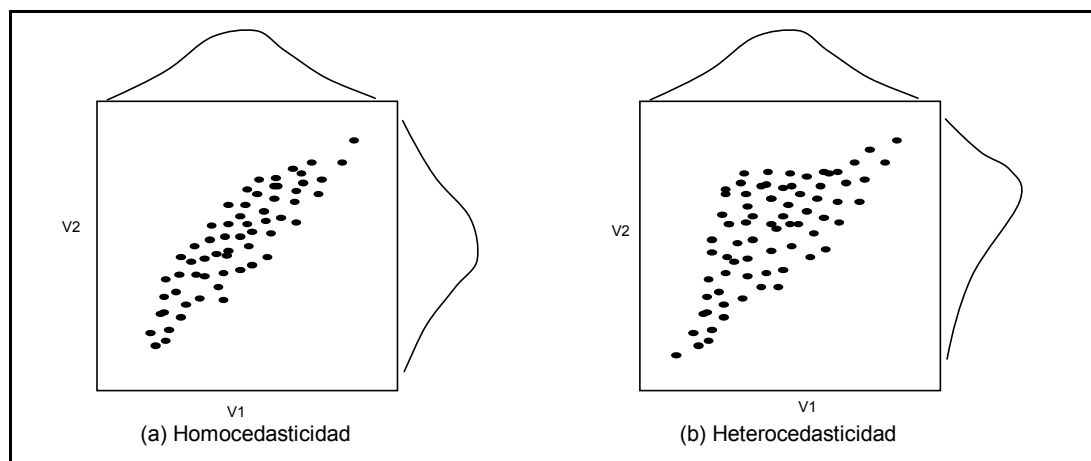


**Figura N° 1**

### b. HOMOCEASTICIDAD

La homocedasticidad es un supuesto relativo primordialmente a las relaciones de dependencia entre variables. Se refiere al supuesto que las variables dependientes exhiban iguales niveles de varianza a lo largo del rango del predictor de las variables. La homocedasticidad es deseable porque la varianza de la variable dependiente que se esta explicando en la relación de dependencia no debería concentrarse solo en un limitado rango de los valores independientes. Aunque las variables dependientes deben ser métricas, este concepto de igual dispersión de la varianza a lo largo de las variables independientes puede aplicarse cuando las variables son métricas o no métricas.

Con variables independientes métricas, el concepto de homocedasticidad se basa en la dispersión de la varianza de la variable dependiente a lo largo del rango de los valores de la variable independiente, que se encuentra en técnicas como la regresión múltiple. El mismo concepto se aplica también cuando las variables independientes son no métricas.



*Figura N° 2*

### **c. LINEALIDAD**

La linealidad es un supuesto implícito de todas las técnicas multivariantes basadas en medidas de correlación, incluyendo la regresión múltiple, regresión logística, análisis factorial y los modelos de ecuaciones estructurales. Dado que las correlaciones representan solo la asociación lineal entre variables, los efectos no lineales no estarán representados en el valor de la correlación. Como resultado, es siempre prudente examinar todas las relaciones para identificar cualquier desplazamiento de la linealidad que pueda impactar la correlación.

**Identificación de relaciones no lineales:** La forma más común de evaluar la linealidad es examinar los gráficos de dispersión de las variables e identificar cualquier pauta no lineal en los datos. Una aproximación alternativa es ir a un análisis de regresión múltiple y examinar los residuos. Los residuos reflejan la parte no explicada de la variable dependiente: por tanto, cualquier parte no lineal de la relación quedará reflejada en los residuos.

### **TRANSFORMACIONES DE LOS DATOS:**

Las transformaciones de los datos proporcionan un medio de modificar variables por una o dos razones: corregir el incumplimiento de los supuestos estadísticos subyacentes a las técnicas multivariantes o mejorar la relación (correlación) entre variables. La transformación de los datos puede basarse en razones tanto “teóricas” como “derivadas de los datos”. Así, en cada caso el investigador debe proceder muchas veces por ensayo y error, ponderando la mejora frente a la necesidad de transformaciones adicionales.

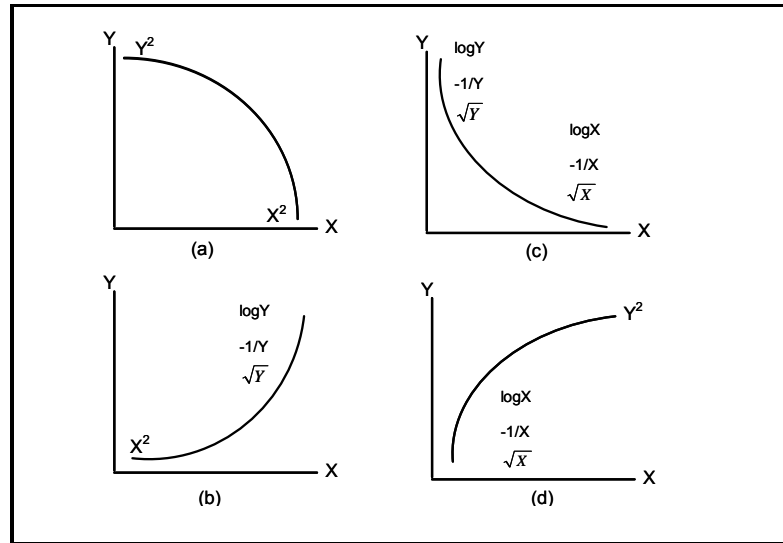
*Transformaciones de los datos para conseguir la normalidad y la homocedasticidad:*

Las transformaciones de los datos proporcionan el medio principal de corregir la no normalidad y heterocedasticidad. En ambos casos, la forma de las variables sugiere transformaciones específicas. Para las distribuciones no normales, las dos formas más comunes son las distribuciones “planas” y las distribuciones “asimétricas”. Para la distribución plana, las transformaciones más comunes son la inversa (es decir,  $1/Y$  o  $1/X$ ). Las distribuciones asimétricas pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable.

La heterocedasticidad es un problema asociado a la normalidad, y en muchos casos la solución del problema tiene que ver también con los problemas de normalidad. La heterocedasticidad se debe también a la distribución de la variable.

*Transformaciones para conseguir la linealidad:*

Existen numerosos procedimientos para conseguir la linealidad entre dos variables, pero las relaciones no lineales más simples pueden clasificarse en cuatro categorías (figura 3). En cada cuadrante, se muestran las transformaciones potenciales para variables dependiente e independiente. Por ejemplo. Si las relaciones locales son como las de la Figura 3 (a) se aplica la raíz cuadrada para conseguir la linealidad. Cuando se muestran las posibilidades de transformación múltiple se empieza con el método más adecuado para cada cuadrante para después bajar hasta que se consigue la linealidad.



*Figura N° 3*

### 1.3.2. TIPOS DE TECNICAS MULTIVARIANTES:

El análisis multivariante es un conjunto de técnicas de análisis de datos en expansión. Entre las técnicas mas conocidas tenemos:

1. Regresión múltiple y correlación múltiple
2. Análisis discriminante múltiple
3. Componentes principales y análisis factorial común
4. Análisis multivariante de varianza y covarianza
5. Correlación canónica
6. Análisis cluster
7. Análisis multidimensional, y
8. Análisis conjunto.

Entre las técnicas más emergentes tenemos:

9. Análisis de correspondencias
10. Modelos de probabilidad lineal como logit y probit; y
11. Modelos de ecuaciones simultaneas/estructurales.

Existe varios software estadísticos uno de los más conocidos el STATGRAPHICS Plus, el cual contiene un Módulo de Métodos Multivariantes, el cual consta de cinco tipos de análisis:

- Análisis Cluster.
- Análisis Factorial.
- Análisis de los Componentes Principales.
- Análisis Discriminante, y
- Análisis de Correlación Canónica.

En el caso de la presente tesis, se va a trabajar con la Aplicación SCAN, el cual trabaja con el Método Multivariante de los Componentes Principales.

#### **1.4. ANALISIS GRAFICO Y DATOS ATIPICOS**

El examen de los datos es un paso necesario, que lleva tiempo, y que habitualmente se descuida por parte de los analistas de datos. Las tareas previas implícitas en el examen previo de los datos pueden parecer insignificantes y sin consecuencias a primera vista, pero son parte esencial del análisis multivariante.

Existen cuatro fases distintas para un examen de datos:

- a. Un examen gráfico de la naturaleza de las variables a analizar y las relaciones que forman las bases del análisis multivariante.
- b. Un proceso de evaluación para entender el impacto que pueden tener los datos ausentes sobre el análisis.
- c. Las técnicas que mejor se ajustan para la identificación de casos atípicos.
- d. Los métodos analíticos necesarios para evaluar adecuadamente la capacidad de los datos para cumplir los supuestos estadísticos específicos de muchas técnicas multivariantes.

### **1.4.1. HISTOGRAMAS Y DIAGRAMAS DE DISPERSION**

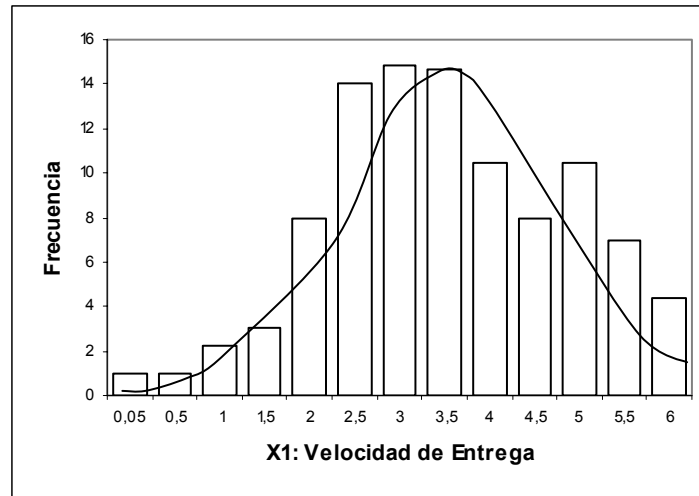
#### **a. ANALISIS DE LA FORMA DE LA DISTRIBUCION:**

El punto de partida para entender la naturaleza de cualquier variable es caracterizar la forma de su distribución. Muchas veces el investigador puede obtener una perspectiva adecuada de la variable a través de un histograma.

Un histograma es una representación grafica de los datos que muestra la frecuencia de los casos (valores de los datos) en categoría de datos. Las frecuencias se representan para examinar la forma de la distribución de las respuestas.

Si el rango de respuestas va de 1 a10, el investigador puede construir un histograma contando el número de respuestas que fueron un uno, un dos, etc. Para las variables continuas se forman categorías, dentro de las cuales la frecuencia de los valores de datos esta tabulada. Por ejemplo según la figura 4 las categorías de puntos intermedios son 0, 0.5, 1, 1.5.....hasta 6.0. La altura de las barras representa la frecuencia de los valores de los datos en cada categoría. Si el examen de la distribución tiene como objetivo evaluar su normalidad, se puede superponer la curva normal sobre la distribución, como se ha hecho en la Figura 4.

El histograma puede utilizarse para examinar cualquier tipo de variable, desde los valores originales a los residuos de una técnica multivariante



*Una representación gráfica de una distribución universal: El histograma.*

***Figura N° 4***

**b. ANÁLISIS DE RELACIÓN ENTRE VARIABLES:**

El método mas popular de análisis de las relaciones bivariantes es el GRÁFICO DE DISPERSIÓN, un gráfico de puntos de datos basados en dos variables. Se presenta una variable en el eje horizontal y la otra en el vertical. El patrón de los puntos representa la relación entre las variables. Cuando los puntos se organizan a lo largo de una línea recta, tenemos una relación lineal de correlación. Un conjunto de puntos curvados puede indicar una relación no lineal.

Existen muchos tipos de gráficos de dispersión, pero un formato que se ajusta particularmente a las técnicas multivariantes es la MATRIZ DEL GRÁFICO DE DISPERSIÓN. Aquí se representa el gráfico de dispersión para todas las combinaciones de variables en la porción inferior de la matriz. Las diagonales contienen los histogramas de las variables. En la parte superior de la matriz se incluyen las correlaciones correspondientes para que el investigador pueda valorar la correlación representada en cada grafico de dispersión.



### 1.4.2. DATOS ATÍPICOS

Los casos atípicos son observaciones con una combinación única de características identificables que les diferencia claramente de las otras observaciones. Los casos atípicos no pueden ser caracterizados categóricamente como benéficos o problemáticos sino que deben ser contemplados en el contexto del análisis y deben ser evaluados por los tipos de información que pueden proporcionar.

Cuando son benéficos, los casos atípicos, aunque diferentes a la mayor parte de la muestra, pueden ser indicativos de las características segmento de la población que se llegarían a descubrir en el curso normal del análisis. Por el contrario, los casos atípicos problemáticos no son representativos de la población y están en contra de los objetivos del análisis. Los casos atípicos problemáticos pueden distorsionar seriamente los test estadísticos. Debido a la variabilidad en la evaluación de los casos atípicos, se hace imperativo que el investigador examine los datos en busca de la presencia de los casos atípicos con el fin de averiguar el tipo de influencia que ejercen.

#### **DETECCIÓN DE CASOS ATÍPICOS:**

Los datos atípicos pueden identificarse desde una perspectiva univariante, bivariante o multivariante.

##### *Detección Univariante:*

La perspectiva univariante de identificación de casos atípicos examina la distribución de observaciones, seleccionando como casos atípicos aquellos casos que caigan fuera de los rangos de la distribución. La cuestión principal consiste en el establecimiento de un umbral para la designación como caso atípico. El enfoque típico convierte en primer lugar los valores de los datos en valores estándar, que tienen una media cero y una desviación estándar de uno.

*Detección Bivariante:*

Además de la evaluación univariante, pueden evaluarse conjuntamente pares de variables mediante un gráfico de dispersión. Casos que caigan manifiestamente fuera del rango del resto de las observaciones pueden identificarse como puntos aislados en el gráfico de dispersión. Para ayudar a determinar el rango esperado de las observaciones, se puede superponer sobre el gráfico de dispersión una elipse que represente un intervalo de confianza especificado (variando entre 50 y 90 por ciento de la distribución) para una distribución normal bivariante. Esto proporciona una representación gráfica de los límites de confianza y facilita la identificación de casos atípicos.

*Detección Multivariante:*

Dado que la mayoría de los análisis multivariantes tienen más de dos variables, el investigador necesita una forma de medición objetiva de la posición multidimensional de cada observación relativa a un punto común. La medida  $D^2$  de Mahalanobis puede usarse con este fin. La  $D^2$  de Mahalanobis es una medida de la distancia de cada observación en un espacio multidimensional respecto del centro medio de las observaciones. Debido a que proporciona una medida común de centralidad multidimensional, también tiene propiedades estadísticas que tienen en cuenta las pruebas de significación. Dada la naturaleza de los test estadísticos, se sugiere que se use un nivel muy conservador, quizá 0.001, como valor umbral para la designación como caso atípico.

Una forma de resumir la variabilidad de un conjunto de variables es mediante la traza de su matriz de varianzas y covarianzas. Se define la varianza total de los datos por:

$$T = tr(S) = \sum_{i=1}^p s_i^2 \dots\dots\dots (4)$$

Y la varianza media por:

$$\bar{s}^2 = \frac{1}{p} \sum_{i=1}^p s_i^2 \dots\dots\dots (5)$$

El inconveniente de estas medidas es no tener en cuenta la estructura de dependencia entre las variables.

Una medida mejor de la variabilidad global es la varianza generalizada, concepto debido a Wilks. Se define como el determinante de la matriz de varianzas y covarianzas, es decir:

$$VG = |S| \dots\dots\dots (6)$$

Su raíz cuadrada se denomina desviación típica generalizada, y tiene las propiedades siguientes:

- 1.5. Está bien definida, ya que el determinante de la matriz de varianzas y covarianzas es siempre no negativo.
- 1.6. Es una medida del área (para p=2), volumen (para p=3) o hipervolumen (para p>3) ocupado por el conjunto de datos.

Un procedimiento alternativo para estudiar la variabilidad de las observaciones es utilizar el concepto de distancias entre puntos. En el caso escalar, la distancia entre el valor de una variable x en un punto, x<sub>i</sub>, y la media

de la variable,  $\bar{x}$ , se puede medir mediante  $\sqrt{(x_i - \bar{x})^2}$ , alternativamente, por el

valor absoluto de la diferencia,  $|x_i - \bar{x}|$ . La varianza es un promedio de estas

distancias al cuadrado entre los puntos y su media.

**La Distancia de Mahalanobis:**

Se define la distancia de Mahalanobis entre un punto y su vector de medias por:

$$d_i = \left[ \left( x_i - \bar{x} \right)' S^{-1} \left( x_i - \bar{x} \right) \right]^{1/2} \dots\dots\dots (7)$$

Es frecuente referirse al valor  $d_i^2$  también como la distancia de Mahalanobis, en lugar de cómo cuadrado de la distancia, la cual es una medida muy razonable de distancias entre variables correlacionadas.

**1.5. DISTRIBUCIONES MULTIVARIANTES**

El problema central en el análisis de datos es decir si las propiedades encontradas en una muestra pueden generalizarse a la población de la que proviene. Para poder realizar esta extrapolación necesitamos construir un modelo del sistema generador de los datos, es decir, suponer una distribución de probabilidad para la variable aleatoria en la población.

**1.5.1. DISTRIBUCION DE WISHART**

La distribución de Wishart se utiliza para representar la distribución muestral de las matrices de covarianzas en muestras de variables normales multivariantes. En el caso escalar, la distribución que representa esta incertidumbre es la ji-cuadrado de Pearson,  $\chi_i^2$  y la distribución de Wishart estándar puede considerarse como una generalización multivariante de esta distribución.

Recordemos los resultados univariantes: Si  $(I_1, \dots, I_m)$  es un conjunto de variables aleatorias normales independientes  $N(0, \sigma^2)$ , la suma estandarizada de sus cuadrados,  $\sigma^{-2} \sum_{i=1}^m x_i^2$ , sigue una distribución  $\chi_m^2$ . También decimos que

$w = \sum_{i=1}^m x_i^2$ , sigue una distribución  $\sigma^2 x_m^2$ . La densidad de una distribución  $x_m^2$  es un caso particular de la Gamma con parámetros  $\left(\frac{1}{2}, \frac{m}{2}\right)$  y tiene función de

densidad dada por:

$$f(x^2) = k(x^2) (x^2)^{\frac{m}{2}-1} \exp. \left\{ -\frac{1}{2} x^2 \right\} \dots\dots (8)$$

Donde k es una constante. Por otro lado, la distribución de la variable  $w = \sum_{i=1}^m x_i^2$  será la Gamma con parámetros  $\left(\frac{1}{2\sigma^2}\right) \cdot \frac{m}{2}$ , y su densidad tendrá la forma:

$$f(w) = k(\sigma^2)^{\frac{m}{2}} w^{\frac{m}{2}-1} \exp. \left\{ -\frac{1}{2} \sigma^{-2} w \right\} \dots\dots (9)$$

Considerando ahora un conjunto de “m” vectores aleatorios,  $(x_1, \dots, x_m)$ , de dimensión “p” con la misma distribución  $N_p(0,1)$ . La estimación de su matriz de varianzas y covarianzas se obtendrá de  $\sum_{i=1}^m x_i x_i' / m$ , y el numerador de esta expresión

$$W = \sum_{i=1}^m x_i x_i' \dots\dots\dots (10)$$

que es una matriz cuadrada  $p \times p$ , simétrica y definida positiva, decimos que sigue una distribución Wishart con “m” grados de libertad. Esta afirmación debe

interpretarse en el sentido de que la distribución conjunta de los  $\frac{1}{2}p(p+1)$  elementos distintos de  $W$  es:

$$f(w_{11}, \dots, w_{pp}) = c |W|^{(m-p-1)/2} \exp\left\{-\frac{1}{2}tr(W)\right\} \dots\dots\dots (11)$$

Donde "c" es una constante de normalización para que la integral de la función = 1.

Observemos que para  $p = 1$  se obtiene la ecuación 8. Escribiremos  $W \sim W_p(m)$ , donde "p" indica que se trata de la distribución de los elementos de una matriz cuadrada y simétrica de orden  $p$ , y  $m$  son los grados de libertad. Observemos que esta distribución depende únicamente de las dos medidas escalares del tamaño de la matriz: la traza y el determinante. Por tanto, todas las combinaciones de elementos de la matriz que conduzcan a los mismos valores de estas medidas de tamaño tienen la misma probabilidad.

Consideremos, seguidamente,  $m$  vectores aleatorios  $(x_1, \dots, x_m)$  de una distribución  $N_p(0, \Sigma)$ , donde hemos utilizado el símbolo  $\Sigma$  en lugar de  $V$  (matriz de varianzas y covarianzas o simplemente matriz de covarianzas) para representar la matriz de covarianzas. La distribución de los elementos de la matriz

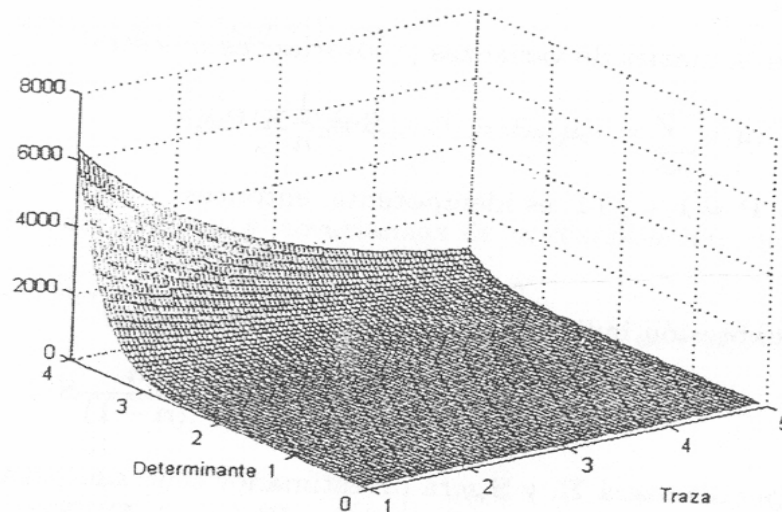
$$W = \sum_{i=1}^m x_i x_i' \dots\dots\dots (12)$$

es la distribución Wishart con  $m$  grados de libertad y matriz de parámetros  $\Sigma$ , dada por

$$f(w_{11}, \dots, w_{pp}) = c / |W|^{(m-p-1)/2} \exp\left\{-\frac{1}{2}tr \Sigma^{-1} W\right\} \dots\dots\dots (13)$$

En general, si una matriz cuadrada y simétrica sigue la distribución (ecuación 9), donde  $\Sigma$  es una matriz simétrica ( $p \times p$ ) no singular definida positiva de componentes constantes, diremos que sigue la distribución Wishart con  $m$  grados de libertad y matriz de parámetros  $\Sigma$ , y escribiremos  $\mathbf{W} \sim W_p(m, \Sigma)$ . Observemos que para  $p = 1$ , esta expresión se reduce a (ecuación 9), y si hacemos  $\Sigma = I$ , la densidad (ecuación 13) se reduce a ecuación 11, y  $W_p(m, I) = W_p(m)$ .

La Figura 5 presenta un ejemplo de esta distribución.



*Figura N° 5*

### 1.5.2. LA $T^2$ DE HOTELLING

Si  $x$  es un vector aleatorio con distribución  $N_p(\mu, V)$ , la variable  $(x - \mu)'V^{-1}(x - \mu)$  es una  $\chi^2$  con  $p$  grados de libertad. Si Sustituimos  $V$  por su estimación  $\hat{S}$ , la matriz de variabilidad muestral dividida por  $n - 1$ , la distribución que se obtiene se denomina  $T^2$  de Hotelling.

En general, si  $x$  se distribuye  $N_p(\mu, V)$  y  $(n-1)\hat{S}$  se distribuye  $W_p(n-1, V)$ , la distribución de la variable escalar:

$$T^2 = (\mathbf{x} - \boldsymbol{\mu})' \hat{S}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \dots\dots\dots (14)$$

que representa la distancia de Mahalanobis entre una variable y su media poblacional, pero calculada con la matriz de covarianza estimada, se denomina distribución  $T^2$  de Hotelling con  $p$  y  $n - 1$  grados de libertad. Diremos que  $T^2 \sim T^2(p, n-1)$ .

Asintóticamente, como  $\hat{S} \rightarrow V$ ,  $T^2$  converge a la distancia de Mahalanobis y la distribución de Hotelling a la distribución  $\chi_p^2$ . Por tanto, para  $n$  grande, la distribución de Hotelling es muy similar a una  $\chi_p^2$ , Para tamaños muestrales mas pequeños tiene una mayor variabilidad que la  $\chi_p^2$ , como consecuencia de la mayor incertidumbre al utilizar la matriz estimada,  $\hat{S}$ , en lugar de la matriz de covarianzas poblacional,  $V$ .

Si  $\mathbf{x}$  es la medida muestral, como  $\bar{x} \sim N_p\left(\mu, \frac{1}{n}V\right)$ , la distribución de

$$(\bar{x} - \boldsymbol{\mu})' \left(\frac{\hat{S}}{n}\right)^{-1} (\bar{x} - \boldsymbol{\mu}) = n(\bar{x} - \boldsymbol{\mu})' \hat{S}^{-1} (\bar{x} - \boldsymbol{\mu}) \dots\dots\dots (15)$$

es también una  $T^2$  de Hotelling. Observemos que si  $p = 1$ , la  $T^2$  se reduce a:

$$T^2 = \frac{n(\bar{x} - \mu)^2}{\hat{S}^2} = t^2 \dots\dots\dots (16)$$

Y coincide con el estadístico  $t$  de Student. Por tanto  $T^2(1, m) = t_m^2$

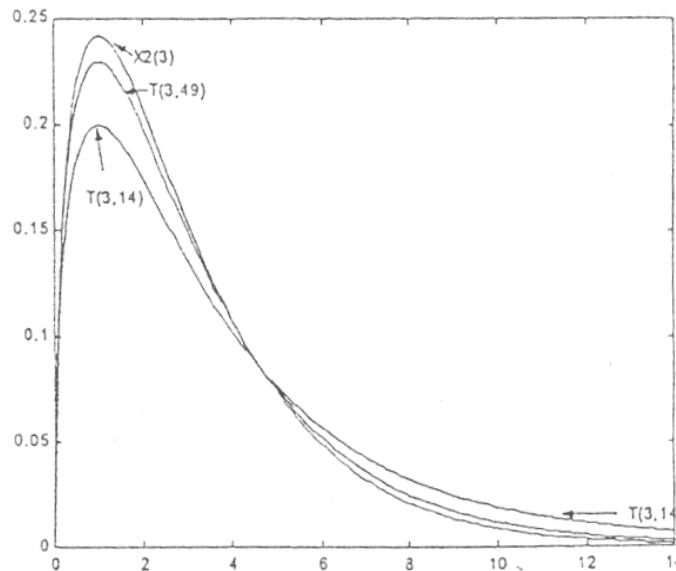
La distribución de Hotelling no se tabula, ya que con una simple transformación se reduce a la distribución  $F$  del análisis de la varianza de Fisher. Se demuestra que:



$$F_{p,n-p} = \frac{n-p}{p(n-1)} T^2(p,n-1) \quad (17)$$

lo que permite calcular la distribución de  $T^2$  en función de las tablas de la distribución F.

Este resultado es consistente con la convergencia de la  $T^2$  a la  $\chi^2$  ya que implica que asintóticamente,  $p F_{p,n-p}$  tiene a una distribución  $\chi_p^2$ . La Figura 6 muestra un ejemplo de la distribución de Hotelling comparada con la  $\chi_p^2$ . Vemos que para tamaño muestral muy pequeño,  $n = 15$ , las colas de la distribución son mas pesadas que las de la ji- cuadrado indicando la mayor incertidumbre existente, pero para  $n = 50$  ambas son ya muy similares. La aproximación depende del cociente  $n/p$ , y si este es grande mayor de 15, podemos aproximar bien la distribución de la Hotelling mediante la ji-cuadrada.



La distribución de Hotelling para dos valores del tamaño muestra y la distribución ji-cuadrada.

**Figura N° 6**

## **1.6. CONTROL ESTADÍSTICO DE PROCESOS CONCEPTOS FUNDAMENTALES**

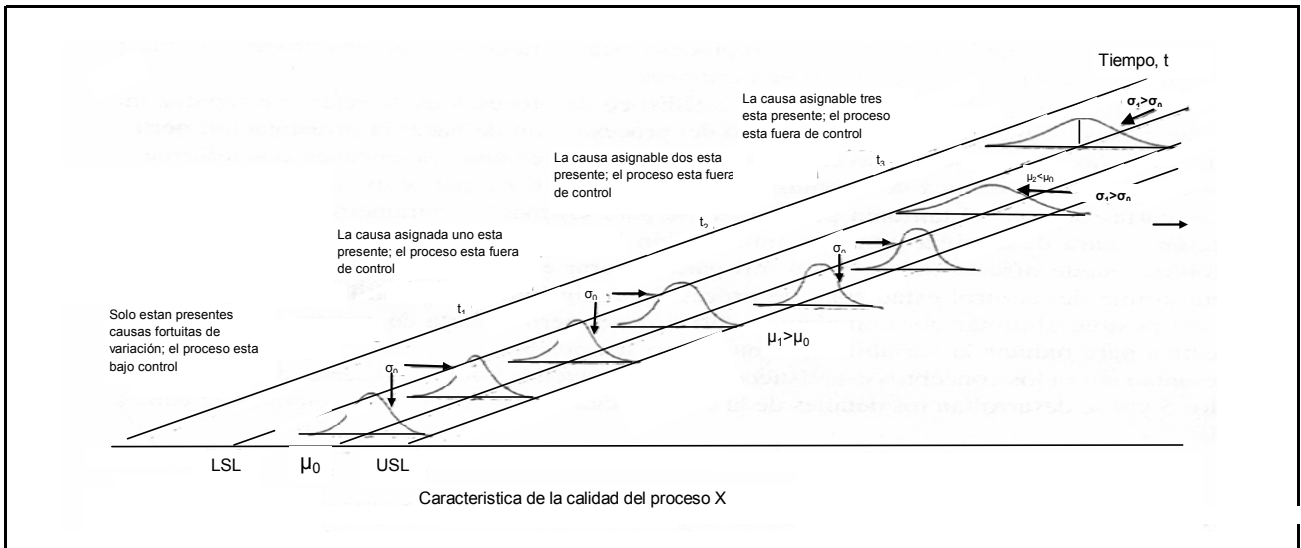
El control estadístico de Procesos (SPC, por sus siglas en inglés) en línea es una herramienta básica para alcanzar el objetivo de un continuo desempeño del proceso y reducir la variabilidad de los parámetros clave. Las cartas de control constituyen el procedimiento de control estadístico de procesos en línea del tipo más simple. Dichas cartas de control se concentran principalmente en la carta de control propuesta por el Doctor Walter A. Shewhart, llamada la carta de control de Shewhart (Juran. Quality Planning and Análisis, 2da edición McGraw-Hill)

En cualquier proceso de producción, independiente de lo adecuado que sea su diseño o de la atención que se preste a su mantenimiento, siempre existirá cierta cantidad de variabilidad inherente o natural. Esta variabilidad natural, o “ruido de fondo”, es el efecto acumulado de muchas causas pequeñas y en esencia inevitables. En el contexto del control estadístico de calidad, a esta variabilidad natural se le denomina un “sistema estable de causas fortuitas”. Se dice que un proceso que opera únicamente con causas fortuitas de variación esta bajo control estadístico. En otras palabras las causas fortuitas son una parte inherente del proceso.

En ocasiones puede estar presente otra clase de variabilidad en la salida de un proceso. Esta variabilidad en las características clave de la calidad se origina de tres fuentes:

- Máquinas ajustadas o controladas incorrectamente.
- Errores del operador.
- Materia Prima defectuosa.

A estas fuentes de variabilidad que no son parte del patrón de las causas fortuitas se les llama “causas asignables”. Se dice que un proceso que opera en presencia de causas asignables esta fuera de control.



**Figura N° 7**

En la figura 7 se ilustran las causas fortuitas y asignables de la variación. Hasta el tiempo  $t_1$  el proceso que se muestra en esta figura esta bajo control; es decir, solo están presentes causas de variación fortuita. Como resultado, tanto la media como la desviación estándar del proceso están en sus valores bajo control (por ejemplo  $\mu_0$  y  $\sigma_0$ ). En el tiempo  $t_1$  ocurre una causa asignable. Como se muestra en la figura 7, el efecto de esta causa asignable es correr la medida del proceso a un nuevo valor  $\mu_1 > \mu_0$ . En tiempo  $t_2$  ocurre otra causa asignable, de la que resulta  $\mu = \mu_0$ , pero ahora la desviación estándar del proceso se ha ocurrido a un valor mas grande  $\sigma_1 > \sigma_0$ . En el tiempo  $t_3$  se presenta otra causas asignable, cuyo resultado es que tanto la media como la desviación estándar toman valores fuera de control. Del tiempo  $t_1$  en adelante, la presencia de causas asignables ha dado como resultado un proceso fuera de control.

Con frecuencia los procesos de producción operaran en el estado bajo control, produciendo productos aceptables durante periodo relativamente prolongados. Con el tiempo, sin embargo, ocurrirán causas asignables, aparentemente al azar, que ocasionaran un “corrimiento” a un estado fuera de control en el que una proporción mayor de la salida del proceso no cumplirá con los requerimientos. Por ejemplo, en la figura 7 se observa que cuando el proceso esta bajo control, la mayor parte de la producción estar entre los limites inferior y superior (LSL y USL, respectivamente, por sus siglas en ingles) de la especificación. Cuando el proceso esta fuera de control, una proporción más alta del mismo se localiza fuera de estas especificaciones.

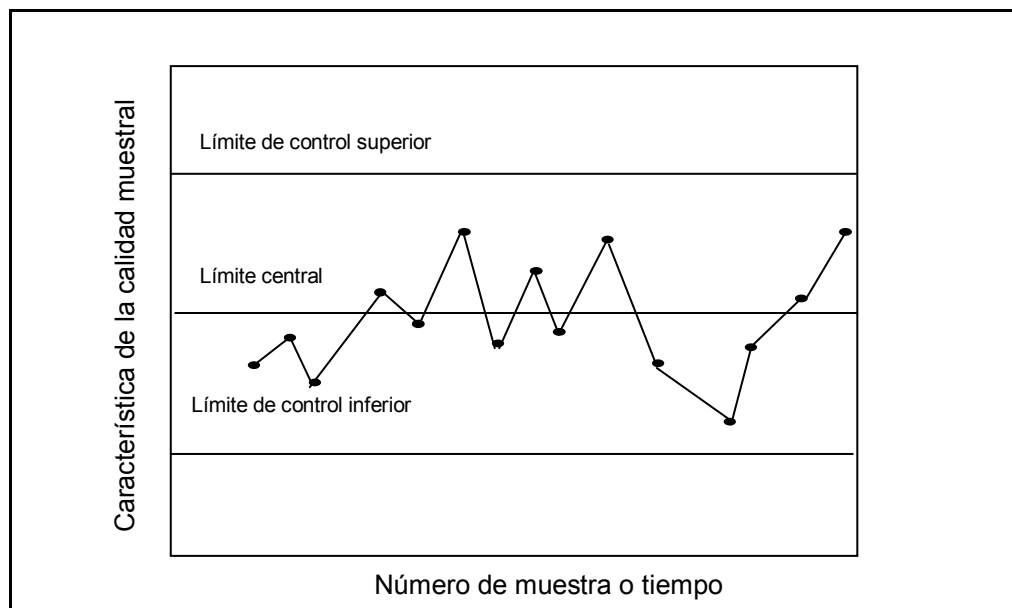
Uno de los objetivos principales del control estadístico de procesos es detectar con rapidez la ocurrencia de causas asignables en el corrimiento del proceso a fin de hacer la investigación pertinente y emprender las acciones correctivas antes de que se fabriquen muchas unidades disconformes.

La carta de control es una técnica del movimiento de procesos en línea que se usa ampliamente para este fin. Las cartas de control también pueden usarse para estimar los parámetros de un proceso de producción y para determinar con esta información la capacidad del proceso. Asimismo, la carta de control puede ofrecer información útil para mejorar el proceso. Por ultimo, recuérdese que la meta última del control estadístico de procesos es eliminar la variabilidad del proceso. Quizás no sea para reducir la variabilidad tanto como sea posible.

### **1.6.1. GRÁFICOS DE CONTROL**

En la figura 8, se muestra una carta de control típica, que es la representación gráfica de una característica de la calidad que se ha medido o calculado a partir de una muestra contra el número de muestra o tiempo. La carta contiene una línea central que representa el valor promedio de la característica de la calidad que corresponde al estado bajo control. También se muestra en la carta otras dos líneas horizontales, llamadas el Límite de Control

Superior (UCL, por sus siglas en ingles) y el límite de control inferior (LCL, por sus siglas en ingles). Estos límites de control se eligen de tal modo que si el proceso esta bajo control, casi todos los puntos muestrales se localizaran entre ellos. En tanto los puntos graficados se localicen dentro de los limites de control, se supone que el proceso esta bajo control y no es necesaria ninguna acción. Sin embargo, un punto que se localiza fuera de los límites de control se interpreta como evidencia de que el proceso esta fuera de control, y se requiere investigación y acción correctiva para encontrar y eliminar la causa o causas asignables responsables de este comportamiento.



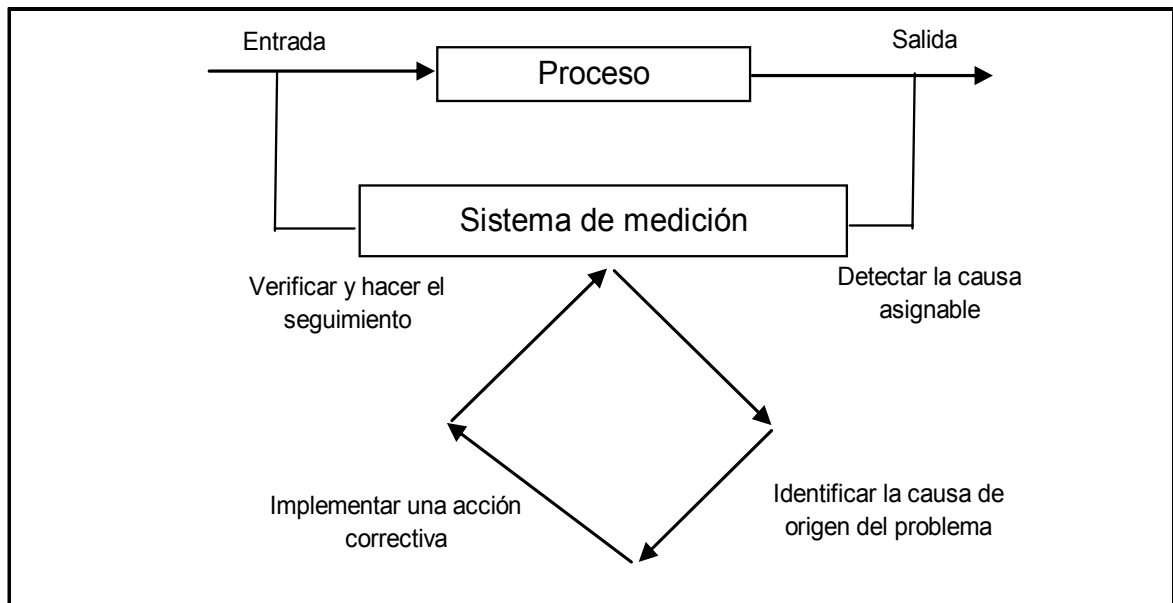
**Figura N° 8: Gráfico de Control**

Puede darse un modelo general para una carta de control. Sea  $\omega$  un estadístico muestral que mide alguna característica de la calidad de interés, y suponer que la media de  $\omega$  es  $\mu$ , y que la desviación estándar de  $\omega$  es  $\sigma_w$ . Entonces la línea central, el límite de control superior (UCL) y el límite de control inferior (LCL) son:

$$\begin{aligned}
 UCL &= \mu_w + L\sigma \\
 \text{LineaCentral} &= \mu_w \quad \dots\dots\dots (18) \\
 LCL &= \mu_w - L\sigma
 \end{aligned}$$

Donde L es la “distancia” de los limites de control a la línea central, expresada en unidades de desviación estándar.

La carta de control es un recurso para describir de manera precisa lo que se pretendió exactamente por medio del control estadístico: como tal, puede usarse en una variedad de formas. En muchas aplicaciones se usa para la vigilancia en línea de un proceso. Es decir, se colectan datos muestrales y se usan para construir la carta de control, y si los valores muestrales de  $\bar{x}$  se localizan dentro de los limites de control y no muestran ningún patrón sistemático, se dice que el proceso esta bajo control en el nivel indicado por la carta.



*Figura N° 9*

## **CARTAS DE CONTROL PARA VARIABLES:**

A una característica de la calidad que se mide en una escala numérica se le llama variable, por ejemplo dimensiones tales como el largo y el ancho, la temperatura y el volumen.

Cuando se trata con una característica de la calidad que es una variable, por lo general es necesario monitorear tanto el valor medio de la característica de la calidad como su variabilidad. El control del promedio del proceso, o nivel de calidad medio, suele hacerse con la carta de control desviación estándar, llamada carta S, o bien con una carta de control para el rango, llamada carta R.

La carta R se usa con mucha frecuencia. Generalmente, se llevan cartas  $\bar{x}$  y R separadas para cada característica de la calidad de interés. Las cartas  $\bar{x}$  y R (o S) se encuentran entre las técnicas estadísticas de monitoreo y control de procesos en línea mas importantes y útiles.

Es importante mantener bajo control tanto la media del proceso como la variabilidad del proceso. En la figura 10 se ilustra la salida de un proceso de producción. En la figura 10(a) la media  $\mu$  y la desviación estándar  $\sigma$  están bajo control en sus valores nominales (por ejemplo  $\mu_0$  y  $\sigma_0$ ); por consiguiente, la mayor parte de la salida del proceso se localiza dentro de los límites de la especificación. Sin embargo, en la figura 10(b) la media se ha corrido a un valor  $\mu_1 > \mu_0$ , dando como resultado una fracción mayor de productos disconformes. En la figura 10(c) la desviación estándar del proceso se ha corrido a un valor  $\sigma_1 > \sigma_0$ . Esto también resulta en una porción caída del proceso más alta, aun cuando la media del proceso se encuentra aun en el valor nominal.

### **1.6.1.1. CARTAS DE CONTROL PARA $\bar{x}$ y R**

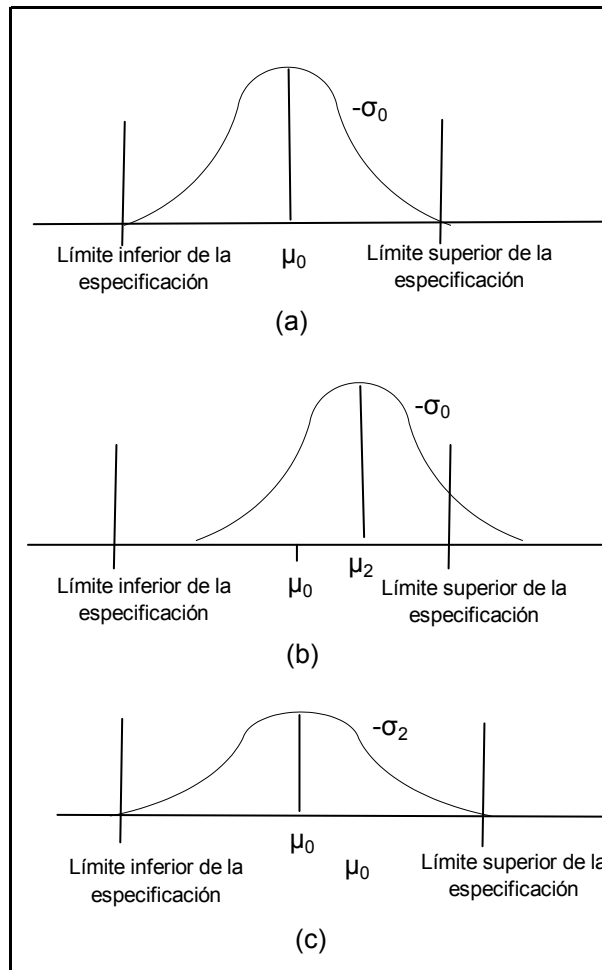
Suponer que una característica de la calidad tiene una distribución normal con medida  $\mu$  y desviación estándar  $\sigma$ . Donde tanto  $\mu$  como  $\sigma$  son conocidas. si

$x_1, x_2, \dots, x_n$  es una muestra de tamaño  $n$ , entonces el promedio de esta muestra es:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \dots\dots\dots (19)$$

y se sabe que  $\bar{x}$  sigue una distribución normal con media  $\mu$  y desviación  $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ . Además, la probabilidad es de  $1 - \alpha$  para que cualquier medida muestral se localice entre:

$$\mu - Z_{\alpha/2} \sigma_{\bar{x}} \leq \bar{x} \leq \mu + Z_{\alpha/2} \sigma_{\bar{x}} \dots\dots\dots (20)$$



**Figura N° 10**



La necesidad de controlar tanto la medida como la variabilidad del proceso, a) La medida y la desviación estándar en sus valores nominales. b) La medida del proceso  $\mu_1 > \mu_{0.c}$ , c) La desviación estándar del proceso  $\sigma_1 > \sigma_{0.c}$  y

$$\mu - Z_{\alpha/2} \sigma_{\bar{x}} = \mu - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \dots\dots\dots (21)$$

Por lo tanto, si  $\mu$  y  $\sigma$  son conocidas, las ecuaciones 20 y 21 podrían usarse como límites de control superior e inferior en una carta de control para las medidas muestrales. Se acostumbra sustituir  $Z_{\alpha/2}$  con 3, a fin de trabajar con límites tres sigmas. Si la media de una muestra se localiza fuera de estos límites, se trata de un indicio de que la media del proceso ha dejado de ser igual a  $\mu$ .

Se ha supuesto que la distribución de las características de la calidad es normal. Sin embargo los resultados anteriores siguen siendo aproximadamente validos cuando incluso cuando la distribución fundamental no es normal, debido al teorema de límite central.

En la practica, generalmente no se conocen los calores de  $\mu$  y  $\sigma$ . Por lo tanto, deben estimarse a partir de muestras o subgrupos preliminares tomados cuando se considera que el proceso esta bajo control. En general, estas estimaciones deberán basarse en al menos 20 o 25 muestras. Suponer que se cuenta con “m” muestras, cada una de las cuales contienen, observaciones de la características de la calidad. De manera típica “n” será pequeña, con frecuencia ya sea 4, 5 o 6. Estos tamaños pequeños de la muestra suelen resultar de la construcción de subgrupos racionales y del hecho de que los costos de muestreo e inspección asociados con la medición de las variables por lo general son relativamente altos. Sean  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$  los promedios de cada muestra. Entonces el mejor estimador de  $\mu$ , el promedio del proceso, el gran promedio: por ejemplo:

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m} \dots\dots\dots(22)$$

Por tanto,  $\bar{x}$  se usara como la línea central (CL, por sus siglas en ingles) en la carta.

Para construir los límites de control es necesaria una estimación de la desviación estándar  $\sigma$ . Este valor de  $\sigma$  puede estimarse sea partir de las desviaciones estándar o bien por lo rangos de las “m” muestras.

Si  $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n$  es una muestra de tamaño n, entonces el rango de la muestra es la diferencia entre las observaciones menor y mayor es decir:

$$R = x_{m\acute{a}x} - x_{m\acute{i}n} \dots\dots\dots (23)$$

Sean:

$R_1 + R_2 + \dots + R_m$  Los rangos de las m muestras. El rango promedio es:

$$\bar{R} \equiv \frac{R_1 + R_2 + \dots + R_m}{m} \dots\dots\dots(24)$$

Ahora pueden darse las fórmulas para construir los límites de control de la carta  $\bar{x}$ . Estos son los siguientes:

Límites de control de la carta  $\bar{x}$

$$\begin{aligned} \text{UCL} &= \bar{x} + A_2 \bar{R} \\ \text{Línea central} &= \bar{x} \\ \text{LCL} &= \bar{x} - A_2 \bar{R} \end{aligned} \quad (25)$$

La variabilidad del proceso puede monitorearse graficando los valores del rango muestral R en una carta de control. La línea y los límites de control de la carta R con los siguientes:

<u>Límites de control de la carta R</u>	
$UCL = D_4 \bar{R}$	
$Línea\ central = \bar{R}$	(26)
$LCL = D_3 \bar{R}$	

#### 1.6.1.2. CARTAS DE CONTROL PARA $\bar{x}$ y S

Aun cuando es muy común la utilización de las cartas  $\bar{x}$  y R, en ocasiones es deseable estimar la desviación estándar del proceso directamente en vez de indirectamente mediante el uso del rango R. Esto lleva a las cartas de control para x y S, donde S es la desviación estándar muestral. En general, las cartas  $\bar{x}$  y S son preferibles a sus contrapartes mas familiares, las cartas  $\bar{x}$  y R, cuando

1. El tamaño de la muestra n es moderadamente grande – por ejemplo,  $n > 10$  o 12 (recuérdese que el método del rango para estimar  $\sigma$  pierde eficiencia estadísticas para muestras de moderadas a grandes).
2. El tamaño de la muestra n es variable.

Si  $\sigma^2$  es la varianza desconocida de la distribución de probabilidad, entonces un estimador insesgado de  $\sigma^2$  es la varianza m.

Sin embargo, la desviación estándar muestral S no es un estimador insesgado de  $\sigma$ . Si la distribución fundamental es normal, entonces S es en realidad una estimación de  $c_4\sigma$ , donde  $c_4$  es una constante que depende del

tamaño de la muestra  $n$ . Además, la desviación estándar de  $S$  es  $\sigma\sqrt{1-c_4^2}$ . Esta información puede utilizar para establecer las cartas de control de  $\bar{x}$  y  $S$ . Considérese el caso en que un valor estándar esta dado por  $\sigma$ . Puesto que  $E(S)=c_4\sigma$ , entonces la línea central de la carta es  $c_4\sigma$ . Los límites de control tres sigma para  $S$  son entonces:

$$UCL = c_4\sigma + 3\sigma\sqrt{1-c_4^2} \dots\dots\dots (27)$$

$$LCL = c_4\sigma - 3\sigma\sqrt{1-c_4^2} \dots\dots\dots (28)$$

Se acostumbra definir las dos constantes

$$B_5 = c_4 - 3\sqrt{1-c_4^2} \dots\dots\dots (29)$$

Y

$$B_6 = c_4 + 3\sqrt{1-c_4^2} \dots\dots\dots (30)$$

Por consiguiente, los parámetros de la carta  $S$  con un valor estándar para  $\sigma$  dado. Se vuelven:

$UCL = B_6 \sigma$ $Línea\ central = c_4 \sigma \quad (31)$ $LCL = B_5 \sigma$
--

En el anexo A se muestran los valores de  $B_5$  y  $B_6$  para varios tamaños de la muestra. Cuando no hay un valor estándar dado para  $\sigma$ , entonces debe estimarse analizando datos pasados. Suponer que se cuenta con  $m$  muestras

preliminares, cada una de tamaño  $n$ , y sea  $S$ , la desviación estándar de la  $i$ -ésima muestra. El promedio de las  $m$  desviaciones estándar es:

$$UCL = \bar{S} + 3 \frac{\bar{S}}{C_4} \sqrt{1 - C_4^2} \dots\dots\dots (33)$$

El estadístico  $S/c_4$  es un estimador insesgado de  $\sigma$ . Por lo tanto, los parámetros de la carta  $S$  serían:

$$\text{Línea central} = \bar{S} \dots\dots\dots (34)$$

$$LCL = \bar{S} - 3 \frac{\bar{S}}{C_4} \sqrt{1 - C_4^2} \dots\dots\dots (35)$$

generalmente se definen las constantes

$$B_3 = 1 - \frac{3}{C_4} \sqrt{1 - C_4^2} \dots\dots\dots (36)$$

y

$$B_4 = 1 + \frac{3}{C_4} \sqrt{1 - C_4^2} \dots\dots\dots (37)$$

Por consiguiente, los parámetros de la carta pueden escribirse como

$UCL = B_4 \bar{S}$ $\text{Línea central} = \bar{S} \quad (38)$ $LCL = B_3 \bar{S}$
---

Obsérvese que  $B_4 = B_6/C_1$  y  $B_3 = B_5/C_4$ .

Cuando se usa  $\bar{S}/c_4$  para estimar  $\sigma$  los límites de control de la carta  $\bar{x}$  correspondiente pueden definirse como:

$$UCL = \bar{x} \dots\dots\dots (39)$$

$$\text{Línea central} = \bar{x} \dots\dots\dots (40)$$

$$\text{UCL} = \bar{x} - \frac{3\bar{S}}{C_4\sqrt{n}} \dots\dots\dots (41)$$

Sea la constante  $A_3 = 3 / (C_4\sqrt{n})$ . Entonces los parámetros de la carta  $\bar{x}$  se vuelve

$\begin{aligned} \text{UCL} &= \bar{x} + A_3\bar{S} \\ \text{Línea central} &= \bar{x} \\ \text{LCL} &= \bar{x} - A_3\bar{S} \end{aligned} \dots\dots\dots (42)$
--

En el anexo A se enlistan las constantes  $B_3$ ,  $B_4$  y  $A_3$  para construir cartas  $\bar{x}$  y S a partir de datos pasados para varios tamaños de la muestra. Obsérvese que se ha supuesto que la desviación estándar muestral se define como

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \dots\dots\dots (43)$$

Algunos autores definen S como n en el denominador de la ecuación 43 en vez de utilizar n-1. Cuando es este el caso, las definiciones de las constantes  $C_4$ ,  $B_3$ ,  $B_4$  y  $A_3$  se alteran. Las constantes correspondientes basadas en el uso de N para calcular S se denominaran  $C_2$ ,  $B_1$ ,  $B_2$  y  $A_1$ , respectivamente.

Tradicionalmente, los ingenieros de calidad han preferido la carta R a la S debido a la simplicidad del cálculo de R de cada muestra. La disponibilidad actual de calculadoras manuales con cálculo automático de S y la disponibilidad creciente de microcomputadoras para implementar directamente las cartas de

control en línea en la estación de trabajo han eliminado cualquier dificultad de cálculo.

### **1.6.2. CAPACIDAD DEL PROCESO**

Las técnicas estadísticas pueden ser útiles en el ciclo de un producto, incluyendo las actividades de desarrollo previas a la manufactura, para cuantificar la variabilidad del proceso, para analizar esta variabilidad respecto de los requerimientos o especificaciones del producto y para ayudar al personal de desarrollo y manufactura a eliminar o reducir en gran medida esta variabilidad. A esta actividad general se le llama Análisis de Capacidad del Proceso.

La capacidad del proceso se refiere a la uniformidad del mismo. Evidentemente, la variabilidad del proceso es una medida de la uniformidad de la salida. Hay dos formas de conceptualizar esta variabilidad:

1. La variabilidad natural o inherente en un tiempo especificado; es decir, la variabilidad “instantánea”.
2. La variabilidad con el tiempo.

El análisis de la Capacidad del Proceso se define como el estudio de ingeniería para estimar la capacidad del proceso. La estimación de la capacidad del proceso puede estar en la condición de una distribución de probabilidad que tenga una forma, centro (media) y dispersión (desviación estándar) especificados.

En el estudio de capacidad del proceso por lo general se miden los parámetros funcionales del producto, no el proceso en sí. Cuando el analista puede observar directamente el proceso y puede controlar o monitorear la actividad de colección de datos, el estudio es un verdadero estudio de capacidad del proceso, ya que al controlar la colección de datos y conocer la secuencia en

el tiempo de los datos, es posible hacer inferencias sobre la estabilidad del proceso con el tiempo.

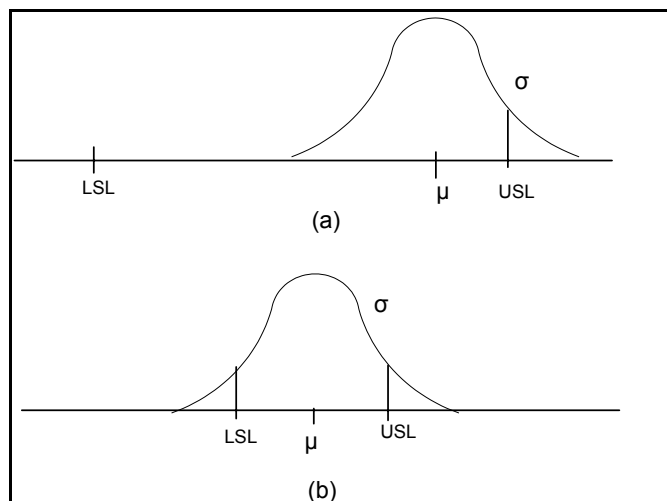
En el análisis de capacidad del proceso se utilizan tres técnicas principales: histogramas o graficas de probabilidad, cartas de control y experimentos diseñados.

### 1.6.2.1. ANALISIS DE CAPACIDAD DEL PROCESO UTILIZANDO UN HISTOGRAMA O UNA GRAFICA DE PROBABILIDAD:

#### Utilizando un Histograma:

El histograma acompañado con el promedio muestral  $\bar{x}$  y la desviación estándar muestral S, proporciona información sobre la capacidad del proceso. Una ventaja de usar el histograma para estimar la capacidad del proceso es que proporciona una impresión visual inmediata del desempeño del proceso.

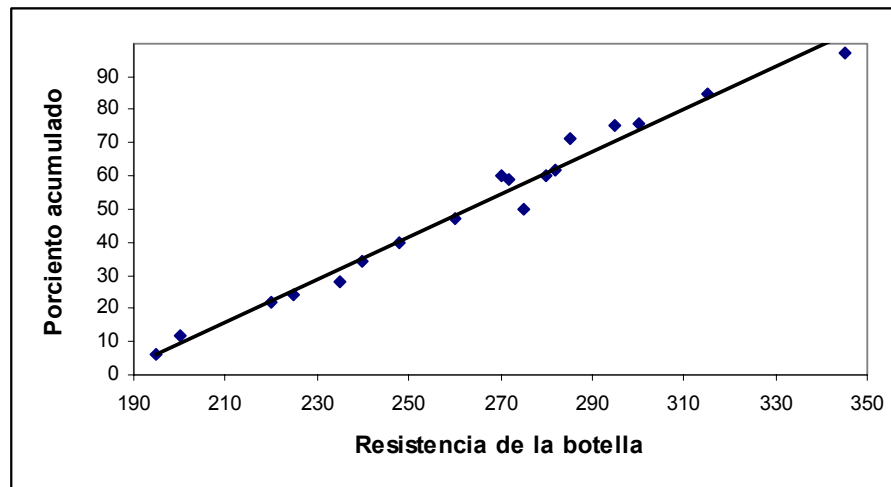
También puede indicar de inmediato la razón de un desempeño pobre del proceso. Por ejemplo en la figura 11 indica un proceso con una capacidad del proceso adecuada, pero el objetivo del proceso tiene una localización deficiente, mientras que la figura 11b indica un proceso con una capacidad pobre que resulta del exceso de variabilidad.





*Algunas razones de la capacidad pobre del proceso a) Centrado del proceso deficiente. b)  
Exceso de variabilidad en el proceso.*

**Figura 11**



*Gráfica de probabilidad normal de los datos de la resistencia al entallamiento de botellas.*

**Figura 12**

### **Graficación de probabilidades:**

La graficación de probabilidades es una alternativa del histograma que puede usarse para determinar la forma, el centro y la dispersión de la distribución. Tiene la ventaja de que no es necesario dividir el rango de la variable en intervalos de clase, y con frecuencia produce resultados razonables para muestras moderadamente pequeñas (lo que el histograma no hará). En general, una gráfica de probabilidad es una grafica de los datos ordenados contra la frecuencia muestral acumulada, con la escala vertical elegida de tal modo que la distribución acumulada del tipo supuesto sea una línea recta.

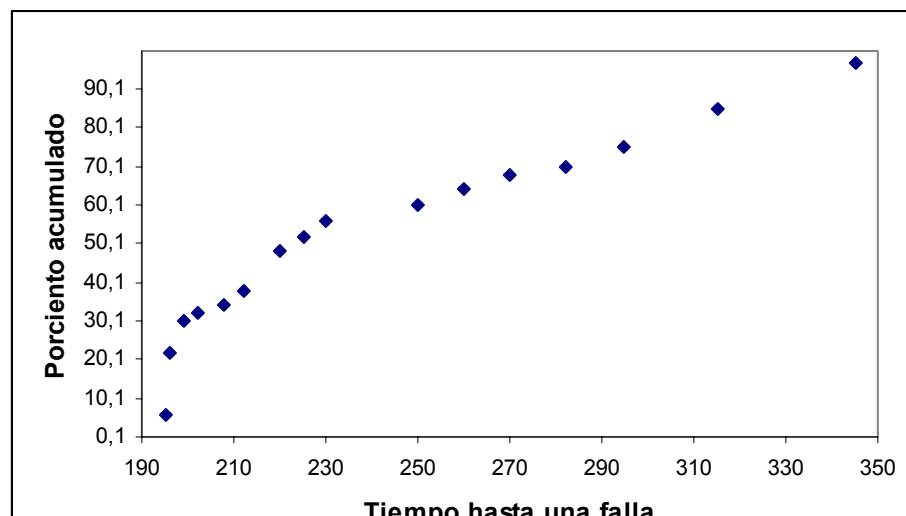
Para ilustrar la utilización de una gráfica de probabilidad normal en un estudio de capacidad del proceso, considérense las 20 observaciones siguientes de la resistencia al entallamiento de las botellas:

197,200,215,221,231,242,245,258,265,265,2371,275,277,278,280,283,290,301, 318 y 346. La figura 12 es la gráfica de probabilidad normal de la resistencia. Obsérvese que los datos se localizan casi sobre una línea recta, lo cual implica que la distribución de la resistencia al entallamiento es normal. La desviación estándar de la distribución es la pendiente de la línea recta. Es conveniente estimar la desviación estándar como la diferencia entre los percentiles octogésimo cuarto y el quincuagésimo. Para los datos de la resistencia, se obtiene:

$$\begin{aligned}\sigma &= 84^{\circ} \text{ percentil} - 50^{\circ} \text{ percentil} \\ &= 298 - 260 \text{ psi} \\ &= 38 \text{ psi}\end{aligned}$$

Obsérvese que  $\mu = 260 \text{ psi}$   $\sigma = 38 \text{ psi}$  no se apartan demasiado del promedio muestral  $\bar{x} = 264.06$  y la desviación estándar muestral  $S = 32.02$ .

La gráfica de probabilidad normal también puede usarse para estimar rendimientos y porciones caídas del proceso. Por ejemplo, la especificación para la resistencia de la botella es  $LSL = 200 \text{ psi}$ .



*Gráfica de probabilidad normal para el tiempo hasta la falla de una válvula.*

***Figura 13***

Por la figura 12 se estimaría que cerca de 5% de las botellas fabricadas con este proceso estallarían debajo de este límite.

Deberá tenerse cuidado al usar las gráficas de probabilidad. Si los datos no provienen de la distribución supuesta, las inferencias sobre la capacidad del proceso que se hagan a partir de la grafica pueden tener serios errores. En la figura 13 se presenta un grafico de probabilidad normal de los tiempos hasta una falla (en obras) de una válvula en una planta química. Al examinar esta grafica, puede verse que la distribución del tiempo hasta una falla no es normal.

Una desventaja evidente de la graficación de probabilidades es que no es un procedimiento objetivo. Existe la posibilidad de que dos análisis lleguen a conclusiones deferentes utilizando los mismos datos. Por esta razón, con frecuencia es deseable complementar las gráficas de probabilidad con pruebas basadas estadísticamente más formales de la bondad del ajuste. Elegir la distribución para ajustar los datos también es un paso muy importante en la graficación de probabilidades. En ocasiones puede usarse el conocimiento personal del fenómeno físico, o la experiencia pasada, para sugerir la elección de la distribución. En otras situaciones, la figura 14 puede ser útil para seleccionar una distribución que describa los datos. Esta figura muestra las regiones en el plano  $\beta_1, \beta_2$  para varias distribuciones de probabilidad estándares, donde  $\beta_1, \beta_2$  son las medidas de sesgo y la curtosis,

respectivamente. Para usar la figura 14, se calculan estimaciones del riesgo y la curtosis de la muestra, por ejemplo:

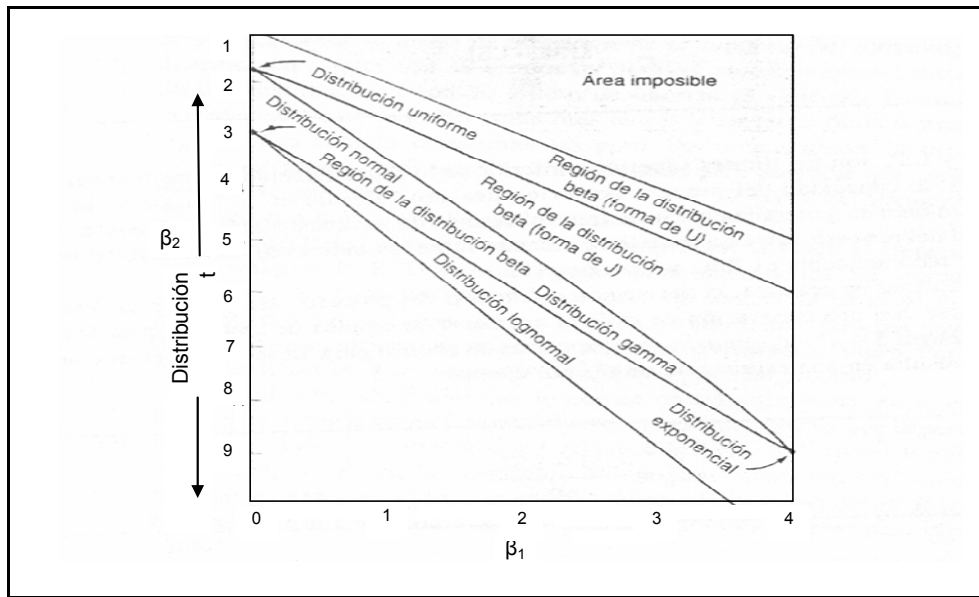
$$\sqrt{\beta_1} = \frac{M_3}{(M_2)^{3/2}} \dots\dots\dots (44)$$

y

$$\beta_2 = \frac{M_4}{M_2^2} \dots\dots\dots (45)$$

Donde

$$M_j = \frac{\sum_{i=1}^n (x_i - \bar{x})^j}{n} \quad j=1,2,3,4 \dots\dots\dots (46)$$



Regiones en el plano  $\beta_1, \beta_2$  para varias distribuciones.

Y se gráfica el **Figura 14**  $(\hat{\beta}_1, \hat{\beta}_2)$  en la gráfica. Si el punto graficado se localiza cerca de un punto, recta o región que corresponde a una de las distribuciones de la figura, entonces esta distribución es una elección lógica para usarla como modelo de los datos. Si el punto se localiza en regiones del plano  $\beta_1, \beta_2$  donde ninguna de las distribuciones parece adecuada, piden requerirse otras distribuciones de probabilidad más generales.

### 1.6.2.2. ÍNDICES DE CAPACIDAD DEL PROCESO:

#### a. Uso e interpretación Cp:

Con frecuencias es conveniente contar con una forma cuantitativa simple para expresar la capacidad del proceso. Una manera de hacerlo es con el índice de capacidad del proceso (PCR, por sus siglas en inglés)  $C_p$ :

$$C_p = \frac{USL - LSL}{6\sigma} \quad (47)$$

Donde USL y LSL son los límites superior e inferior de la especificación, respectivamente. El  $C_p$  y otros índices de capacidad del proceso tienen un uso generalizando en la industria. Su utilización incorrecta también es generalizada. Señalaremos algunos de los abusos más comunes de los índices de capacidad del proceso.

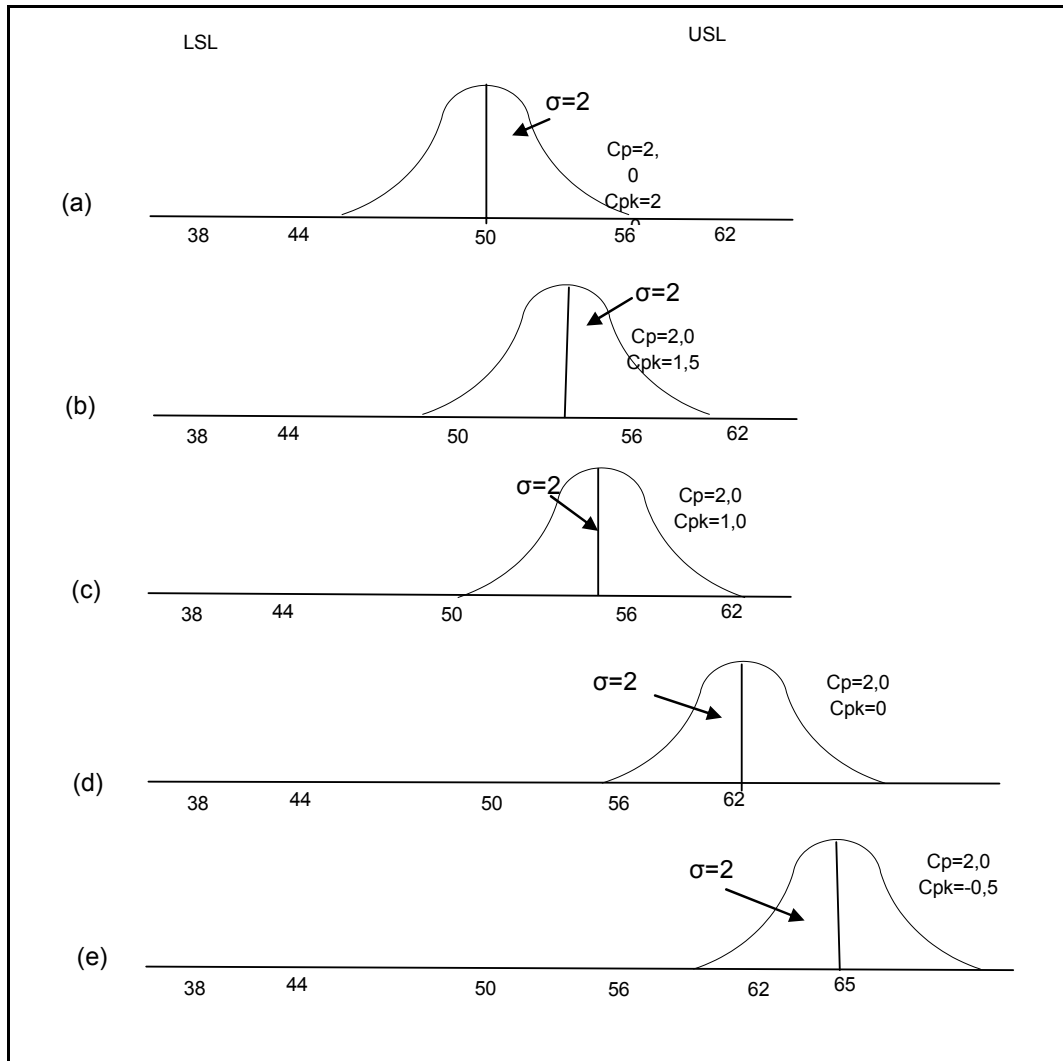
En una aplicación práctica la desviación estándar  $\sigma$  del proceso casi siempre es desconocida y debe sustituirse con una estimación de  $\sigma$ . Para estimar  $\sigma$  se emplea de manera típica la desviación estándar muestral  $S$  o  $R/d_2$  (cuando se usan cartas de control para variables en el estudio de capacidad). Esto resulta en una estimación de  $C_p$ : por ejemplo

$$\hat{C}_p = \frac{USL - LSL}{6\sigma} \quad (48)$$

#### b. Índice de capacidad del proceso para un proceso fuera del centro.

El índice de capacidad del proceso  $C_p$  no toma en consideración donde se localiza la media del proceso respecto a las especificaciones. El  $C_p$  simplemente mide la extensión de las especificaciones en comparación de la dispersión seis sigma del proceso. Por ejemplo las distribuciones normales de la parte superior de la figura 15 tiene  $C_p = 2.0$ , pero el proceso del panel b) de la figura tienen

claramente una capacidad mas baja que el proceso del panel a), debido a que no esta operando en el punto medio del intervalo entre las especificaciones.



Relación de  $C_p$  y  $C_{pk}$   
**Figura 15**

Esta situación puede reflejarse con mayor precisión con mayor índice de capacidad del proceso que tome en consideración el centrado del proceso. Esta cantidad es

$$C_{pk} = \min. (C_{pU}, C_{pL}) \quad (49)$$

Obsérvese que  $C_{pk}$  es simplemente el PCR unilateral para el límite de las especificaciones más próximo al promedio del proceso. Para el proceso ilustrado en la figura 15b, se tendría

$$\begin{aligned} C_{pk} &= \min (C_{pu}, C_{pl}) \\ &= \min \left[ C_{pu} = \frac{62-53}{3(2)} = 1,5, C_{pl} = \frac{53-38}{3(2)} \right] \\ &= 1.5 \end{aligned}$$

En general, si  $C_p, C_{pk}$ , el proceso esta centrado en el punto medio de las especificaciones y cuando  $C_{pk} < C_p$  el proceso esta descentrado.

La magnitud de  $C_{pk}$  respecto a  $C_p$  es una medida directa de que tan apartado del centro esta operando el proceso. Varios casos que es común encontrar se ilustran en la figura 15. Obsérvese que en el panel (c) de la figura 15  $C_{pk} = 1.0$  mientras que  $C_p = 2.0$ .

Por tanto es común decir que  $C_p$  mide la capacidad potencial del proceso, mientras que  $C_{pk}$  mide la capacidad real.

En el panel d) de la figura 15 se ilustra el caso en que la media del proceso es exactamente uno de los límites de la especificación, llevando a  $C_{pk} = 0$ . Como se ilustra en el panel c), cuando  $C_{pk} = < 0$  la implicación es que la a media del proceso se localiza fuera de las especificaciones evidentemente,  $C_{pk} < -1$ . El proceso completo se localiza fuera de los límites de la especificación. Algunos factores definen  $C_{pk}$  como un valor no negativo, por lo que los valores menores a cero se definen como cero.

Muchas autoridades en ingeniería de calidad se han pronunciado en contra de la utilización rutinaria de los índices de capacidad del proceso, tales como  $C_p$  y  $C_{pk}$ , con base en que son una sobre simplificación de un fenómeno complejo. Ciertamente, es posible que cualquier estadístico que contiene información tanto de la localización (la medida y el centrado del proceso) como

de la variabilidad, y que quiera el supuesto de normalidad para que su interpolación tenga sentido, sea mal usado (o se abuse de el).

### 1.6.2.3. ESTUDIOS DE CAPACIDAD DE INSTRUMENTOS Y SISTEMAS DE MEDICION

#### Cartas de control y métodos tabulares

Un aspecto importante de muchos esfuerzos de implementación del control estadístico de procesos es asegurar una capacidad adecuada del sistema de instrumentos de medición e inspección. En cualquier problema que implique mediciones, parte de la variabilidad observada se deberá a la variabilidad en el producto en si, y parte se deberá ala error de medición o variabilidad de los instrumentos.

Expresado en términos matemático,

$$\sigma_{total}^2 = \sigma_{producto}^2 + \sigma_{instrumentos}^2 \quad (50)$$

Donde  $\sigma_{total}^2$  es la varianza observada toral,  $\sigma_{producto}^2$  es el componente de la varianza debido al producto, y  $\sigma_{instrumentos}^2$  es el componente de la varianza debido al error de medición. Pueden usarse cartas de control y otros métodos estadísticos para separar estos componentes de la varianza, así como para dar una determinación de la capacidad de los instrumentos de medición.

## CAPITULO II

### MARCO CONCEPTUAL

---

#### 2.1. CONFIGURACION ACTUAL DE LA PLANTA CONCENTRADORA



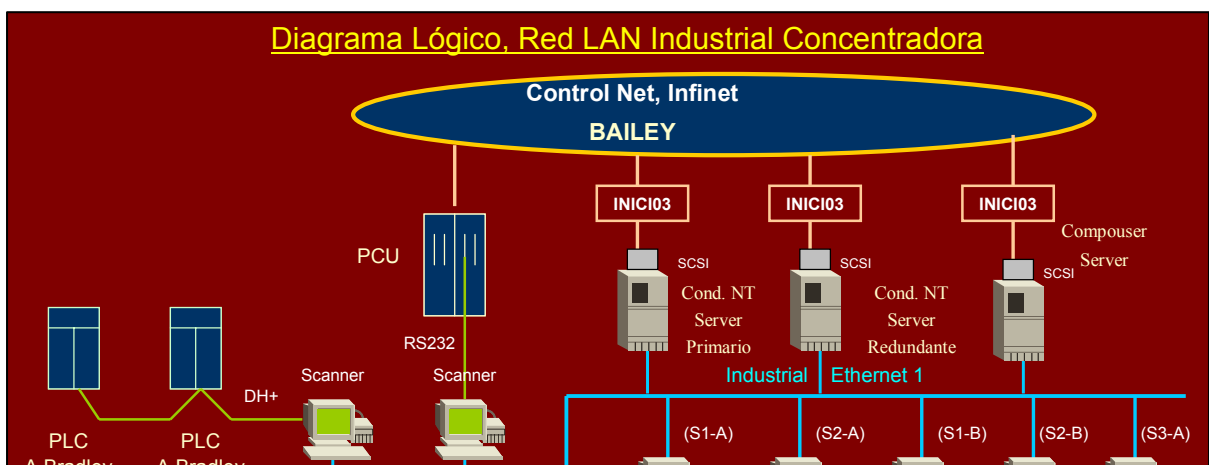
La Planta de Cobre cuenta con un Sistema de Control Distribuido (DCS) de la marca ABB, cuya infraestructura de comunicación es INFI90 y la interfaz con el operador se denomina Conductor NT; cuenta además con un software de configuración llamado Composer.

Las consolas Conductor NT son sistemas cliente – servidor basados en computadoras convencionales ejecutadas en el sistema operativo Windows 2000. Se compone de un servidor que se comunica con la red INFI90, anillo o loop redundante de comunicación donde se cuelgan todas las Unidades de los Procesadores (PCU) y las consolas servidores de interfaz del operador, recaban los datos actuales del proceso y almacenan los datos históricos.

Este servidor sirve como fuente de datos para el cliente, que es una interfaz gráfica mediante el cual el operador realiza acciones necesarias para controlar la operación de los procesos. La comunicación entre el servidor y el cliente es a través de Ethernet.

En la planta se cuenta con 2 servidores o primarios y 4 clientes o secundarios de Conductor NT y 01 servidor y 01 cliente de Composer.

La Planta Concentradora de Cuajone cuenta con un Sistema de Información de Planta (PI System). Este sistema se basa en una arquitectura cliente – servidor que puede ser implementada en diferentes configuraciones y ambientes



*Figura 16*

El servidor PI es la máquina donde reside el sistema de archivos PI (el Data Archive) que provee el almacenamiento centralizado de información de proceso, proporcionada por módulos especializados llamados interfases, que se encargan de adquirir datos desde terreno y desde otras fuentes. Las interfases pueden ser instaladas tanto en el servidor PI (Home Node) como en otras máquinas, denominadas PI-Net y Nodos API.

En el servidor PI o “Home Node” se encuentra, además del Data Archive, una serie de módulos de software (tales como Performance Equations y el driver ODBC) que aumentan las capacidades del sistema y permiten administrar el sistema de archivos y la base de datos de puntos (tags) que mantienen la información de configuración del sistema.

Las interfaces con consideradas herramientas clientes que escriben datos en PI. Una de las herramientas del PI System es el PI ProcessBook, el cual permite el manejo de pantallas que muestra la información del proceso, almacenada en un sistema de Archivos de PI y/o desde una base de datos relacional, de manera gráfica. PI ProcessBook realiza el manejo de pantalla en el computador del “cliente”.

Los datos entre tanto, son recolectados y almacenados en un computador diferente. Este computador proporciona los datos a las aplicaciones de los clientes cuando los usuarios hacen los requerimientos respectivos, por eso se le denomina “servidor”.

Las aplicaciones PI ProcessBook cotienen una o mas colecciones de despliegues, denominadas ProcessBook. Los datos dinámicos y estáticos que se pueden incluir en cada despliegue provienen desde las siguientes fuentes:

- Sistema de Archivos PI. Datos históricos y en el tiempo real desde uno o mas servidores.
- Datos de otras bases de datos que soporten conectividad ODBC.
- Objetos enlazados o vinculados provenientes de aplicaciones que soporten OLE automation. Estos objetos pueden ser hojas de cálculo, gráficos, etc.

Estas herramientas han hecho que recurrir al PI System sea una rutina obligada para revisar los diferentes parámetros y variables operativas del proceso, revisando en línea como se esta trabajando en la Concentradora Cuajone.

El anexo C muestra la Arquitectura PI de la Concentradora Cuajone, la forma como se enlazan al servidor PI los distintos equipos y sistemas de control para almacenar de manera automática sus principales datos operacionales configurados previamente mediante tags.

Dentro de las herramientas tecnológicas con que se cuenta en Concentradora Cuajone se tiene a los PSM

El analizador de tamaños de partículas (Particle Size Monitor) Multipoint PSM 400, es un analizador en línea que sirve para determinar el tamaño de partículas y el porcentaje de los sólidos dentro de un lodo acuoso durante las 24 horas de operaciones diarias.

Proporciona resultados de fracción de tamaños múltiples para mejorar la información que se presenta en la distribución del tamaño de molienda.

Es un equipo cuya medición se basa en emisión de impulsos de energía ultrasónica en seis frecuencias hacia la muestra. Los parámetros de atenuación múltiple que se derivan de estos impulsos son llamados factores de pulpa y están directamente relacionados a lo fino, grueso y a la concentración de las partículas del lodo acuoso. La medición de la atenuación de los pulsos ultrasónicos en frecuencias múltiples se utiliza para determinar el tamaño de partículas y el porcentaje de los sólidos. Las frecuencias de operación se eligen basadas en el intervalo del tamaño de partículas, la gravedad específica y la densidad del lodo acuoso. Los factores de pulpa se usan como variables en los modelos de calibración para cada fracción de tamaño.

El multipoint PSM 400 es un instrumento robusto en línea que bombea sus propias muestras a una velocidad aproximadamente de 75 lpm. Esta muestra es extraída del proceso mediante un sistema de vacío, se acondiciona y se le elimina todas las burbujas de aire dentro del tanque del Eliminador de Aire.

Luego pasa para la medición a través de unos sensores ultrasónicos diametralmente opuestos instalados en el módulo de análisis de muestras.

El PSM Multipoint 400 cuenta con cuatro componentes:

- Un sistema de eliminación de burbujas de aire de la muestra antes de ser analizada.
- Un módulo provisto de sensores ultrasónicos por donde pasa la muestra para sus análisis.

- Una estación de ingeniería desde donde el operador monitorea la operación del equipo y hace los cambios de parámetros según los requerimientos operacionales.
- Un equipo de acondicionamiento del agua para estandarización de los sensores (regresar a cero sus lecturas antiguas).

## **2.2. PROCEDIMIENTOS DE OPERACION**

Para que la planta opere de la manera en que fue diseñada, el operador debe controlar cuidadosamente ciertas variables del proceso. Estas variables incluyen datos de proceso tales como presión, temperatura, caudal, nivel y densidad, por nombrar sólo algunos. Cada variable que se debe controlar ha sido dividida en lazos de control separados. Cada lazo de control consta de una descripción escrita, de un diagrama de bloques y de un diagrama de lazos. La descripción por escrito se divide en las siguientes partes:

- El propósito de controlar la variable.
- El método utilizado para controlar la variable automáticamente.
- El método utilizado por el operador para controlar la variable manualmente.
- En el caso de las válvulas de control, el tipo de válvula y si la válvula falla cerrada o abierta ante la pérdida de energía.

Tanto la capacidad productiva como la recuperación metalúrgica de la planta dependen en gran medida de la rigurosidad que el operador controla estas variables.

### **2.2.1. LAZOS DE CONTROL DE MOLIENDA**

#### **a.- CONTROL DE LA TASA DE ALIMENTACIÓN AL MOLINO DE BOLAS 1C**

El sistema de control de la tasa de alimentación al Molino de bolas 1C monitorea y controla la tasa de alimentación de mineral al molino de bolas. El sistema de control de alimentación a los Molinos de bolas 1D y 2E es idéntico. El ajuste de la velocidad de dos fajas de alimentación de recuperación de velocidad ajustable fija la tasa de alimentación de mineral al molino.

Existen tres alimentadores para cada molino de bolas. Los alimentadores exteriores están equipados con controladores indicadores de velocidad individuales. El alimentador central está equipado con una transmisión de velocidad fija. Los alimentadores equipados con transmisión de velocidad aumentan o disminuyen la velocidad según se requiera para mantener la tasa de carga deseada en la faja de mineral fino.

Existen dos métodos que se usan para controlar automáticamente la tasa de alimentación al molino de bolas. La tasa de alimentación se puede controlar según el nivel en el cajón de alimentación a ciclones del molino de bolas o se puede controlar según el rango de peso del mineral que va al molino.

**TABLA N° 1 - REFERENCIA CRUZADA DE LA ALIMENTACIÓN AL MOLINO**

Molino de Bolas	Instrumentación de Control de Nivel IE/LT/LIC	Instrumentación de control de Rango de Peso WT/WIC	Controladores de la velocidad de la Faja de Alimentación ST/SIC/SZ
IC 5835-1103	5835-1163-1	5835-1043-1	5835-1007-01 5835-1009-01
ID 5835-1104	5835-1164-1	5835-1044-1	5835-1010-01 5835-1012-01
IE 5835-1105	5835-1165-1	5835-1045-1	5835-1013-01 5835-1015-01
2A 5835-2101	5835-2161-1	5835-2041-1	5835-2001-01 5835-2003-01
2B 5835-2102	5835-2162-1	5835-2042-1	5835-2004-01 5835-2006-01
2C 5835-2103	5835-2163-1	5835-2043-1	5835-2007-01 5835-2009-01
2D 5835-2104	5835-2164-1	5835-2044-1	5835-2010-01 5835-2012-01
2E 5835-2105	5835-2165-1	5835-2045-1	5835-2013-01 5835-2015-01

## CONTROL AUTOMÁTICO

Existen dos modos de control automático posibles. En el modo Cascada, el operador ingresa el set point de nivel del cajón de alimentación a ciclones en el controlador indicador de densidad para fijar la tasa de alimentación deseada de mineral en el molino de bolas. En el modo Automático, el operador ingresa un set point de peso en el controlador indicador de peso de la tasa de alimentación de mineral para mantener la tasa de alimentación de mineral deseada en el molino de bolas.

### Modo Cascada

En el caso de control de nivel del cajón de alimentación a ciclones, la tasa de alimentación del molino varía según el nivel en el cajón de alimentación a ciclones. Un sensor de nivel montado sobre el sumidero mide el nivel de la

pulpa y esta medición se transmite (LE/LT-5835-1163-1) hacia un controlador indicador de nivel (LIC-5835-1163-01). El controlador indicador de nivel compara el set point de nivel ingresado por el operador con el nivel medido y, en caso de que haya diferencia, el controlador indicador de nivel cambia el set point remoto que se transmite al controlador indicador de peso (WIC-5835-1043-1).

Esta señal representa la tasa de alimentación requerida para cumplir con el set point del operador. El controlador indicador de peso transmite entonces la señal de la nueva tasa de alimentación a los controladores indicadores de velocidad (SIC-5835-1007-01 t SIC-5835-1009-01) de los dos alimentadores exteriores.

Cada controlador de velocidad transmite esta señal a la transmisión de velocidad variable de su faja de alimentación. Los transmisores de velocidad variable aumentan o disminuyen la velocidad de sus respectivos alimentadores, según se requiera. Si el nivel en el sumidero es mayor que el set point, las velocidades de las fajas de alimentación disminuyen. Si el nivel en el sumidero es menor que el set point, las velocidades de la faja de alimentación aumentan.

### **Modo Automático**

En este modo de operación, el proceso de control no considera el nivel del sumidero. El operador cambia el set point de la tasa de alimentación de mineral según la observación visual de la demanda de la carga de circulación del molino, así como también de otras variables en las operaciones de la planta.

La tasa de alimentación al molino se mide con un elemento de peso y transmite junto con una lectura de la velocidad de la faja hacia un transmisor de peso (WT-5835-1043-1). El transmisor de peso calcula la tasa de alimentación de mineral. Éste envía la señal de tasa de alimentación de mineral ala



controlador indicador de peso de la tasa de alimentación de mineral (WIC-5835-1043-1).

El controlador indicador de peso de la tasa de alimentación transmite la señal de la tasa de alimentación a los controladores indicadores de velocidad de las fajas de alimentación (SIC-5835-1007-01 y 1009-01). Cada controlador de velocidad transmite esta señal a las unidades de transmisión de velocidad ajustable del alimentador de faja de recuperación de mineral (SZ-5835-1007-1 y 1009-1). Cada transmisión de velocidad ajustable compara la señal del controlador de velocidad con la velocidad medida (ST-5835-1007-1 y 1009-1) y aumenta o disminuye la velocidad de su alimentador, según se requiera, para mantener el set point de la tasa de alimentación. Normalmente, el operador ajusta el set point de la tasa de alimentación para mantener la tasa en el rango prescrito por el metalúrgico de la planta.

### **CONTROL MANUAL**

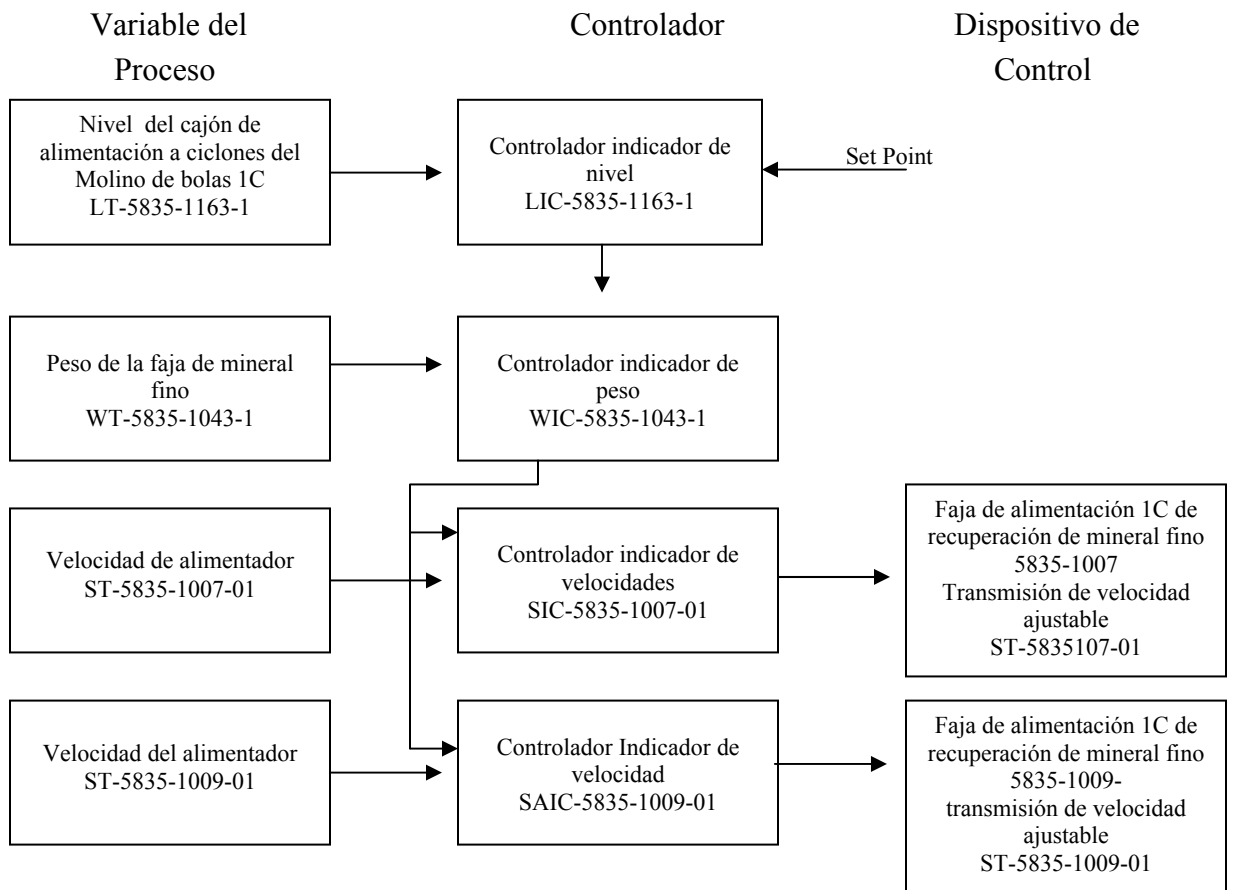
El operador puede cambiar el controlador indicador de peso (WIC-5835-1043-1) al modo Manual. Luego el operador puede ajustar la salida del controlador el que, a su vez, ajusta las velocidades de los alimentadores. Si los controladores indicadores de velocidad también están en modo Manual, la tasa de alimentación de mineral se puede cambiar, ajustando las velocidades de los alimentadores. El operador debe ser muy cuidadoso al operara el modo Manual.

Alternativamente, con el switch Local/Remoto (Local/Remoter) del panel de control de terreno de cada alimentador colocado en la posición Local, el operador ajusta manualmente la velocidad de cada alimentador en terreno.

En ambos casos, el operador debe monitorear la tasa de alimentación con mucha más precisión que en el modo Automático para asegurar que la tasa de alimentación deseada corresponda a la tasa de alimentación real. La operación en el modo, Manual no es recomendable en circunstancias normales.

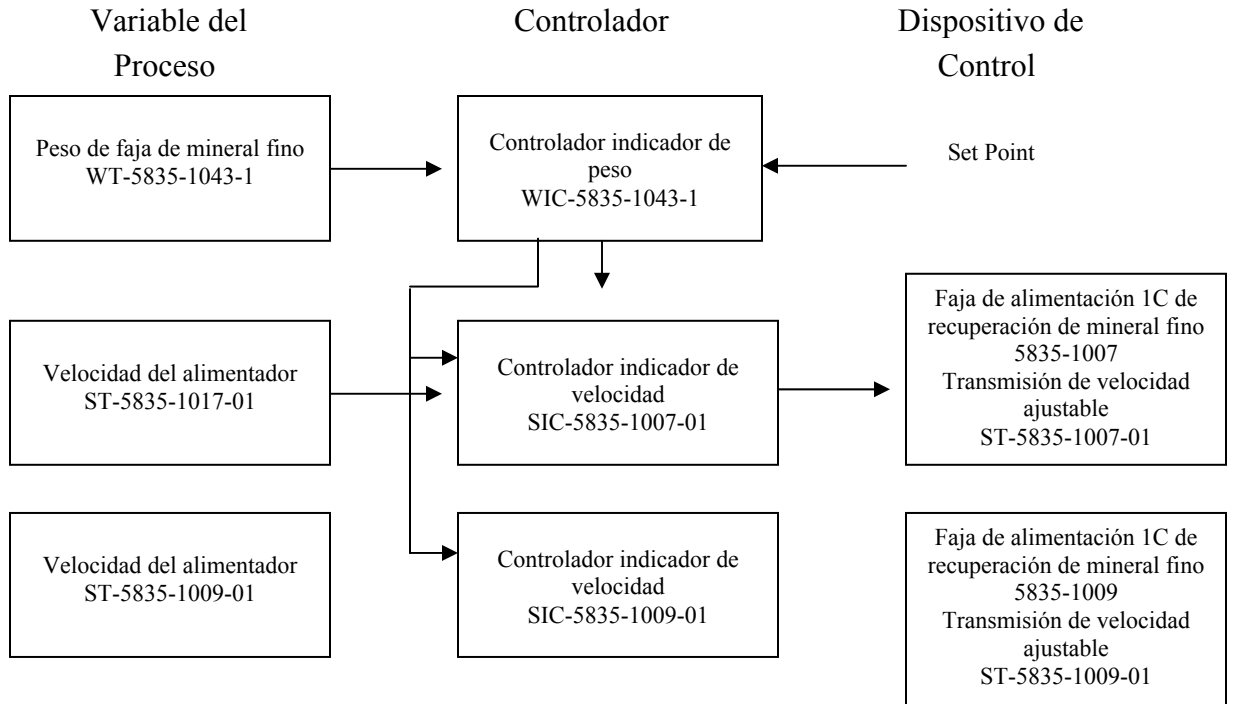
**DIAGRAMA DE BLOQUES**

**Modo Cascada**



*Figura 17*

**Modo Automático**



*Figura 18*

**b.- CONTROL DEL NIVEL DEL TANQUE DE PULPA DEL POLVO DE RECOLECCIÓN DE POLVO DE RECUPERACIÓN DE MINERAL FINO**

El objetivo del lazo de control del nivel del tanque de pulpa del polvo de recolección de polvo de recuperación de mineral fino es mantener el nivel adecuado de pulpa en el tanque que recolecta la pulpa de descarga del colector de polvo y la bombea a los molinos de bolas.

Este tanque recibe pulpa de mineral desde los colectores de polvo norte y sur. La pulpa del colector de polvo del lado sur descarga en el tanque de transferencia de pulpa de polvo y se bombea al tanque de pulpa de polvo. La pulpa del colector de polvo del lado norte también descarga en el tanque de

pulpa de polvo. La pulpa combina se bombea a los Molinos de molienda 1C a 2E, donde el flujo controlado de pulpa se entrega a los molinos.

El operador debe mantener alto nivel del tanque. Al mantener el nivel alto en el tanque se asegura el mantenimiento del sello y que el aire no derive a un colector de polvo. Al mantener un nivel del tanque relativamente alto también mejora la eficiencia de la bomba. Una bomba centrífuga debe funcionar más rápido en los niveles de tanque inferiores que en los niveles de tanque superiores para bombear la misma cantidad de material. Al mantener un nivel de tanque amplio reduce el desgaste de la bomba y la demanda de energía de la bomba.

### **CONTROL AUTOMÁTICO**

El nivel en el tanque para pulpa de polvo (5835-2302) se mide con un transmisor de nivel (LT-5835-1300-1) y el valor medido se transmite al controlador indicador de nivel del tanque (LIC-5835-1300-01). El controlador indicador de nivel compara el valor medido con un set point ingresado por el operador. Si el nivel es diferente al set point, la salida del controlador indicador de nivel ajusta la posición de la válvula de agua fresca (LV-5838-1300-1) en la línea hacia el tanque para cambiar el nivel en el tanque a un valor consecuente con el set point.

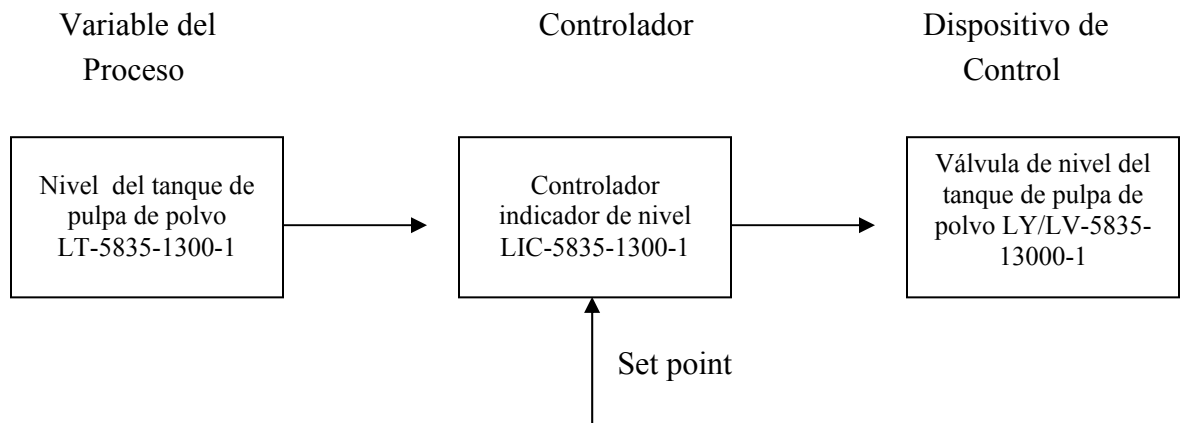
A modo de ejemplo, si el nivel en el tanque está por sobre el set point, la salida del controlador de nivel disminuye la abertura de la válvula de agua fresca. Esto tiene el efecto de disminuir el flujo de agua hacia el tanque y, por lo tanto, disminuye el nivel de la pulpa hasta que alcanza el nivel del set point. Si el nivel en el tanque está bajo el set point, la salida del controlador de nivel aumenta la abertura de la válvula de agua fresca. Esto tiene el efecto de

aumentar el flujo de agua hacia el tanque y así aumentar el nivel en el tanque hasta que alcance el nivel del set point.

### CONTROL MANUAL

El lazo de control de nivel del tanque para pulpa de polvo se puede operar en modo *Manual*. En este modo, el operador manipula directamente la salida del controlador para fijar la posición de la válvula en cualquier punto de su rango de operación. En el modo *Manual*, la válvula no responde en forma automática a los cambios en el nivel de la pulpa en el tanque y el colector de polvo puede perder el sello.

### DIAGRAMA DE BLOQUES



*Figura 19*

**TABLA N° 2 - VÁLVULAS DE CONTROL DEL NIVEL DEL TANQUE DE PULPA**

Válvula	Tipo	Actuador
Válvula de nivel LV-5835-1300-1	Válvulas de compuerta	Neumático

### c.- CONTROL DEL FLUJO DE PULPA DE POLVO A LOS MOLINOS DE BOLAS 1C A 2E

Los colectores de polvo de mineral fino de los lados norte y sur producen una pulpa que se bombea al círculo de molienda. El polvo recolectado por el colector del lado sur se bombea al tanque de pulpa de polvo del colector de polvo del lado norte. La pulpa combinada se bombea hacia un múltiple de distribución de los Molinos de bolas 1C a 2E. Las válvulas de control de flujo en las líneas de pulpa de polvo hacia estas canaletas de underflow a los ciclones del molino controlan la cantidad de pulpa recibida desde los colectores de polvo.

Los flujo metros miden los caudales de pulpa de polvo hacia los chutes de alimentación al molino. Las válvulas automáticas de control de flujo de pulpa controlan el flujo hacia el chute de alimentación. El control de flujo de pulpa de polvo al Molino de bolas 1C se analiza más adelante. El control del flujo de pulpa de polvo hacia los demás molinos es el mismo. Consulte la tabla de referencia cruzada donde encontrará los números de rótulo para equipos e instrumentos.

**TABLA N° 3 - REFERENCIA CRUZADA DEL CONTROL DE FLUJO DE PULPA**

Molino de Bolas	Instrumentación de Control de Flujo y de Pulpa de Polvo FE/FT/FIC/FY/FCV
1C-5835-1103	5835-1103-3
1D-5835-1104	5835-1104-3
1E-5835-1105	5835-1105-3
2 <sup>a</sup> -5835-2101	5835-2101-3
2B-5835-2102	5835-2102-3
2C-5835-2103	5835-2103-3
2D-5835-2104	5835-2104-3
2E-5835-2105	5835-2105-3

### CONTROL AUTOMÁTICO

En el modo *Automático*, el operador ingresa un set point de flujo de pulpa de polvo en el controlador indicador de flujo para mantener el caudal de pulpa de

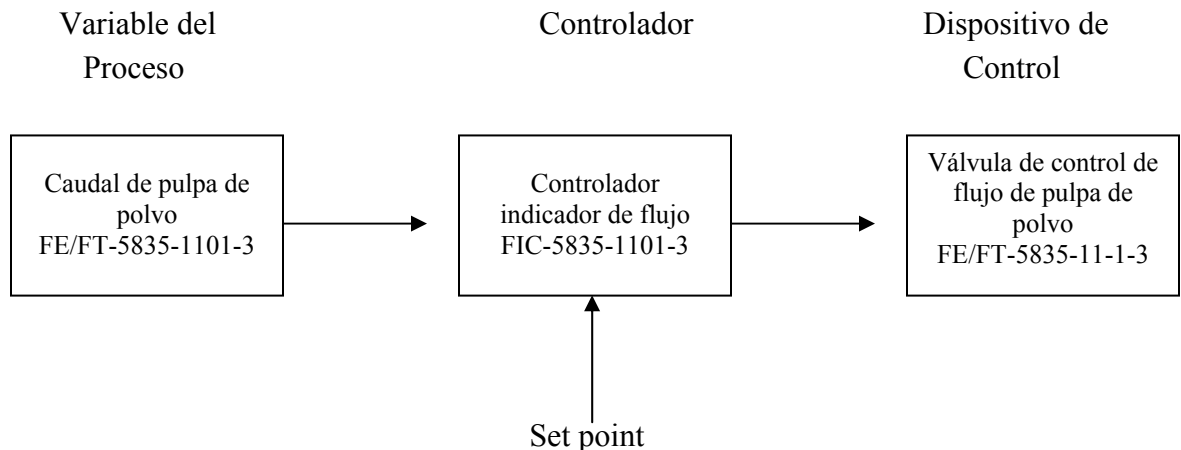
polvo deseado en el molino de bolas. Un sensor de flujo (FE-5835-1103-3) en la línea hacia el molino de bolas mide el flujo de pulpa de polvo. El controlador indicador de flujo recibe la medición del caudal de pulpa, del flujo metro y transmisor de flujo en la línea de agua. El controlador compara esta medición de flujo con el set point ingresado por el operador. Si los valores son diferentes, el controlador indicador de flujo ajusta la posición de la válvula de control de flujo para igualar los valores del flujo deseado y del flujo medido.

Si el caudal medido es superior al set point de flujo, la salida proveniente del controlador indicador de flujo le indica a la válvula de control de flujo de pulpa (FCV-5835-1103-3) en la canaleta de underflow a los ciclones que reduzca el flujo de pulpa, cerrándose parcialmente. Si el caudal medido es inferior al set point de flujo, la válvula de control de flujo se abre para aumentar el flujo.

### CONTROL MANUAL

El operador también puede seleccionar el modo de salida *Manual*. En este modo el operador cambia directamente el controlador indicador de flujo. Esto le permite colocar la válvula de control de flujo de pulpa en cualquier posición dentro de su rango de operación.

### DIAGRAMA DE BLOQUES



**Figura 20**

**TABLA N° 4 - VÁLVULAS DE CONTROL DEL FLUJO DE PULPA**

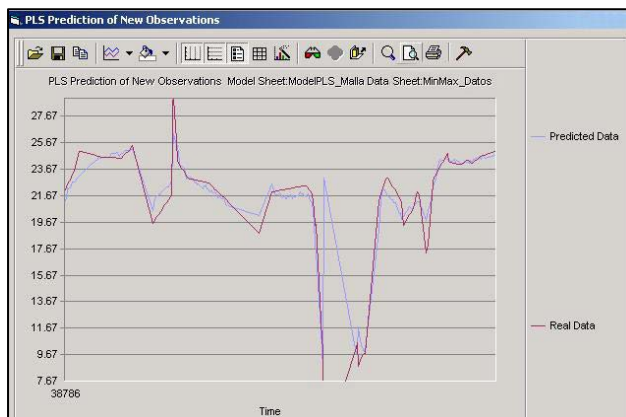
Válvula	Tipo	Actuador
Válvulas de control de flujo FCV-5835-1103-3 FCV-5835-1104-3 FCV-5835-1105-3 FCV-5835-2101-3 FCV-5835-2102-3 FCV-5835-2103-3 FCV-5835-2104-3 FCV-5835-2105-3	Pinch	Neumático

### 2.3. HIPOTESIS: INCREMENTAR TONELAJE EN CIRCUITOS DE MOLIENDA TENIENDO A LA MALLA +65 COMO VARIABLE DE CALIDAD

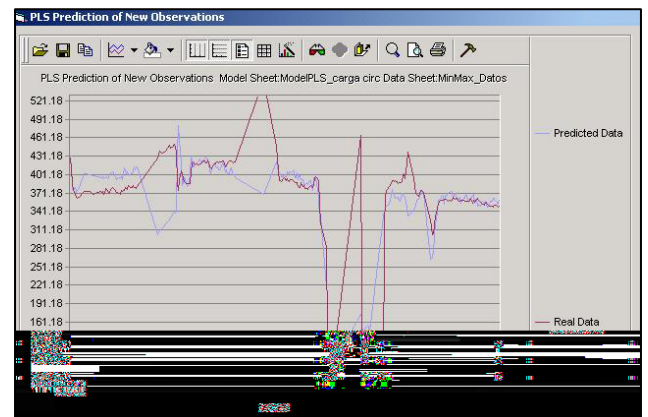
Se realizaron pruebas de análisis multivariante de los molinos primarios 1C y 1D haciendo uso del SCAN, los resultados obtenidos fueron los siguientes.

#### Molino 1C – Prueba N° 1:

Con la aplicación del SCAN en el Excel, se tomaron datos de las variables que actúan en el proceso de molienda. Se realizó modelos de predicción para la Malla +65 del O/F, Carga Circulante, Potencia y Tonelaje. Los resultados obtenidos se muestran a continuación:



(a) Modelo Predictivo del %Malla+65

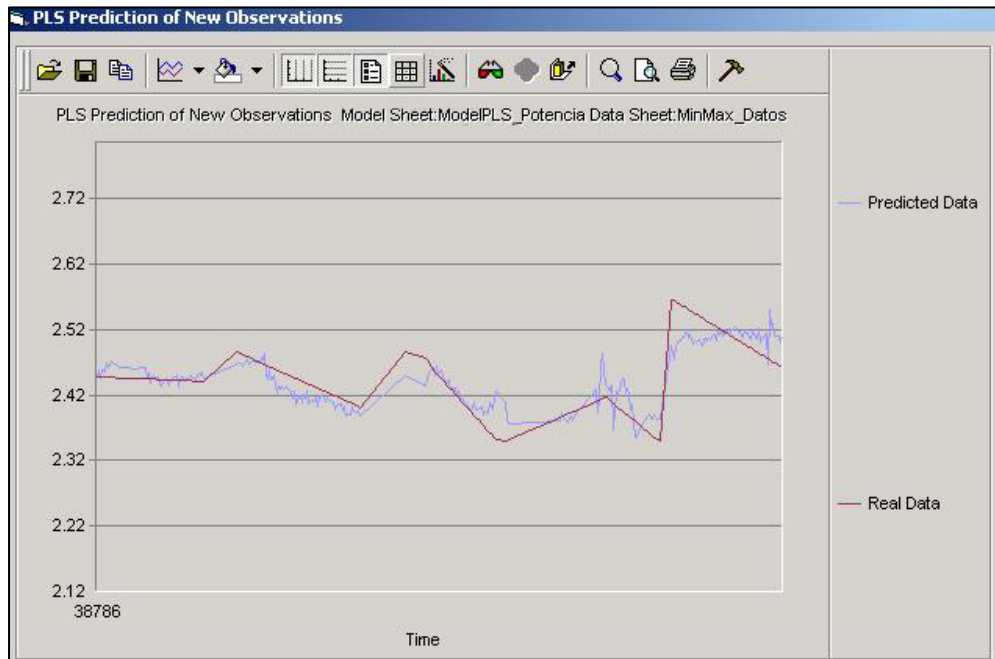


(b) Modelo Predictivo de la Carga Circulante





(c) Modelo Predictivo del Tonelaje



(d) Modelo Predictivo de la Potencia

**Figura 21**

ESTADISTICA DESCRIPTIVA DEL MOLINO 1C

10 DE MARZO DEL 2006

DATOS DE 12:00 AM – 13:40 PM

**TABLA N° 5**

	<b>Tonelaje</b>	<b>Malla +65</b>	<b>% Solidos</b>	<b>Presión ciclones</b>	<b>% Solidos alim ciclón</b>	<b>Amperaje</b>	<b>Nivel Cajón</b>	<b>% Carga Circ</b>	<b>Potencia</b>	<b>Flujo desc ciclón</b>	<b>Agua colectores</b>	<b>Agua Recup</b>
	<i>WIC1043-1</i>	<i>AI1103-35</i>	<i>DI1103-35</i>	<i>PI1173-01</i>	<i>DIC1173-1</i>	<i>II1173-05</i>	<i>LIC1163-01</i>	<i>XY1103-1</i>	<i>JI1103-1</i>	<i>XY1103-2</i>	<i>FIC1103-3</i>	<i>FIC1103-1</i>
Mean	216.919	22.042	35.639	2.089	62.865	75.019	35.433	380.517	2.443	8703.729	812.960	526.745
Standard Error	0.898	0.198	0.219	0.030	0.000	0.176	0.288	4.489	0.006	88.366	0.220	1.728
Median	214.758	22.507	35.812	2.012	62.863	74.195	36.186	380.563	2.443	8988.127	811.452	510.729
Mode	214.301	#N/A	#N/A	2.057	62.863	79.369	#N/A	#N/A	#N/A	#N/A	810.572	#N/A
Standard Deviation	16.307	3.597	3.982	0.550	0.002	3.202	5.230	81.549	0.100	1605.256	3.992	31.384
Sample Variance	265.907	12.935	15.857	0.303	0.000	10.250	27.357	6650.284	0.010	2576848.064	15.937	984.976
Kurtosis	16.839	8.958	10.275	2.431	-0.903	-1.373	0.016	15.220	155.637	14.987	0.680	-1.422
Skewness	2.469	-2.624	-2.833	1.006	0.870	-0.028	0.587	-1.037	-10.184	-3.373	0.334	0.604
Range	170.199	27.682	28.900	4.155	0.006	11.706	25.350	956.194	1.789	12481.010	27.473	88.604
Minimum	167.661	1.603	16.967	0.400	62.863	68.408	27.300	3.737	0.933	71.490	797.619	490.417
Maximum	337.860	29.285	45.867	4.555	62.869	80.114	52.650	959.931	2.722	12552.500	825.093	579.020
Sum	71583.175	7273.746	11760.780	689.395	20745.398	24756.215	11692.993	125570.775	806.130	2872230.697	268276.842	173825.889
Count	330.000	330.000	330.000	330.000	330.000	330.000	330.000	330.000	330.000	330.000	330.000	330.000
Confidence Level(95.0%)	1.766	0.389	0.431	0.060	0.000	0.347	0.566	8.831	0.011	173.835	0.432	3.399

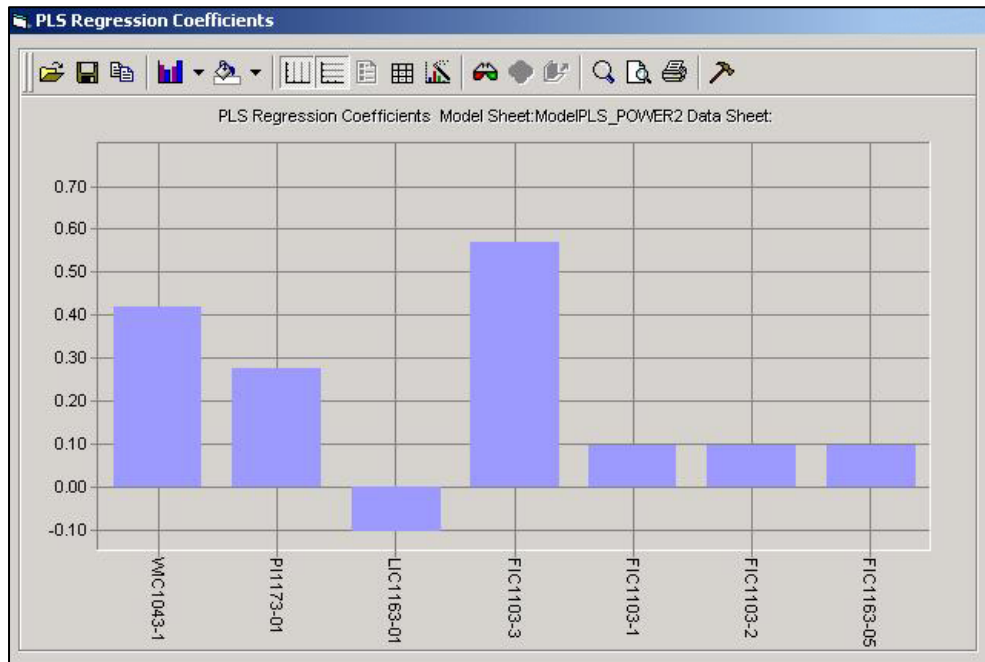
En esta primera prueba se consiguió lograr un incremento de tonelaje, aumentando el agua de colectores como indicaba el Modelo PLS obtenido por el SCAN.

**TABLA N° 6: MOLINO 1C PRUEBA N° 1**

<b>INCREMENTO TONELAJE</b>	
TONELAJE INICIAL	224.18 TM
HORA INICIO	3/10/06 2:25 PM
HORA FINAL	3/10/06 5:05 PM
INCREMENTO	12.37 TM

**Molino 1C – Prueba N° 2:**

Esta Segunda prueba se llevo a cabo el día 12 de Marzo; los datos evaluados corresponden al rango desde las 7:00 am a 11:00 am. El objetivo era incrementar el tonelaje y disminuir la carga circulante, para ello primeramente se busco incrementar la Potencia del Motor. Los modelos PLS obtenidos fueron:



**Figura 22**

*Gráfico de Coeficientes de Regresión de la Potencia*

Coeficientes de Regresión del Modelo PLS obtenidos por el SCAN, expresados en peso:

**TABLA N° 7: MOLINO 1C PRUEBA N° 2**

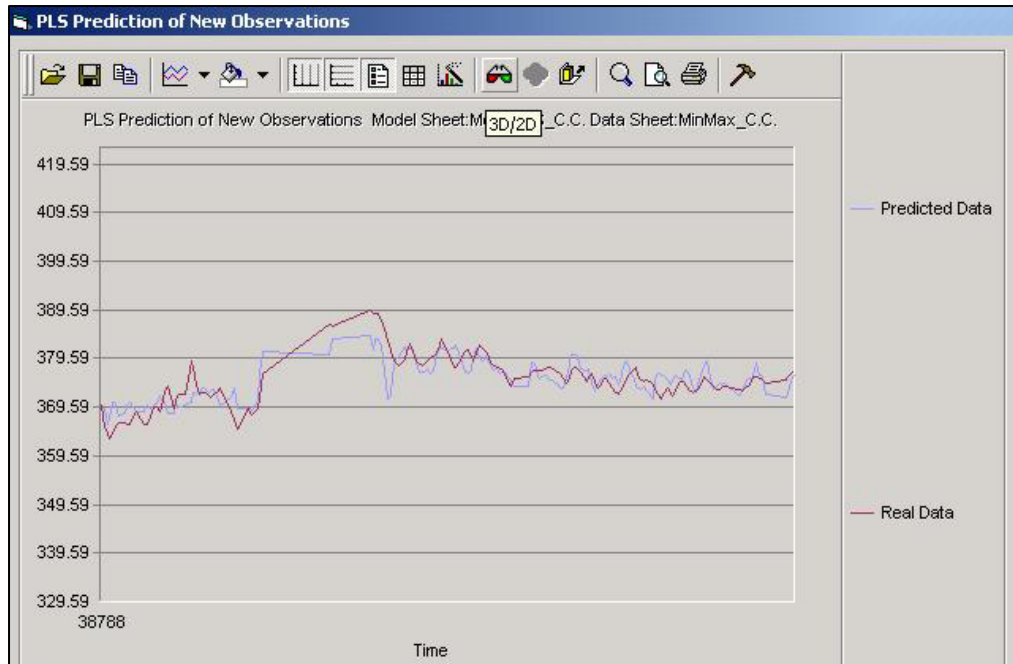
	<b>Variable</b>	<b>Peso</b>	<b>Peso Absoluto</b>	<b>%</b>
Agua Colectores	<b>FIC1103-3</b>	<b>0.57</b>	<b>0.57</b>	<b>34.49</b>
Tonelaje	<b>WIC1043-1</b>	<b>0.42</b>	<b>0.42</b>	<b>25.20</b>
Presión Ciclón	<b>PI1173-01</b>	<b>0.28</b>	<b>0.28</b>	<b>16.74</b>
Nivel Cajón	<b>LIC1163-01</b>	<b>-0.11</b>	<b>0.11</b>	<b>6.39</b>
Agua Fresca	FIC1103-2	0.09	0.09	5.73
Agua Dilución	FIC1163-05	0.09	0.09	5.73
Agua Recuperada	FIC1103-1	0.09	0.09	5.72

Con estos resultados se procedió a incrementar las siguientes aguas: Agua Fresca, Agua de Dilución y Agua Recuperada; la primera a las 11:01am y las otras dos a las 11:52am

**TABLA N° 8: VARIACIÓN DE LAS AGUAS**

	<b>VALOR ACTUAL</b>	<b>VARIACION</b>	<b>VALOR SUGERIDO</b>	<b>VALOR REAL</b>
Agua Colectores	700.00	399.48	1099.48	800
Agua Fresca	588.41	55.75	644.16	624.90
Agua Recuperada	510.02	48.28	558.30	571.00

Al efectuar dichas variaciones se observó el incremento de la potencia y una disminución de la caga circulante, entonces se procedió a incrementar el tonelaje:



Modelo Predictivo del Tonelaje - Molino 1C

**Figura 23**

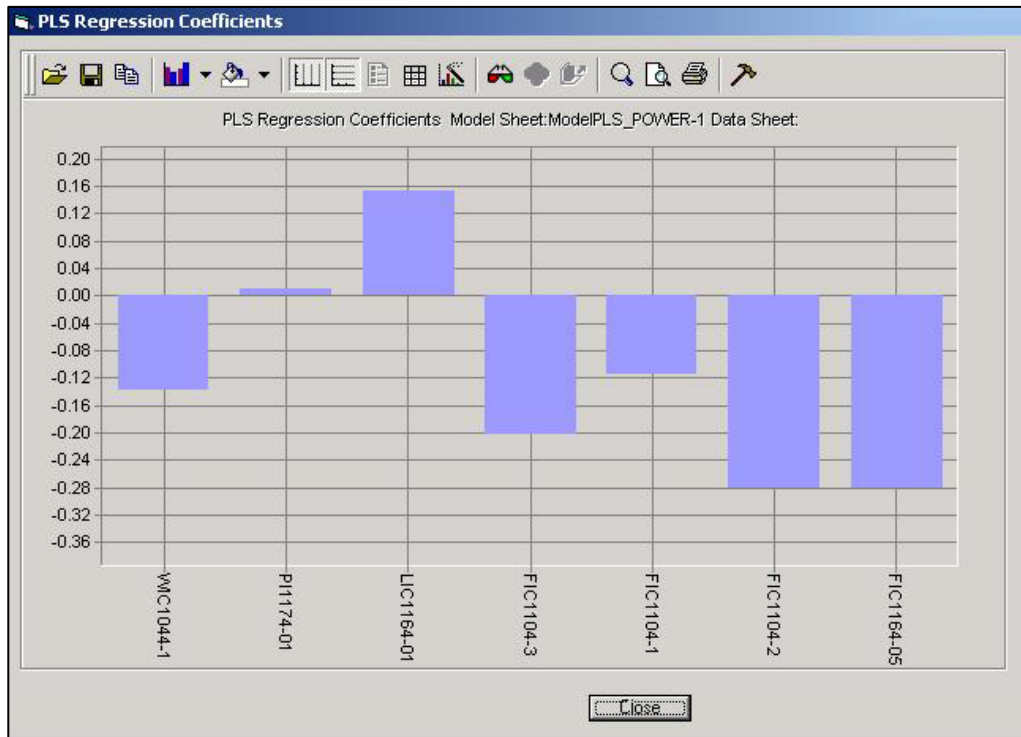
En esta segunda prueba durante las variaciones de las variables se realizó una parada del feeder 11-1C-1 por un tiempo de 1 hora, el incremento de tonelaje fue:

**TABLAN ° 9: RESULTADOS DE LA PRUEBA N°2**

<b>INCREMENTO TONELAJE</b>	
TONELAJE INICIAL	257.4 TM
HORA INICIO	3/12/06 1:10 PM
HORA FINAL	3/12/06 4:00 PM
INCREMENTO	17.78 TM

**Molino 1D – Prueba N° 3:**

El objetivo era también incrementar el tonelaje buscando en primera instancia incrementar la potencia. Esta prueba se llevó a cabo el 12 de Marzo con los datos extraídos del PI desde las 5:00 a.m. a 8:00 a.m.



Regresión de Coeficientes de la Potencia

**Figura 24**

**TABLA N° 10: Tabla de los Coeficientes de Regresión – Molino 1D**

	<b>Variable</b>	<b>Peso</b>	<b>Peso Absoluto</b>	<b>%</b>
Agua Fresca	<b>FIC1104-2</b>	<b>-0.28</b>	<b>0.28</b>	<b>23.83</b>
Agua Dilución	<b>FIC1164-05</b>	<b>-0.28</b>	<b>0.28</b>	<b>23.83</b>
Agua Colectores	<b>FIC1104-3</b>	<b>-0.20</b>	<b>0.20</b>	<b>17.10</b>
Nivel Cajón	<b>LIC1164-01</b>	<b>0.15</b>	<b>0.15</b>	<b>13.05</b>
Tonelaje	<b>WIC1044-1</b>	<b>-0.14</b>	<b>0.14</b>	<b>11.68</b>
Agua Recuperada	FIC1104-1	-0.11	0.11	9.73
Presión Ciclones	PI1174-01	0.01	0.01	0.79

Con estos resultados se procedió a incrementar el agua de colectores:

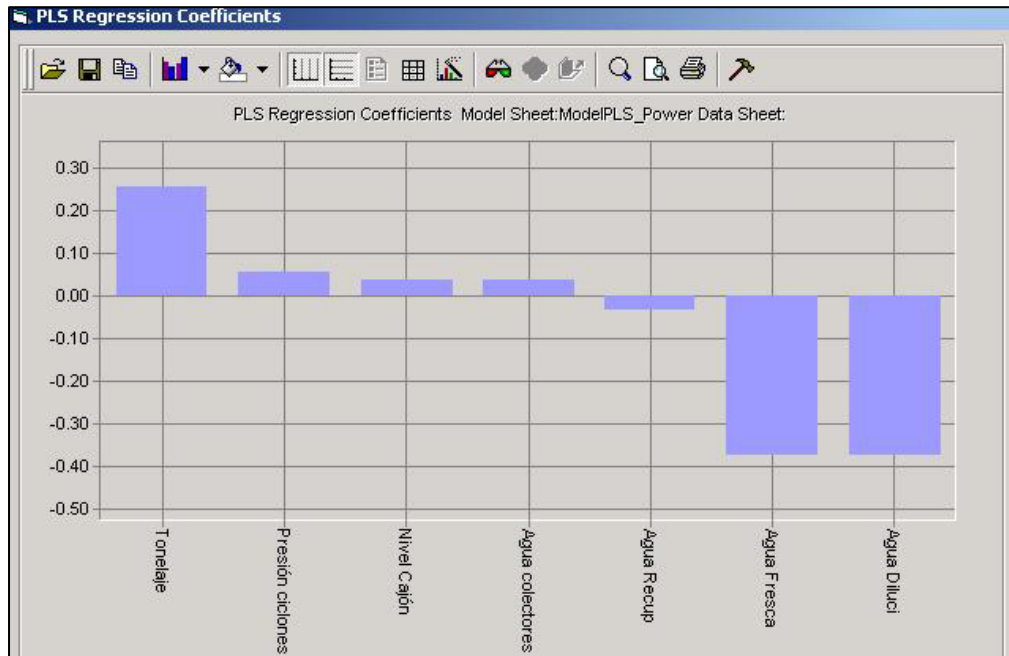
**TABLA N° 11: MOLINO 1D PRUEBA N°3**

	V.ALOR ACTUAL	VARIACION	VALOR SUG.	VALOR REAL
Agua Colectores	814.00	-164.38	649.62	750.00

Observando el incremento de la potencia, entonces se procedió a incrementar el tonelaje:

**Molino 1D – Prueba N° 4:**

Luego se realizó un Segundo modelo para lograr incrementar aun más el tonelaje:



*Regresión de Coeficientes de la Potencia*

**Figura 25**

**Tabla de los Coeficientes de Regresión – Molino 1D:**

**TABLA N° 12**

Variable	Peso	Peso Absoluto	%
<b>Agua Fresca</b>	<b>-0.38</b>	<b>0.38</b>	<b>31.99</b>
<b>Agua Dilución</b>	<b>-0.38</b>	<b>0.38</b>	<b>31.99</b>
<b>Tonelaje</b>	<b>0.26</b>	<b>0.26</b>	<b>21.87</b>
Presión ciclones	0.06	0.06	4.87
Agua colectores	0.04	0.04	3.23
Nivel Cajón	0.04	0.04	3.20
Agua Recuperación	-0.03	0.03	2.84

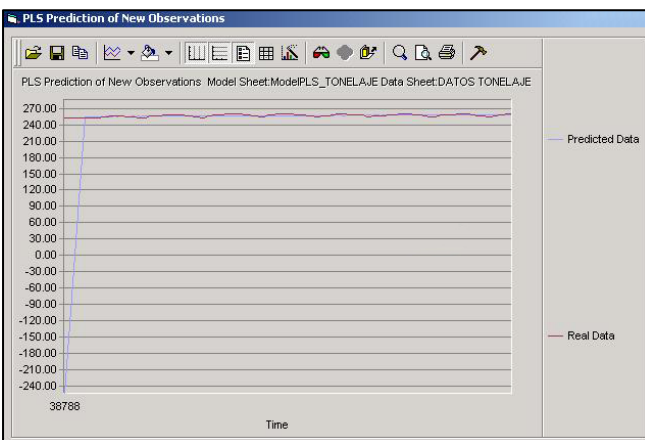
Las aguas a modificar fueron:

**TABLA N° 13: MOLINO 1D PRUEBA N° 4**

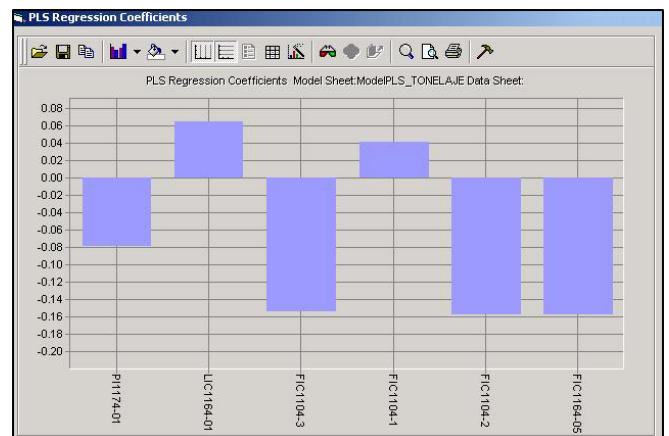
	VALOR ACTUAL	VARIACION	VALOR SUG.	VALOR REAL
Agua Fresca	303.32	-114.15	189.17	284.00
Agua Dilución	4558.13	-1715.44	2842.68	4510.00

**Molino 1D – Prueba N° 5:**

A continuación se realizó un modelo para el tonelaje buscando un mayor incremento:



(a) Modelo Predictivo del Tonelaje



b) Gráfico de Regresión de Coeficientes del Tonelaje

**Figura 26**



**Tabla de los Coeficientes de Regresión – Molino 1D:**

**TABLA N° 14**

	<i>Variable</i>	<i>Peso</i>	<i>Peso Absoluto</i>	<i>%</i>
<i>Agua Fresca</i>	<b>FIC1104-2</b>	<b>-0.16</b>	<b>0.16</b>	<b>24.06</b>
<i>Agua Dilución</i>	<b>FIC1164-05</b>	<b>-0.16</b>	<b>0.16</b>	<b>24.06</b>
<i>Agua Colectores</i>	<b>FIC1104-3</b>	<b>-0.15</b>	<b>0.15</b>	<b>23.65</b>
<i>Presión de Ciclones</i>	<b>PI1174-01</b>	<b>-0.08</b>	<b>0.08</b>	<b>12.03</b>
<i>Nivel de Cajón</i>	LIC1164-01	0.06	0.06	9.91
<i>Agua Recuperada</i>	FIC1104-1	0.04	0.04	6.29

Las aguas modificadas fueron:

**TABLA N° 15: VARIACIÓN DE LAS AGUAS PRUEBA N°4**

	<b>VALOR ACTUAL</b>	<b>VARIACION</b>	<b>VALOR SUG.</b>	<b>VALOR REAL</b>
<i>Agua Fresca</i>	283.39	-44.45	238.94	241.70
<i>Agua Colectores</i>	749.63	-115.59	634.04	643.00

Obteniéndose un incremento final de:

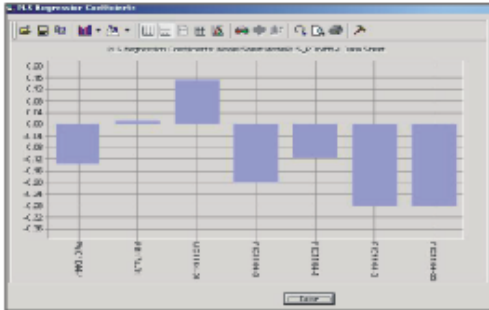
**TABLA N° 16: RESULTADOS DE LA PRUEBA N° 4**

<b>INCREMENTO TONELAJE</b>	
<i>TONELAJE INICIAL</i>	240.7 TM
<i>HORA INICIO</i>	3/12/06 9:00 AM
<i>HORA FINAL</i>	3/12/06 3:48 PM
<i>INCREMENTO</i>	93.86 TM

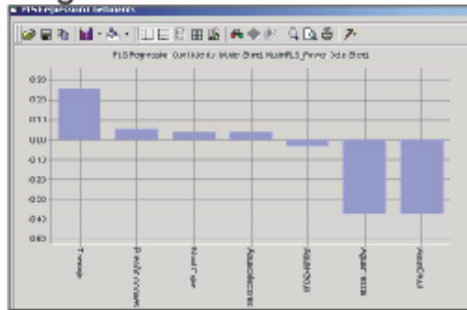
# INCREMENTO DE TONELAJE – CONCENTRADORA CUAJONE

## Coeficientes de Regresion de los Modelos PLS:

### Primer Modelo



### Segundo Modelo



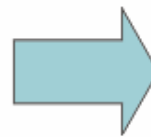
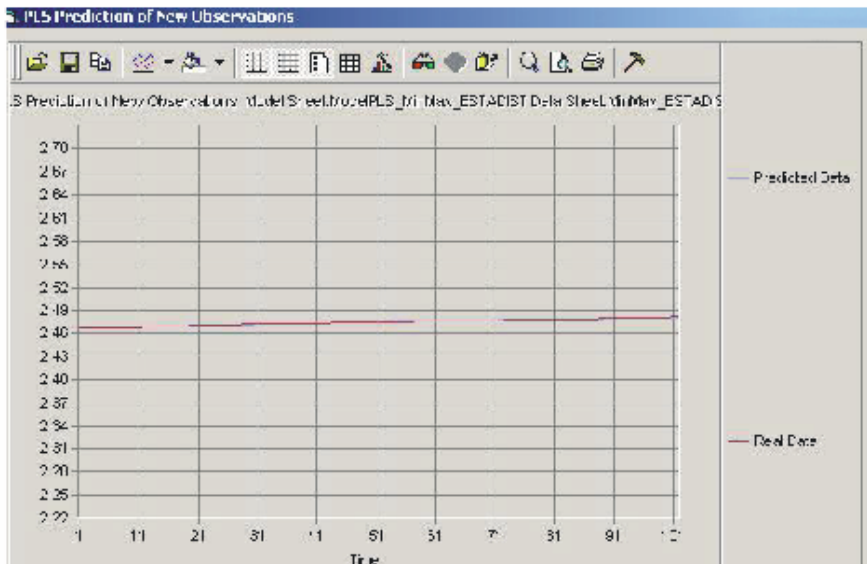
	Variable	Peso	Peso Absoluto	%
Agua Fresca	FIC1104-2	-0.28	0.28	23.83
Agua Dilución	FIC1164-05	-0.28	0.28	23.83
Agua Colectores	FIC1104-3	-0.2	0.2	17.1
Nivel Cajón	LIC1164-01	0.15	0.15	13.05
Tonelaje	WIC1044-1	-0.14	0.14	11.68
Agua Recuperada	FIC1104-1	-0.11	0.11	9.73
Presión Ciclonas	PI1174-01	0.01	0.01	0.79

- Modelos con PLS para la Potencia indicaron que se debería disminuir las aguas de Colectores, Fresca y Dilución

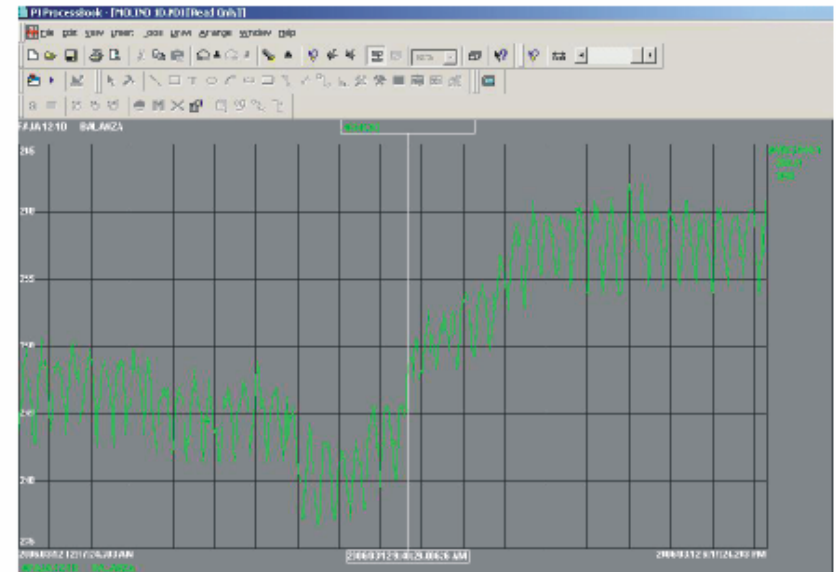
	Flujo Normal (LPM)	Flujo Corregido con SCAN (LPM)
Agua Colectores	814	643
Agua Fresca	303	241
Agua Dilución	4558	4510

- Se logro incrementar 93.86TM monitoreando un tiempo de 8h.

## Gráfico de PLS para la Potencia:



## Tendencia visualizada en el Process Book:



## 2.4. JUSTIFICACION

Para lograr el objetivo de la Hipótesis de la Tesis: Incrementar el Tonelaje en la Molienda, se hace necesario adaptar, integrar y conciliar diversos componentes de la tecnología disponible en la Planta Concentradora Cuajone. El logro del incremento del tonelaje, requirió el desarrollo de un Soft Sensor (Sensor Virtual) del %malla +65 como un variable de control de calidad y del % de Sólidos en el OverFlow. La justificación del desarrollo de estos Soft Sensor fue para continuar con la medición de las variables (%malla +65 y %Sólidos del OverFlow) cuando exista alguna falla o mantenimiento del PSM.

La información recolectada de las variables de operación del circuito de molienda clasificación, permitió elaborar modelos multivariantes con el uso de la Aplicación SCAN, los cuales pudieron ser ingresados de dos formas al PI System:

1. Mediante los Data Set, en forma manual.
2. Mediante el PI ACE, en forma automática.

El ingreso ya sea en forma manual o automática puede ser observado con ayuda del Process Book, el cual mostrará los modelos virtuales de las principales variables de operación. Los modelos generados fueron para el %Carga Circulante, Potencia, %Malla +65, %Sólidos en el OverFlow y el Tonelaje.

Dichos modelos permitirán realizar correcciones de las desviaciones de las condiciones de la operación, las cuales justificaran el logro del incremento del tonelaje. Toda esa información retornará al DCS, el cual permitirá realizar las correcciones pertinentes para mantener un Balance de Masa logrando incrementar el Tonelaje sin afectar el %Malla+65.

Un esquema de la teoría planteada se muestra a continuación:

PLANTEAMIENTO DEL DESARROLLO DE LA HIPOTESIS:

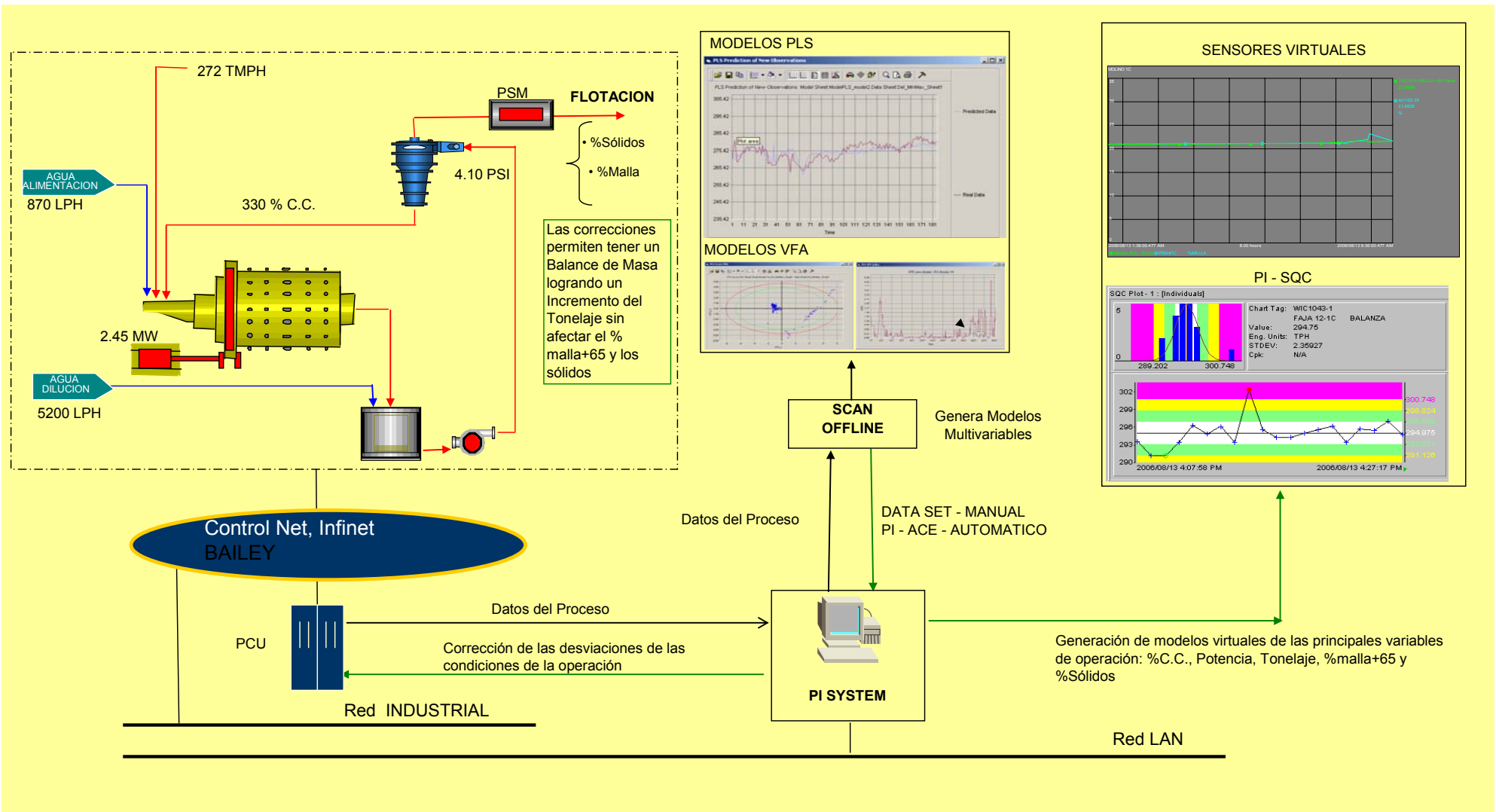


Figura 27

## CAPITULO III

### VALIDACION DEL METODO

---

La verificación de la Hipótesis requiere del uso de dos herramientas de control multivariante para la aplicación Scan-PI - el test analítico VFA y el test analítico PLS. Con dichos test se procedió a realizar los modelos predictivos de análisis multivariante con el uso del Scan.

#### 3.1. TEST ANALITICOS ANALISIS DE FACTORES DE VARIABILIDAD - VFA

##### 3.1.1. ANALISIS FACTORES DE VARIABILIDAD

El Análisis de Factores de Variabilidad (VFA), también llamado Análisis de Componentes Principales (PCA) ha sido la base del análisis de datos multivariante. VFA fue formulado por primera vez en estadística por Pearson, quien describió el análisis como *“la búsqueda de líneas y planos de acercamiento a sistemas de puntos en el espacio”*. Esta breve descripción dice mucho acerca de lo que trata VFA. El más importante uso de VFA es para representar set de datos multivariantes K-dimensionales en sub-espacios A-dimensionales, tal que  $A < K$ . Esta reducción de dimensión del problema original comprime y extrae toda la información relevante del voluminoso conjunto de datos obtenido. Esta representación del set de datos, a través de VFA, podría revelar clusters o agrupamientos de observaciones entorno a una característica común y descubrir relaciones ocultas entre variables.

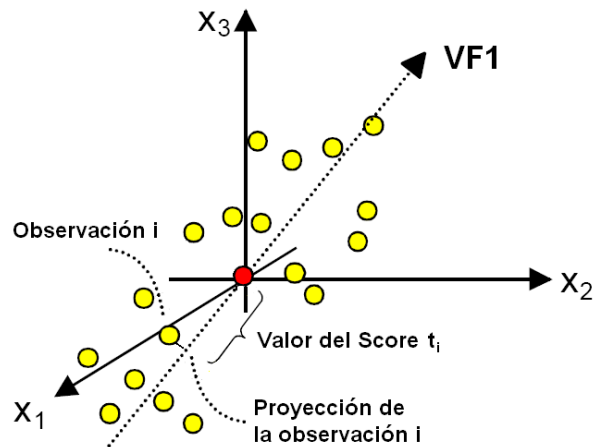
El número de factores de variabilidad que se consideran en un modelo puede elegirse de forma que la variabilidad de la matriz E (señal de ruido) sea menor que un 10% de la variabilidad propia de la matriz X (matriz de datos). Generalmente esto se logra con la determinación de los dos primeros componentes principales (VF1 y VF2).

El objetivo de VFA es descomponer la matriz- $X$  de datos para detectar fenómenos ocultos. Es fundamental asumir en VFA que en el espacio  $K$ -dimensional, las direcciones de máxima variabilidad de la nube o sistema de puntos está más o menos vinculada a estos fenómenos ocultos. Así, el cálculo de variables latentes o implícitas, llamadas *Componentes Principales o Factores de Variabilidad (VF)*, que coinciden con estas direcciones de máxima variabilidad, puede ser usado como una eficiente herramienta para entender e interpretar los datos. Los VF's corresponden a los vectores propios de la matriz de covarianza construida a partir de la matriz- $X$  de datos y marcan los ejes de máxima varianza en los datos.

A continuación se dará una interpretación geométrica de VFA, mostrando cómo se calculan los factores de variabilidad, cómo se generan modelos VFA a partir de ellos y la manera de recoger e interpretar la información que entregan estos modelos a partir de herramientas apropiadas de diagnóstico.

### **Cálculo del primer componente principal (VF1)**

Este primer componente es la línea en el espacio  $K$ -dimensional que mejor aproxima la dirección de máxima variabilidad de la nube de puntos de acuerdo al método de los mínimos cuadrados parciales. Esta línea atraviesa el centro de gravedad de la nube de puntos o punto promedio de los datos y no es paralela a ningún eje coordenado. La Figura 28 ilustra este concepto.



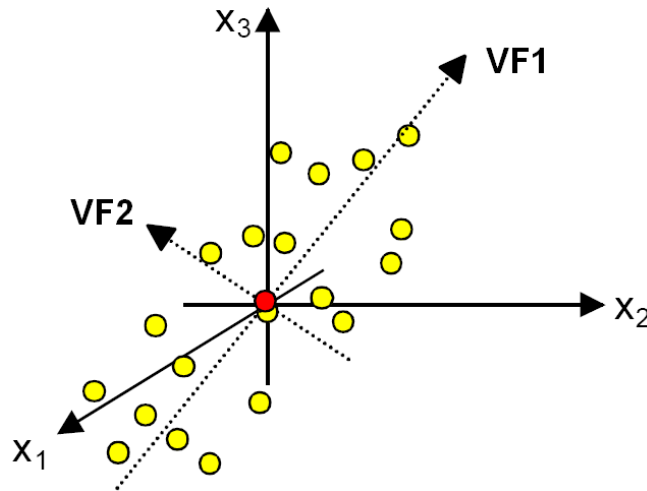
*Interpretación geométrica del primer componente principal VF1. Corresponde a la línea que mejor aproxima la dirección de máxima variabilidad de los datos en el espacio K-dimensional.*

**Figura 28**

Cada observación puede ser proyectada sobre esta línea para obtener un nuevo valor de coordenada respecto a un nuevo sistema de eje coordenado constituido por VF1. Esta nueva coordenada se conoce como *Score* y para la observación  $i$  se representa por el símbolo  $t_i$ .

### **Cálculo del segundo componente principal (VF2)**

Normalmente, sólo un componente principal es insuficiente para modelar adecuadamente un set de datos multivariantes. Por lo tanto, un segundo componente principal VF2 debe ser calculado. El segundo VF está representado por la línea que, al igual que VF1, atraviesa el centro de gravedad o punto promedio de los datos y representa la mejor aproximación a la segunda mayor dirección de máxima variabilidad de los datos, pero manteniendo siempre la ortogonalidad respecto al primer VF, tal como lo muestra la Figura 29.



*Interpretación geométrica del segundo componente principal VF2. Esta orientado de manera tal que representa la mejor aproximación a la segunda mayor dirección de máxima variabilidad de los datos manteniendo la ortogonalidad respecto al primer VF.*

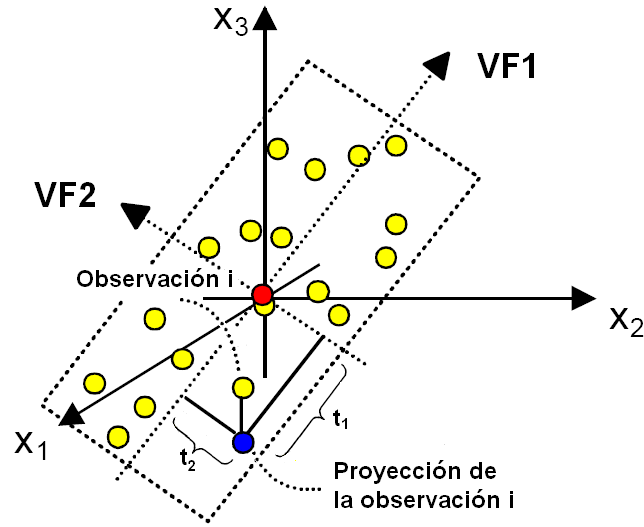
**Figura 29**

Al igual que el primer VF, la proyección del set de datos sobre el segundo VF genera otro nuevo valor de coordenadas pero esta vez contenidos sobre el nuevo eje coordenado VF2.

### **Modelo generado por el primer y segundo componente principal**

Cuando dos componentes principales han sido calculados, juntos definen un plano, un sub-espacio de menor dimensión dentro del espacio K-dimensional original de la matriz X de datos. La Figura 30 muestra la interpretación geométrica del plano formado por VF1 y VF2, el que al igual que los componentes principales anteriores pasa por el centro de gravedad o punto promedio medio de los datos. La construcción de este nuevo sub-espacio divide la nube de puntos en dos, dejando puntos sobre y bajo el plano.





*Plano formado por VF1 y VF2. Este plano es un sub-espacio de menor dimensión dentro del espacio K-dimensional original. Cada observación puede ser proyectada sobre este plano obteniendo nuevas coordenadas (scores) en el nuevo sistema de ejes coordenados formado por VF1 y VF2. La proyección de la observación i sobre el plano se muestra en azul.*

**Figura 30**

Cada punto u observación puede ser proyectado sobre este plano, generando para cada una de las proyecciones dos nuevas coordenadas o scores ( $t_1, t_2$ ) en el nuevo sistema de ejes coordenados formados por VF1 y VF2. Estas proyecciones representan la ubicación real de cada observación. A esta representación de la realidad a través de la generación de nuevos sub-espacios de menor dimensión dentro del espacio K-dimensional original se le llama Modelo VFA.

## ANÁLISIS DE FACTORES DE VARIABILIDAD (VFA)

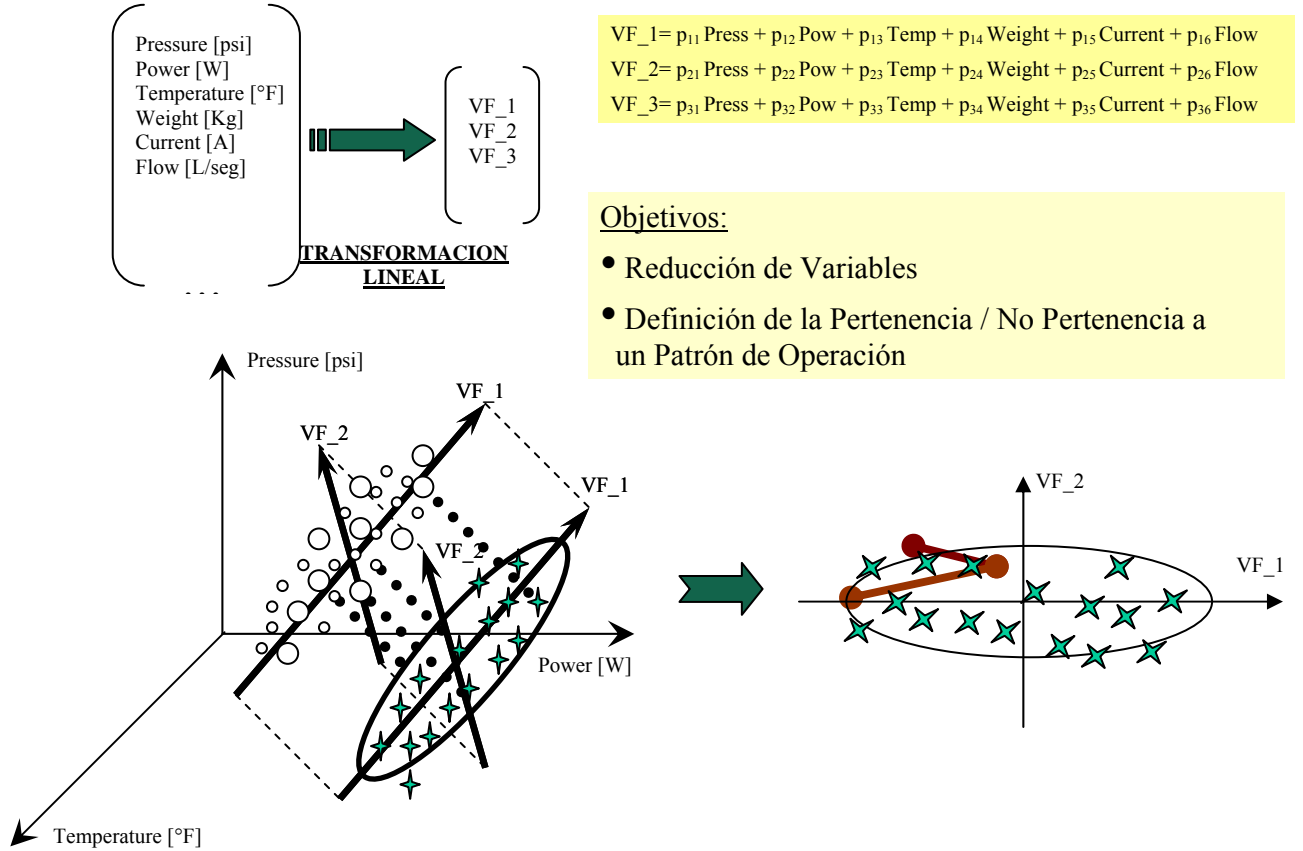


Figura 31

Matemáticamente, la descomposición VFA para una matriz X de datos con n muestras y k variables (centrados y escalados apropiadamente) es:

$$X = t_1 VF_1T + t_2 VF_2T + \dots + t_q VF_qT \dots \dots \dots (51)$$

Donde  $q = \min \{n, k\}$  y los términos  $t_i VF_i$  están ordenados por la cantidad de varianza que logran explicar.

$$VF_1 = p_{11} X_1 + p_{12} X_2 + p_{13} X_3 + \dots p_{1k} X_k \dots \dots \dots (52)$$

$$VF\_2 = p_{21} X_1 + p_{22} X_2 + p_{23} X_3 + \dots p_{2k} X_k \dots \dots \dots (53)$$

VFA descompone los datos de un conjunto de datos en una parte de señal y en una parte de ruido (E):

$$X = t_1 VF\_1^T + \dots t_k VF\_k^T + \dots t_a VF\_a^T + E \dots \dots \dots (54)$$

Para cada variable de salida:

$$X_1 = c_{11} VF\_1 + c_{12} VF\_2 + \dots c_{1a} VF\_a + E_1 \dots \dots \dots (55)$$

$$X_2 = c_{21} VF\_1 + c_{22} VF\_2 + \dots c_{2a} VF\_a + E_2 \dots \dots \dots (56)$$

El número de VF que se consideran en el modelo anterior puede elegirse de forma que la variabilidad de la matriz E sea menor que un 10% de la variabilidad propia de la matriz X.

De este análisis matemático se pueden obtener dos tipos de gráficos fundamentales que entregan información sobre el proceso analizado:

- ✓ Loading plot
- ✓ Score plot

**Loading Plot**

A partir de las siguientes ecuaciones que relacionan las VFs con las variables originales del proceso, se pueden graficar los coeficientes de VF1 v/s coeficientes de VF2 para cada variable:

$$VF\_1 = p_{11} X_1 + p_{12} X_2 + p_{13} X_3 + \dots p_{1k} X_k \dots \dots \dots (57)$$

$$VF\_2 = p_{21} X_1 + p_{22} X_2 + p_{23} X_3 + \dots p_{2k} X_k \dots \dots \dots (58)$$

**Score Plot**

- A partir de las siguientes ecuaciones que relacionan las VFs con los datos del proceso, se pueden formar gráficos de proyección de los datos en las VFs:

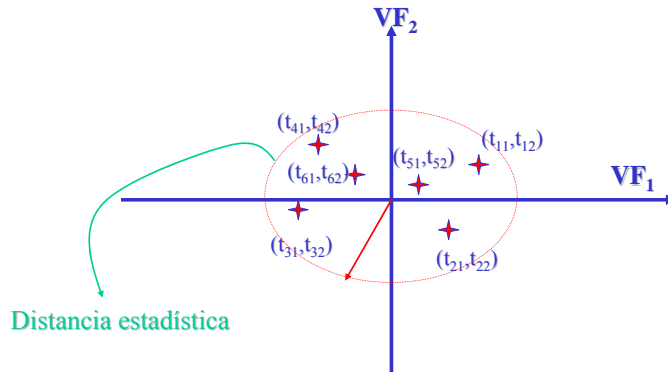
$$X(1) = t_{11} VF_{1T} + \dots t_{12} VF_{2T} + \dots t_{1a} VF_{aT} + E \dots \dots \dots (59)$$

$$X(2) = t_{21} VF_{1T} + \dots t_{22} VF_{2T} + \dots t_{2a} VF_{aT} + E \dots \dots \dots (60)$$

$$X(3) = t_{31} VF_{1T} + \dots t_{32} VF_{2T} + \dots t_{3a} VF_{aT} + E \dots \dots \dots (61)$$

$$X(k) = t_{k1} VF_{1T} + \dots t_{k2} VF_{2T} + \dots t_{ka} VF_{aT} + E \dots \dots \dots (62)$$

$$X(k) = t_{k1} VF_{1T} + \dots t_{k2} VF_{2T} + \dots t_{ka} VF_{aT} + E$$



*Figura 32*

**3.1.2. INDICE DE HOTELLING T<sup>2</sup>**

**CONCEPTO DE DISTANCIA ESTADÍSTICA**

Se necesita una herramienta que permita medir la pertenencia de los datos a un patrón de operación determinado.

Un conjunto de datos pertenece a un patrón de operación si estos datos están dentro de un rango definido por el patrón.

Para definir el rango de validez de un patrón de operación se hace necesario incorporar el concepto de “distancia estadística” dentro de un set de datos. De este modo, es posible definir un límite máximo para la distancia de un punto respecto al centro de los datos, en el que se asegure una operación dentro de un patrón esperado.

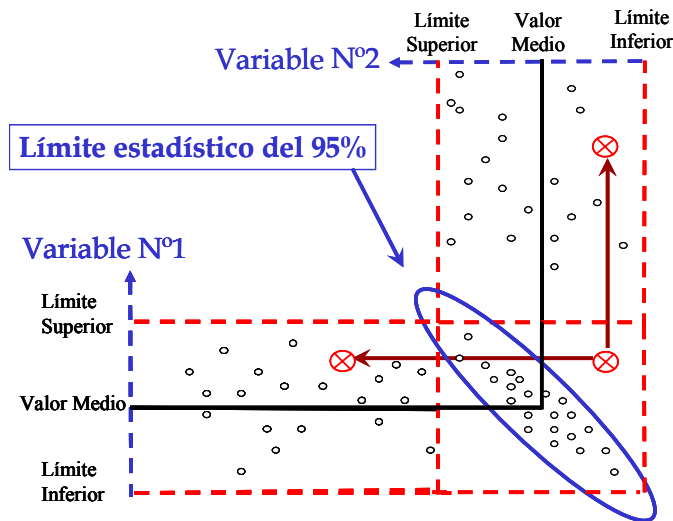


Figura 33

Para encontrar la máxima distancia aceptable, se utiliza el Test de Hotelling, que establece que el cociente entre la distancia al cuadrado desde un punto cualquiera al centro de los datos y la varianza total del set distribuyen Fisher:

$$T^2_i = \sum_{a=1}^A t_{ia}^2 / s_{ta}^2 \dots\dots\dots (63)$$

$$T^2_i > A(N^2 - 1) / (N(N - A)) * F_{critico}(\alpha) \rightarrow \text{ANORMALIDAD}$$

Donde:

**N:** Número de datos

**A:** Número de dimensiones de las proyecciones

De este modo, los límites de la elipse de control están dados por la siguiente expresión:

$$EJE_x = (s_{tx}^2 * F_{2, N-2, \alpha} * 2 * (N^2 - 1) / (N(N-2)))^{1/2} \dots\dots\dots (64)$$

$$EJE_y = (s_{ty}^2 * F_{2, N-2, \alpha} * 2 * (N^2 - 1) / (N(N-2)))^{1/2} \dots\dots\dots (65)$$

Si el proceso tiene más de dos variables, en teoría bastaría con realizar este análisis agrupando todas las variables del proceso de a pares

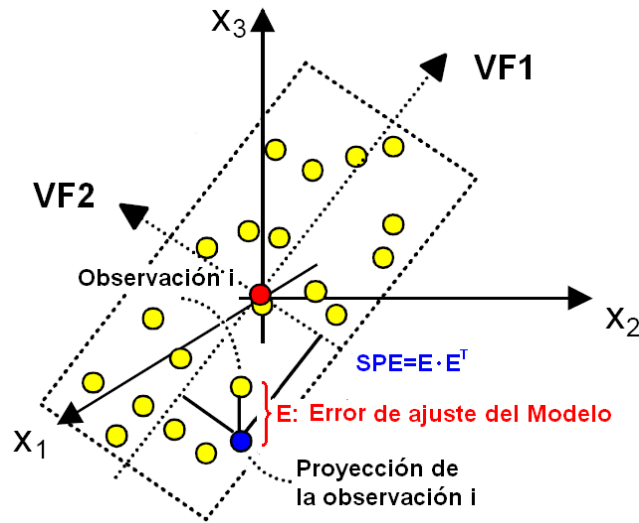
El estadístico  $T^2$  de Hotelling, que indica la pertenencia de los datos a la elipse definida por el patrón de operación:

$$T_i^2 = t_i \lambda^{-1} t_i^T = x_i VF_k \lambda^{-1} VF_k x_i^T \dots\dots\dots (66)$$

Si este índice es muy alto, quiere decir que los puntos están más lejos que la distancia estadística del centro de los datos, que define la pertenencia al patrón de operación.

**3.1.3. INDICE ERROR CUADRATICO DE PREDICCION - SPE**

Al igual que el Hotelling, es posible obtener una medida del error de predicción del modelo en forma global abarcando todos los conjuntos de scores plot construidos a partir del número total de VF's calculadas para el modelo. A esta medida global del error se le llama SPE (Error cuadrático de predicción) y corresponde a la suma cuadrática de los errores de cada observación. El error de predicción para cada observación corresponde a la distancia comprendida entre la observación y el Plano definido por el modelo. La Figura 34 ilustra este concepto.



Interpretación geométrica del error de predicción del modelo y del término SPE.

Figura 34

Graficando contra el tiempo de manera análoga al Hotelling global de la Figura 34, para un alto nivel del SPE de una observación en particular, significa que hay una pérdida de homogeneidad en los datos y que las correlaciones no son como las explica el modelo y por lo tanto, que hay un nuevo factor externo al modelo no considerado inicialmente que provoca esta desviación en el comportamiento de la data analizada.

El índice de ajuste del modelo (SPE), que indica la suma cuadrática de los errores E del modelo VFA:

$$SPE_i = e_i e_i^T = x_i (I - VF_k VF_k^T) x_i^T \quad \dots\dots\dots (67)$$

Si este índice es muy alto, implica que las correlaciones entre las variables no son como las explica el modelo

## Correlación positiva (Es necesario revisar todas las proyecciones)

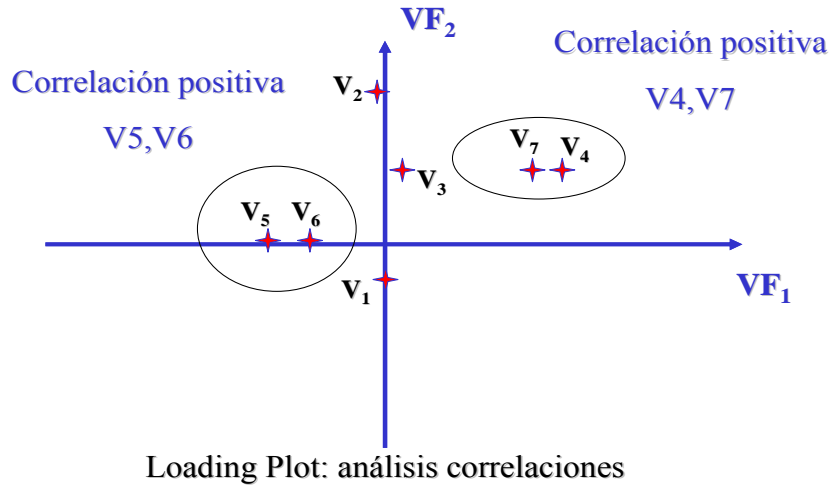
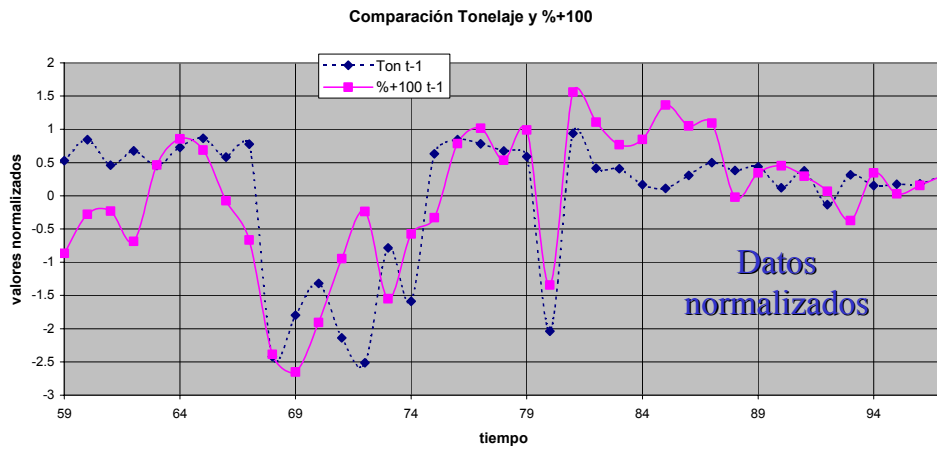


Figura 35

## Correlación positiva tonelaje y %+100



Loading Plot: análisis correlaciones

Figura 36



**Correlación negativa**  
**(Es necesario revisar todas las proyecciones)**

**No Correlación**

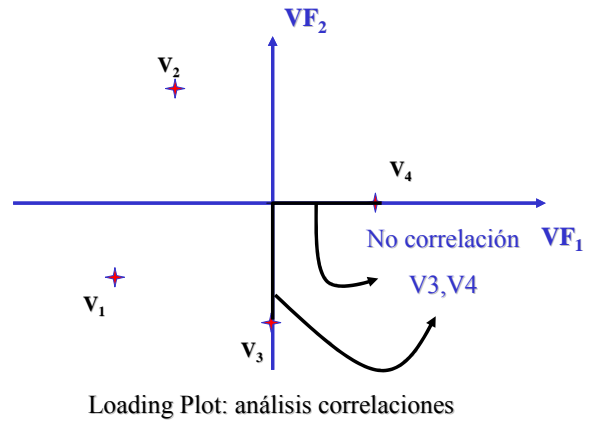
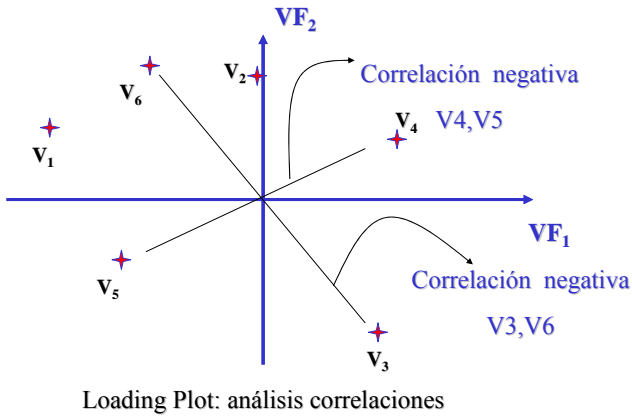


Figura 37

**Score plot: existencia de grupos**

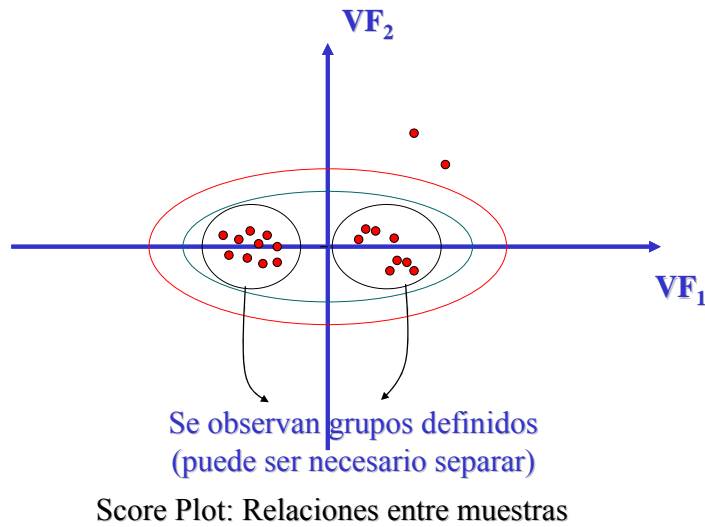
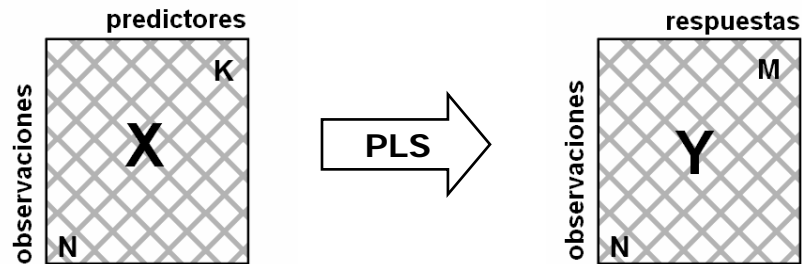


Figura 38

### 3.2. TEST ANALITICOS PROYECCION DE ESTRUCTURAS LATENTES - PLS

El análisis PLS se utiliza cuando se requiere conectar la información de una matriz- $X$  de datos de entrada a un proceso, con una Matriz  $-Y$  de datos de salida, a través de un modelo lineal multivariable. Figura 39.



*Figura 39: Para un determinado proceso, el análisis PLS relaciona el conjunto de datos de entrada  $X$  con el conjunto de datos de salida  $Y$ . El conjunto de datos de entrada  $X$ , corresponde a la matriz  $X$  de datos con  $K$  variables predictoras y el conjunto de datos de salida  $Y$  corresponde a la matriz  $Y$  de datos con  $M$  variables de respuesta. Así, a cada observación le corresponde una fila en  $X$  y otra en  $Y$ .*

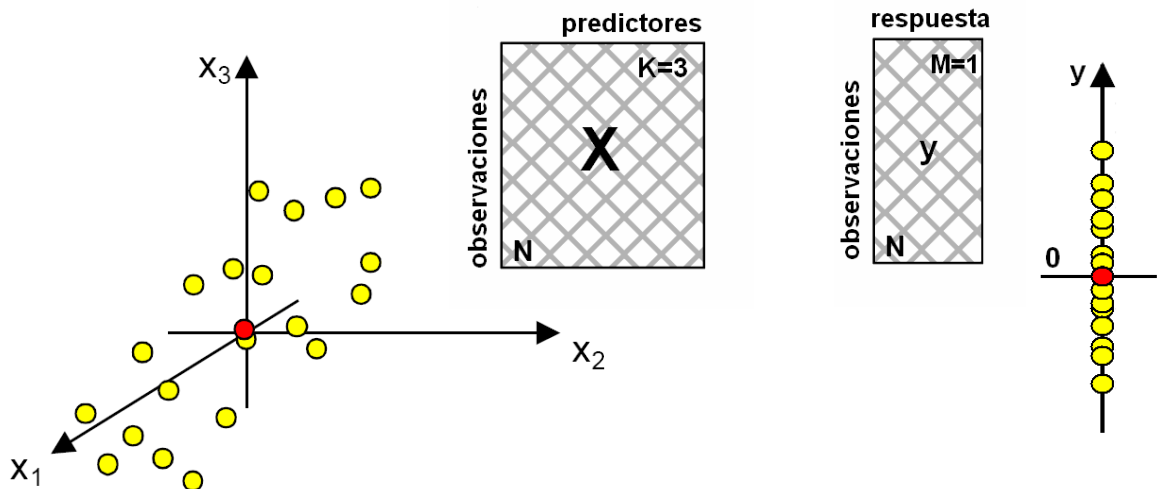
PLS es el acrónimo (en inglés) de *proyección a estructuras latentes por el método de mínimos cuadrados parciales*. El análisis fue creado alrededor de 1975 por Herman Wold y co-autores para la modelación de complicados set de datos en términos de conectar o encadenar información entre matrices de datos. Wold desarrollo una manera eficiente de estimar parámetros en estos modelos a través de un método (algoritmo) llamado NIPALS (Nonlinear Iterative Partial Least Squares) [Rencher, 2002; Eriksson et al., 1999].

PLS puede ser visto como una efectiva técnica de regresión general para modelar la asociación entre  $X$  e  $Y$ . PLS tiene propiedades similares a los modelos termodinámicos, formando modelos entre variables macroscópicas sin la necesidad de usar directamente en forma detallada la teoría microscópica en el modelo. Esto no quiere decir que sea inconsistente con la teoría

fenomenológica de un determinado proceso, ya que el análisis PLS usa la regularidad en la data para construir relaciones estructurales dentro de los modelos, en vez de formar funciones hipotéticas derivadas de alguna teoría fundamental.

Esta manera de generar modelos tal como el análisis PLS, funciona también en casos donde no existe una buena teoría de principios fundamentales y además sirve para proveer un chequeo independiente de la validez de los modelos derivados de la teoría existente en el campo de la aplicación en estudio.

A continuación, se dará una interpretación geométrica del análisis PLS de forma similar a la interpretación anterior para el análisis VFA. Por simplicidad se considerará la modelación en el caso de una respuesta única ( $M=1$ ) y tres variables de entrada predictoras ( $K=3$ ). La Figura 40 muestra el sistema de coordenadas y la representación gráfica de las observaciones escaladas y centradas tanto en el espacio  $X$  como en el espacio  $Y$ .



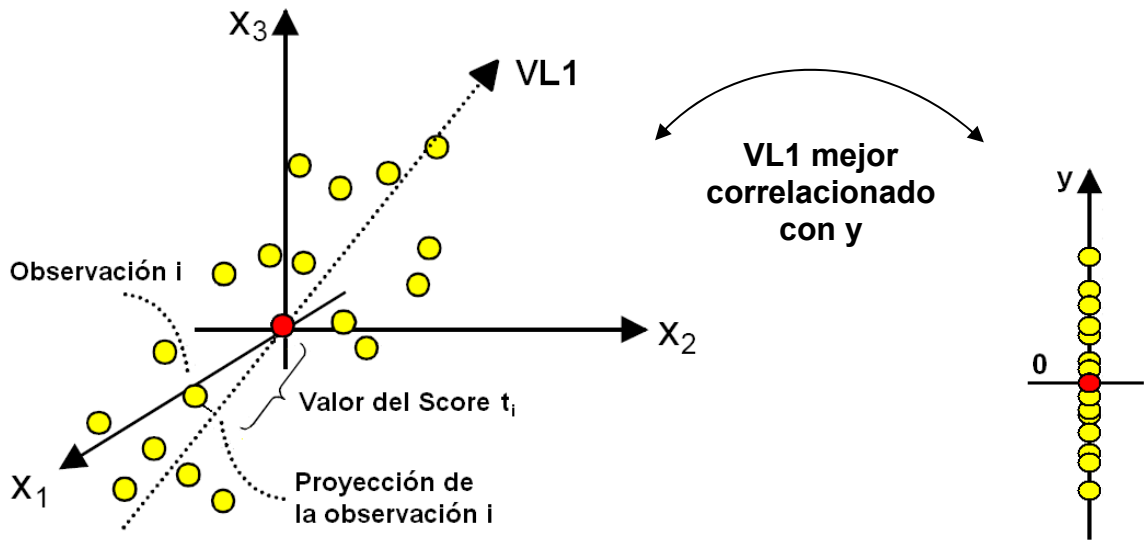
*Figura 40: Sistema de coordenadas para el espacio  $K$ -dimensional de  $X$  y para el espacio unidimensional de  $Y$ .*

Al igual que VFA cada observación puede ser representada como un punto en el espacio, sin embargo, la gran diferencia en PLS es que a cada observación de la tabla de datos le corresponde dos puntos; uno en el espacio X y otro en el espacio Y. Luego para una gran cantidad de observaciones se generan dos nubes de puntos, cada una ubicada en el correspondiente espacio dimensional. Una vez representado los espacios dimensionales de ambas matrices el problema a resolver es como descubrir la relación entre la forma en que las observaciones se agrupan en el espacio X predictor y la manera en que ellas se propagan en el espacio Y de respuestas.

La manera en que PLS relaciona la forma en que las observaciones se agrupan en el espacio X y la forma en que la(s) variable(s) de respuesta(s) se propaga(n) en el espacio Y, es como sigue. PLS al igual que VFA calcula componentes principales que indican la dirección de máxima variabilidad en la matriz X de datos, llamadas *Vectores Latentes (VL)*, pero escoge para el análisis sólo aquellas cuya covarianza con la variable de respuesta a predecir sea máxima. Así, PLS define variables latentes VL's que mejor describen la distribución de datos en el espacio X y a la vez predicen de mejor manera el valor de la variable de respuesta y en el espacio Y.

### **3.2.1. Cálculo del primer vector latente (VL1).**

El primer vector latente, geoméricamente se puede interpretar en el espacio X, como la línea que pasando por el origen mejor ajusta la distribución de la nube de puntos y a la vez correlaciona de mejor manera la variable **y** en el tiempo. Al igual que en VFA, la proyección de las observaciones sobre esta línea generan nuevas coordenadas llamadas scores  $t_i$ , tal como lo muestra la Figura 41.

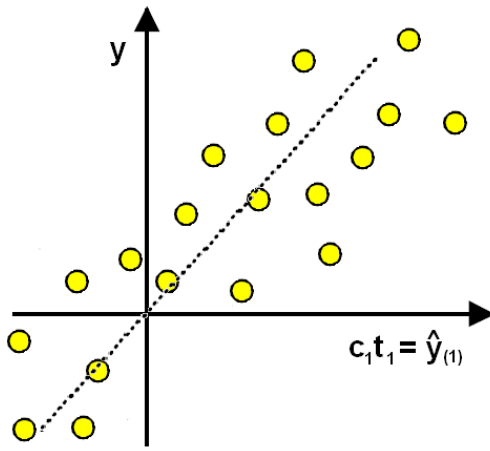


*Figura 41: Interpretación geométrica del primer VL. Es el componente principal que mejor describe la distribución de la variabilidad de la nube de puntos y a la vez predice de mejor manera la variable de respuesta  $y$ .*

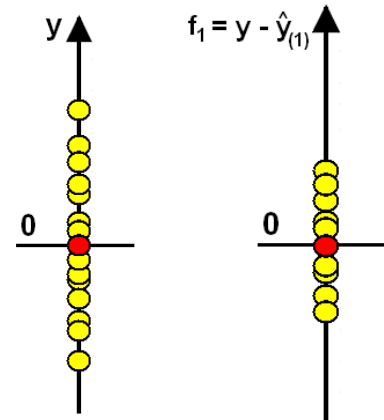
Los scores de todas las observaciones obtenidas desde el primer VL forman el primer vector score  $t_{i1}$ , el que refleja sólo la información en la matriz  $X$  original que es relevante para modelar y predecir la variable de respuesta  $y$ . Este vector score  $t_{i1}$ , puede ser usado para obtener una aproximación de la variable  $y$ , la que se denominará  $\hat{y}_{(1)}$ . Esto se lleva a cabo multiplicando el vector  $t_{i1}$  con un vector de peso o weight  $c_1$  de la matriz  $y$  de datos. El vector de pesos  $c_1$  de la data  $y$ , se obtiene a partir del algoritmo de NIPALS.

La Figura 42 muestra el gráfico de la data  $y$  observada versus su aproximación  $\hat{y}_{(1)}$  después del cálculo del primer vector latente. Las diferencias entre las respuestas observadas y estimadas después del primer vector latente se agrupan en un vector residual  $f_1$ , ver Figura 42.

Se observa que tiene una longitud menor que el vector  $y$ , esto significa que el primer vector latente VL1 explica gran parte de la variabilidad de la data  $y$ .



**Figura 42**



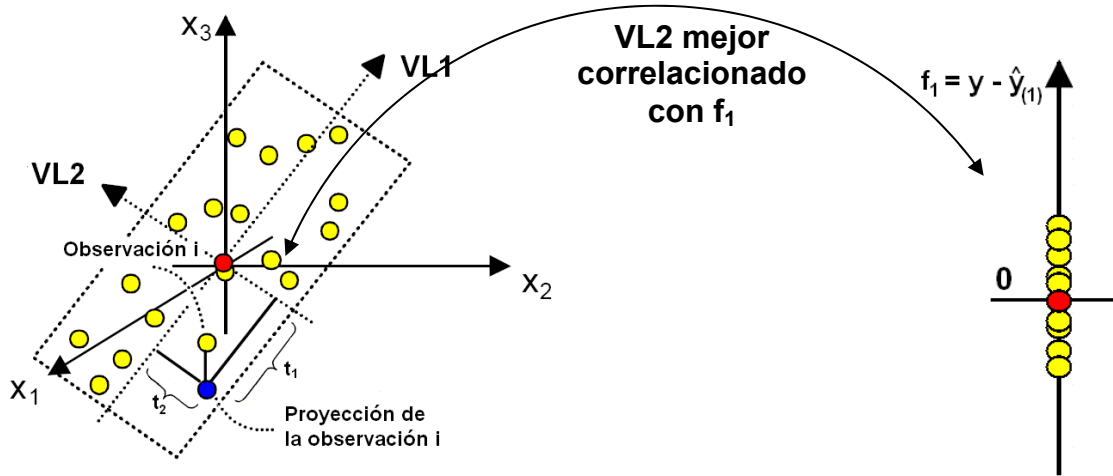
**Figura 43**

Fig. 42 (izquierda) Gráfico de la data  $y$  observada versus su aproximación  $\hat{y}_{(1)}$  estimada después del cálculo del primer VL1. La dispersión de puntos alrededor de la línea diagonal ideal permite estimar el nivel de ajuste del modelo PLS con el primer VL1. Fig. 43: (derecha) Comparación entre la data  $y$  observada y los residuos  $f_1$  resultados después de calcular el primer VL1. El vector residual  $f_1$  se obtiene substrayendo  $\hat{y}_{(1)}$  a  $y$ , y tiene una longitud menor que el vector  $y$ , por lo que se concluye que el primer VL explica gran parte de la variabilidad de la data  $y$ .

### 3.2.2. Cálculo del segundo vector latente (VL2)

Normalmente un vector latente es insuficiente para representar adecuadamente la variabilidad de la respuesta  $y$ . Es posible aumentar la capacidad predictiva del Modelo PLS expandiendo el modelo con un segundo vector latente.

El VL2 se interpreta como una línea que pasa través del origen en el espacio  $X$ , mejor describe la distribución de la variabilidad de la nube de puntos en forma ortogonal al primer VL1 y entrega la mejor correlación con los residuos  $f_1$  resultantes después de la substracción de la primera aproximación  $\hat{y}_{(1)}$  a la data  $y$  observada, tal como lo muestra la Figura 44.



**Fig. 44**

**Fig. 45**

*Interpretación geométrica del segundo VL. Es el componente principal que mejor describe la distribución de la variabilidad de la nube de puntos en forma ortogonal a VL1 y a la vez predice de mejor manera el vector de residuos  $f_1$ .*

Para obtener los valores de los scores de las observaciones una vez calculado el segundo vector latente, todas las observaciones deben ser proyectadas sobre el plano definido por VL1 y VL2, tal como lo muestra la parte izquierda de la Figura 44. Una vez calculado el vector  $t_{i2}$ , en forma análoga al caso anterior, es posible aproximar el vector de residuos  $f_1$ , a través de la multiplicación del vector score  $t_{i2}$  con el nuevo vector de pesos (weight)  $c_2$  de la matriz  $y$  de datos. La Figura 45 muestra el gráfico del vector de residuos calculado  $f_1$  versus su aproximación calculada después del cálculo del segundo componente principal. Las diferencias entre los residuos calculados y su aproximación después del primer vector latente se agrupan en un vector residual  $f_2$ , ver Figura 46. Se observa que el vector de residuos  $f_2$  tiene una longitud menor que el vector de residuos original  $f_1$ , lo que significa que el segundo vector latente explica gran parte de los residuos que el primer vector latente no pudo explicar, aumentando así el poder predictivo del Modelo PLS.

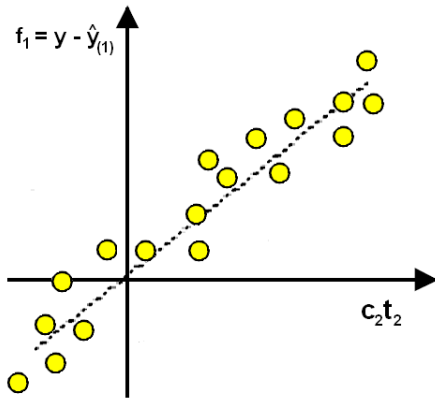


Fig. 46

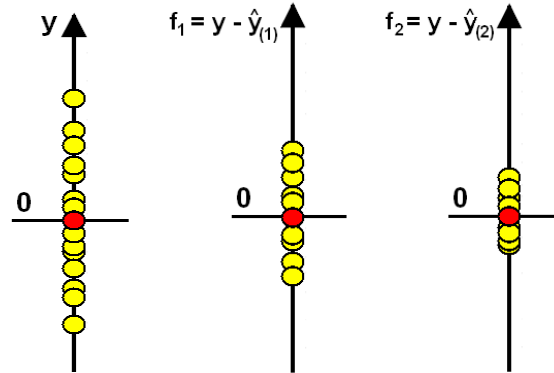


Fig. 47

Fig. 46:(izquierda) Gráfico del vector de residuos  $f_1$  obtenido después del cálculo del primer vector latente versus su aproximación calculada después del segundo vector latente. Fig.47: (derecha) Comparación entre las longitudes de la data  $y$  observada, la aproximación a los residuos  $f_1$  y la aproximación a los residuos  $f_2$ . Se observa que el vector  $f_2$  tiene una longitud menor que el vector  $f_1$ , lo que significa que el segundo vector latente explica gran parte de los residuos que el primer vector latente no pudo explicar.

Finalmente el análisis PLS permite combinar la capacidad predictiva de los vectores scores  $t_{i1}$  y  $t_{i2}$  para modelar y predecir la variable de respuesta  $y$ . Una estimación de  $y$  después de dos vectores latentes, denominada  $\hat{y}_{(2)}$ , se obtiene a través del cálculo de  $c_1t_1 + c_2t_2$ , tal como lo muestra la Figura 48.

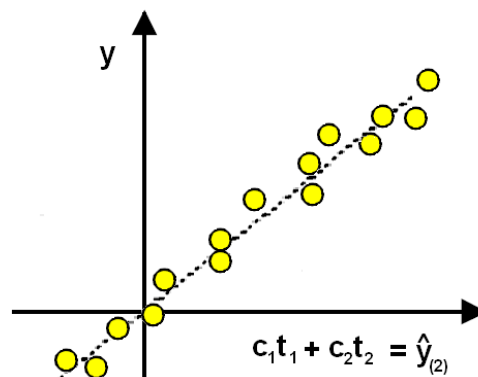


Fig. 48



Una estimación de  $\mathbf{y}$  después de dos vectores latentes  $\hat{\mathbf{y}}_{(2)}$ . Es obtenida por el cálculo de  $c_1t_1 + c_2t_2$ . Cuando la predicción entre las variables estimada y observada de la data  $\mathbf{y}$  es buena, la dispersión de los puntos alrededor de la línea punteada es pequeña. Si se compara con la figura 48 se observa que la dispersión es menor con dos vectores latentes.

PLS extrae factores de variabilidad que, al igual que VFA, son funciones de las variables originales. Pero PLS extrae factores tanto de X como de Y

Las matrices X e Y se modelan de la siguiente manera (datos normalizados y escalados):

$$\mathbf{X}=\mathbf{TP}'+\mathbf{E} \quad (t_1p_1'+t_2p_2'+\dots+\mathbf{E}) \quad \dots\dots\dots (68)$$

$$\mathbf{Y}=\mathbf{UQ}'+\mathbf{F} \quad (u_1q_1'+u_2q_2'+\dots+\mathbf{F}) \quad \dots\dots\dots (69)$$

T y U corresponden a las matrices de scores

P y Q corresponden a las matrices de loadings

La relación entre X e Y es una regresión lineal entre U y T

$$\mathbf{U}=\mathbf{bT} \quad (u_j=b_jt_j) \quad \dots\dots\dots (70)$$

El primer set de loadings vectors  $p_1$  y  $q_1$  se obtienen **maximizando la covarianza entre X e Y.**

La proyección de los datos X e Y en  $p_1$  y  $q_1$  entregan el primer set de scores  $t_1$  y  $u_1$  respectivamente ( $u_1=t_1b_1$ )

El proceso se repite para X' e Y' recursivamente:

$$\mathbf{X}'=\mathbf{X}-t_1p_1' \quad \mathbf{Y}'=\mathbf{Y}-u_1q_1' \quad \dots\dots\dots (71)$$

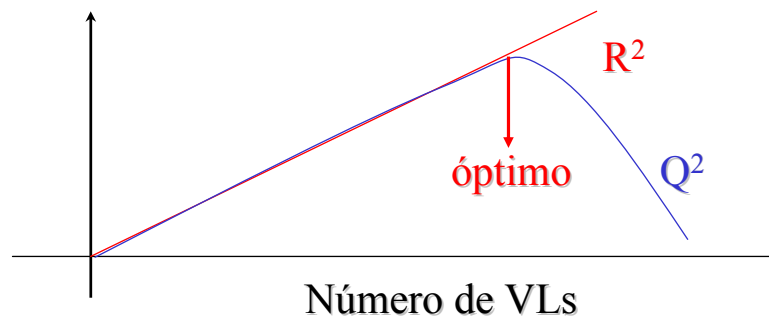
PLS descompone un problema de regresión multivariante en una serie de regresiones univariantes.

Cada regresión univariante define un vector latente (VL), definidos en función de mayor a menor covarianza entre X e Y. Las VLs son independientes entre si y son una importante fuente de información sobre la relación entre X e Y, a través de los score plots y loading plots.

El número de VLs debe ser escogido de acuerdo a trade-off entre identificación y predicción

Índices para escoger VLs:

- $R^2$  (suma cuadrática de errores de identificación)
- $Q^2$  (suma cuadrática de errores de predicción)



*Fig. 49: La primera VLs es la más importante y explica la tendencia de Y. El resto de las VLs explica las variaciones sobre esa tendencia*

Una vez definida la base de VLs, se puede obtener una matriz **B** (dependiente de las matrices **P**, **Q** y una matriz de pesos **W**) tal que  $Y = X B$

Las filas de B son los coeficientes de regresión de cada  $Y_i$ . **El modelo PLS se convierte en un modelo lineal de fácil uso para predicción** (datos normalizados):

$$Y_1 = X_1 B_{11} + X_2 B_{12} + \dots + X_K B_{1K} \quad \dots \quad (72)$$

Nota: los coeficientes B son generalmente **no** independientes

La información que se puede obtener de los análisis PLS son:

1. Loading plot: coeficientes de VLs
2. Coeficientes de regresión
3. Score plot t/t, t/u
4. Gráficos de residuos: DModX y DModY
5. Predicción de datos nuevos

### ***Loading Plot (Matriz de Pesos W)***

Las W son la nueva base de X, son combinaciones lineales de las variables de entrada. El gráfico de coeficientes de W nos entrega información sobre la correlación entre las variables X y las variables Y. A diferencia de VFA, en donde VF1 es la dirección de máxima variabilidad, en PLS, el vector W1 es la dirección de X que mejor explica Y. Las demás W explican los residuos.

Los pesos (valores de los vectores  $W_i$ ) son esenciales para entender:

- Cuales variables X son importantes (las cantidades de  $W_i$  grandes en valor absoluto)
- Cuales variables X entregan la misma información o partes repetidas (dos o más variables que tienen coeficientes parecidos para una o más  $W_i$ )

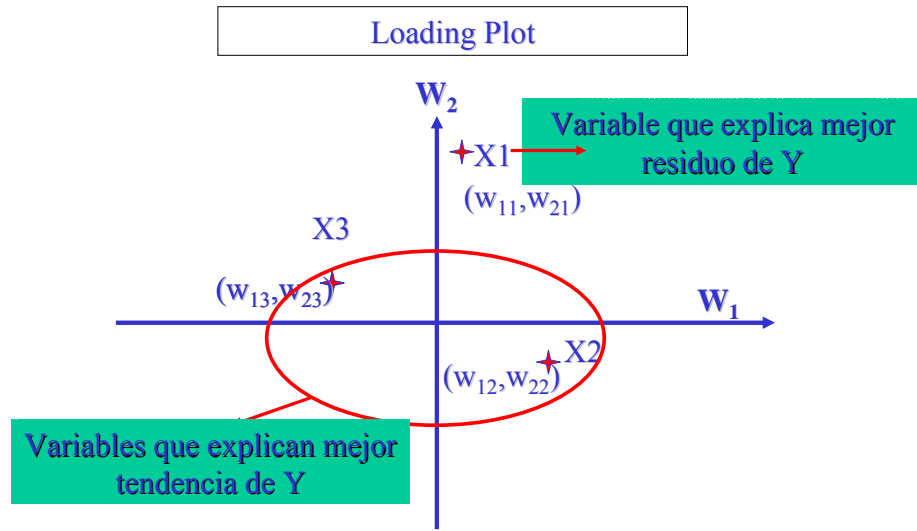


Figura 50

**Coefficientes de Regresión**

El modelo PLS, mediante una transformación de las VLs se convierte en:

$$Y1 = X1 B11 + X2 B12 + \dots + XK B1K \quad \dots \dots \dots (73)$$

Los coeficientes  $B1i$  resume la influencia total de la variable  $Xi$  en  $Y1$   
En general no son independientes entre si, a menos que se cumplan ciertas condiciones especiales.

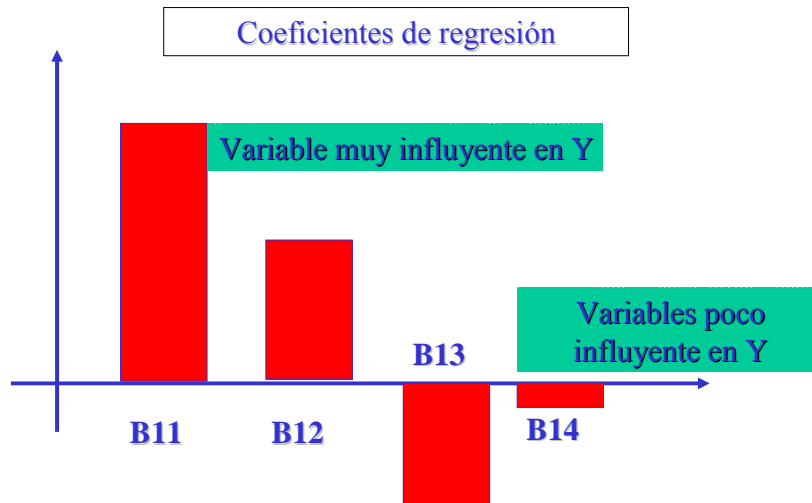


Figura 51

### Score plot (t/t)

Con el fin de identificar clusters y outliers se puede graficar los  $t_i$  proyecciones de  $X$  en las VLs. Muestras cercanas entre si corresponden a condiciones parecidas de proceso, muestras lejanas a condiciones distintas

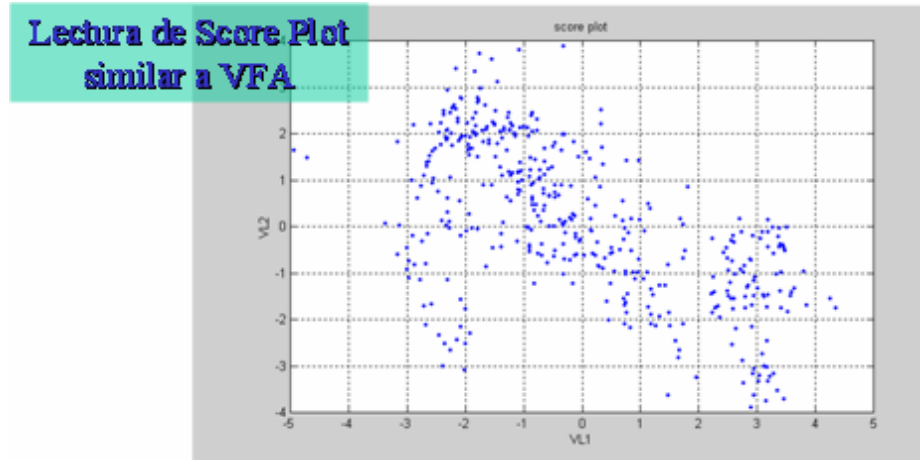


Fig. 52

### Prediction vs Real Output: Score plot t/u

A través de este gráfico también es posible detectar clusters y outliers. La relación entre  $t_i$  y  $u_i$  en una identificación óptima es lineal (si la distribución de puntos no es lineal, es necesario hacer transformaciones a las variables).

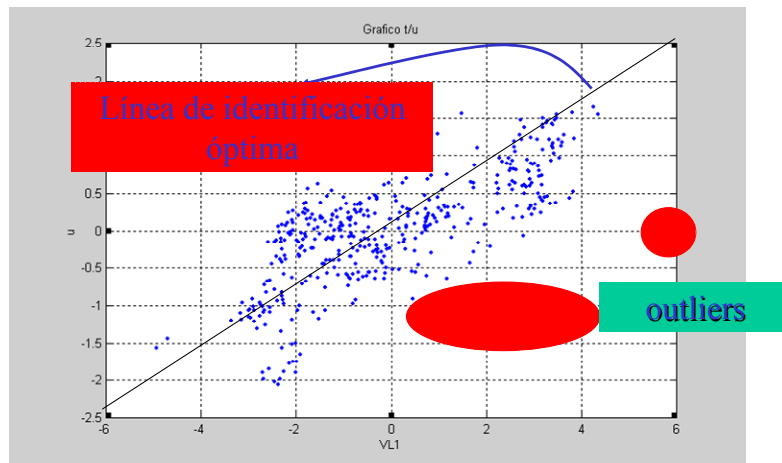


Figura 53

**Prediction vs Real Output: Score plot t/u**

Hay casos en que es clara la necesidad de dividir el set de datos (existencia de sub-patrones)



**Figura 54**

**Índices de Residuos**

Existen dos expresiones para medir el ajuste del modelo PLS:

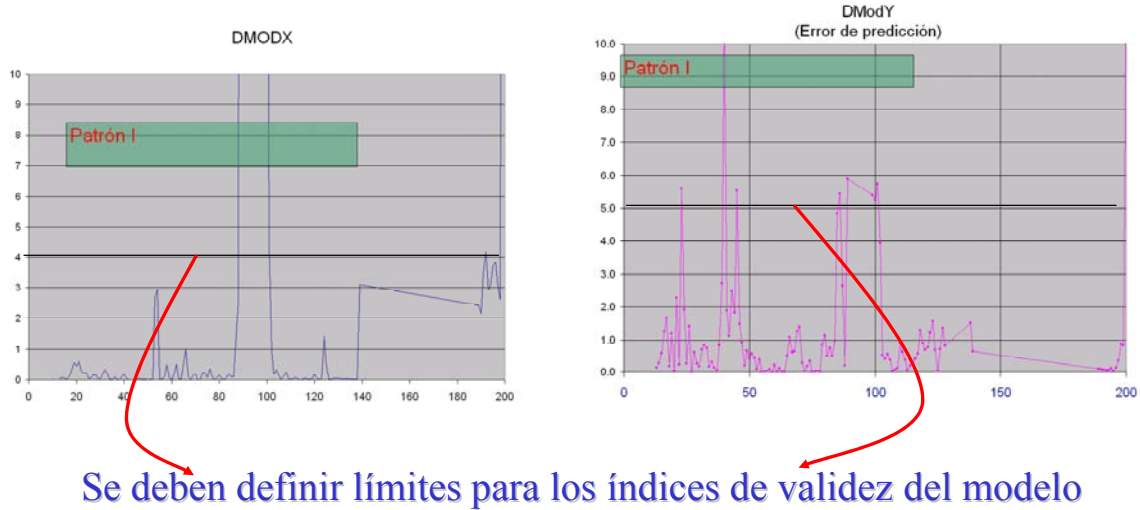
- DModY: error de predicción de Y

$$DModY(k) = (Y(k) - Y_{pred}(k))^2 \quad \dots\dots\dots (74)$$

- DModX: error de modelación de X

$$DModX(k) = \sum \text{fila}(k) \frac{(X - XP^T P)^2}{No.(VarX) - No.(VLs)} \quad \dots\dots\dots (75)$$

Estos indicadores entregan información complementaria sobre el ajuste del modelo.



*Figura 55*

El conocimiento de las propiedades físico - químicas de los procesos y lazos de control existentes en la planta cobra especial relevancia a la hora de estudiar la relación causa - efecto.

De esta forma, puede establecerse un conjunto de variables que permite “explicar” el comportamiento de otra. Las relaciones matemáticas que vinculan a la variable en estudio con su set de variables explicativas se conoce como “modelo”.

En el caso de un “Modelo de Regresión Lineal”, se considera que el valor de la variable en un instante dado de tiempo se puede obtener al sumar los aportes ponderados de las variables explicativas.

$$Y = a_1 \cdot X_1 + a_2 \cdot X_2 + \dots + a_n \cdot X_n \quad \dots \dots \dots (76)$$

Los valores de los coeficientes se calculan a partir de un conjunto de datos que resume la operación “normal” del proceso. El algoritmo que los calcula intenta disminuir el “Error Cuadrático” para ese set de datos.

El error cuadrático corresponde a la suma de todos los errores de predicción que se obtienen al fijar en algún valor los parámetros  $a_1, a_2, \dots, a_n$ , elevados al cuadrado.

$$EC = \sum (Y - Y_{pred})^2 \dots\dots\dots (77)$$

Se elevan al cuadrado para evitar que los errores positivos se cancelen con los negativos. Mientras mayor es el error cuadrático, peor es la predicción. Este método funciona muy bien cuando las variables explicativas son independientes y tengan un vínculo real con la variable estudiada, en caso contrario son poco coherentes.

El modelo que es definido a través de los métodos anteriores es estático, es decir depende de los valores actuales de las variables explicativas y no de la historia anterior.

Cuando se desea incorporar la componente temporal en un modelo, se introduce el concepto de “Modelo Dinámico”. Para esto, se debe agregar, como variables auxiliares, las variables originales retardadas en 1 o más muestras.

$$X^*=[X_1 \ X_{1t-1} \ X_{1t-2} \ X_{1t-3} \ X_2 \ X_{2t-1} \ \dots] \dots\dots\dots (78)$$

$$Y=X^*B \dots\dots\dots (79)$$

Las desviaciones con respecto a modelos de referencia (dinámicos o estáticos) generan RESIDUOS. Los residuos corresponden a causas no explicadas por los modelos, luego son fuente de información (Concepto de Innovación).



### 3.3. MODELOS PREDICTIVOS

El molino de estudio fue el Molino Primario 1C y como se aprecia en el anexo D, las principales variables (que son expresadas mediante Tags) en un circuito de molienda y que fueron las que se emplearon para la generación de los modelos predictivos son:

**TABLA N° 17: TAGS DEL MOLINO 1C**

<i>Variable</i>	<i>Tags</i>
Tonelaje ( TMH)	WIC1043-1
%Malla +65 O/F	AI1103-35
%Sólidos O/F	D11103-35
Presión Ciclonos (PSI)	PI1173-01
Amperaje de Bomba ciclonos (A)	II1173-05
Nivel Cajón Descarga %	LIC1163-01
Carga Circulante %	XY1103-1
Potencia Motor (MW)	JI1103-1
Agua de Alimentación – Agua Colectores (GPM)	FIC1103-3
Agua de Alimentación – Agua Fresca (GPM)	FIC1103-1
Agua de Alimentación – Agua Recuperada (GPM)	FIC1103-2
Agua de Descarga – Agua Recuperada (GPM)	FIC1163-05

Se agruparon las variables por semanas pudiendo obtener hasta un total de 1008 datos por semana.

Se tomo todo una data de 11 semanas desde el Mes de Enero hasta el mes de Marzo del 2006, con el objetivo de examinar los principales parámetros estadísticos de las variables. Haciendo uso de los Gráficos de Control se encontró la media y los rangos de estas variables que son una de las condiciones de ingreso en el SCAN.

Un resumen de los resultados encontrados se muestra en la Tabla N° 18

**TABLA N° 18: ESTADISTICA DEL MOLINO 1C**

WEEK	MEDIA	DES. EST	RANGO	MINIMO	MAXIMO	COUNT
1	255,37	17,32	71,84	223,63	295,47	144
2	286,20	18,62	109,76	224,64	334,40	974
3	274,95	22,60	105,38	225,73	331,10	977
4	291,78	17,55	82,50	245,48	327,98	998
5	285,43	14,63	75,45	233,00	308,45	992
6	285,63	18,06	85,35	235,73	321,08	994
7	279,35	18,46	88,47	232,24	320,70	958
8	279,22	14,93	63,75	247,89	311,64	918
9	265,39	17,97	78,05	220,94	298,99	894
10	250,80	17,00	78,56	215,14	293,70	876
11	236,41	17,93	84,54	200,19	284,73	909

A continuación se determinó los Límites de Control mediante el uso de dos métodos:

**Metodo 1**

LIMITES DE CONTROL PARA X

UCL =  $X_p + A_2 \cdot R_p$

Línea Central =  $X_p$

LCL =  $X_p - A_2 \cdot R_p$

ara n = 10 ----->  $A_2 = 0,300$

LIMITES DE CONTROL PARA R

UCL =  $D_4 \cdot R_p$

Línea Central =  $X_p$

LCL =  $D_3 \cdot R_p$

ra n = 10 ----->  $D_4 = 1,78$

$D_3 = 0,223$

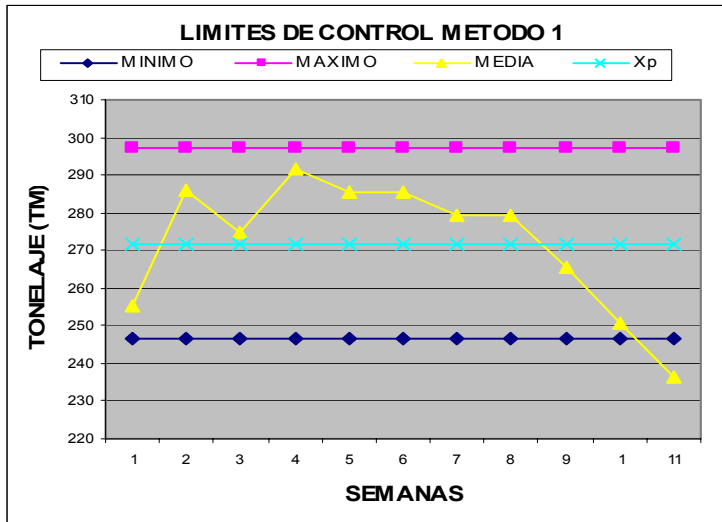
**Metodo 2**

UCL =  $X_p + 3 \cdot \text{DESV.EST} / \text{RAIZ}(N)$

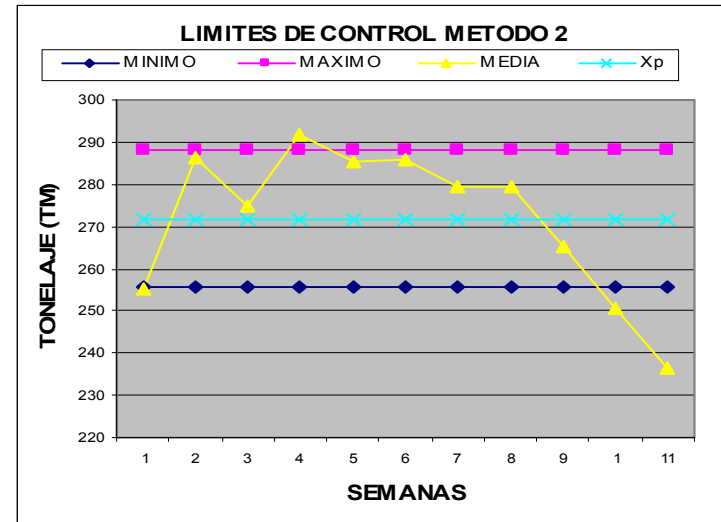
Línea Central =  $X_p$

LCL =  $X_p - 3 \cdot \text{DESV.EST} / \text{RAIZ}(N)$

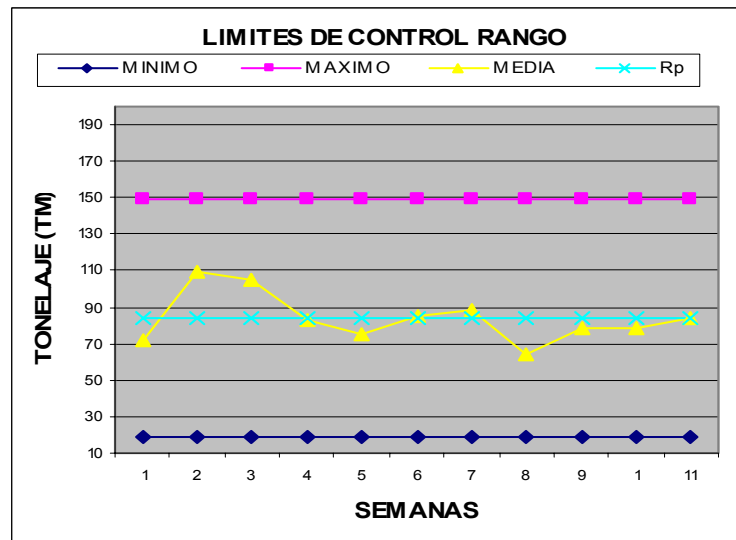
3.3.1. DETERMINACION DE LOS LÍMITES DE CONTROL



(a)



(b)



(c)

FIGURA 56

### 3.3.2. DETERMINACION DE LA CAPACIDAD DEL PROCESO:

La capacidad del proceso es determinada mediante la siguiente ecuación:

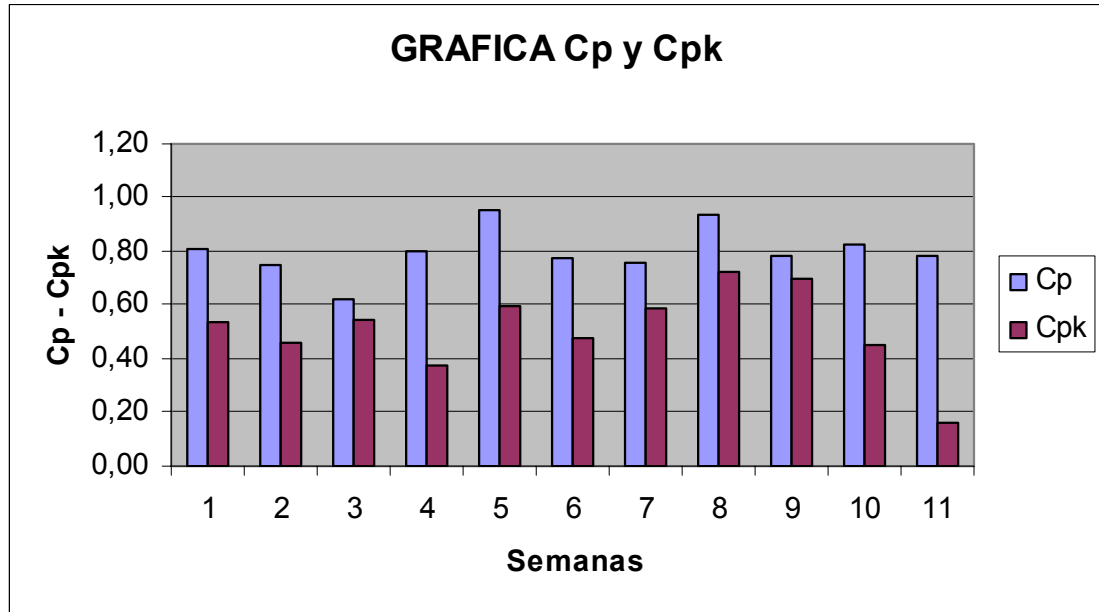
$$C_P = \frac{USL - LSL}{6\sigma}$$

**TABLA N° 18:**  
**DETERMINACIÓN DE LA CAPACIDAD DEL PROCESO**

<i>WEEK</i>	<i>MEDIA</i>	<i>DES. EST</i>	<i>RANGO</i>	<i>LSL</i>	<i>USL</i>
<i>PROM</i>	<b>271,87</b>	17,73	<b>83,97</b>	<b>227,69</b>	<b>311,66</b>

**TABLA N° 19: Cp y Cpk**

<i>WEEK</i>	<i>Cp</i>	<i>Cpk</i>
1	0,81	0,53
2	0,75	0,46
3	0,62	0,54
4	0,80	0,38
5	0,96	0,60
6	0,77	0,48
7	0,76	0,58
8	0,94	0,72
9	0,78	0,70
10	0,82	0,45
11	0,78	0,16
<b><i>PROM</i></b>	<b>0,80</b>	<b>0,51</b>



*Figura 57*

### 3.3.3. MODELO PREDICTIVO %MALLA +65 COMO CONTROL DE CALIDAD:

Una de las condiciones para la elaboración de los modelos con el SCAN es tener a una variable como el parámetro de calidad, dicha variable según el estudio fue el % de Malla +65. Para ello se estableció una serie de muestreos para calibrar el PSM (Particle Monitor Size) del Molino 1C.

El método de calibración consistía en tomar una muestra del PSM y anotar la lectura de este junto a la hora exacta del muestreo para rescatar el valor de las principales variables de operación en el Molino en el PI; luego se determinaba el % de malla +65 en el Laboratorio y se realizaban los ajustes correspondientes.

Los resultados obtenidos fueron los siguientes:

**CONTROL ESTADISTICO MULTIVARIANTE  
EN CIRCUITO DE MOLIENDA**

**CRISTHIAN CURO LORO**

#	MOLINO	FECHA	Hora	Tonelaje	Malla	Laboratorio	PSM	%Conf.	Diferencia	#	MOLI	FECHA	Hora	Tonelaje	Malla	Laboratorio	PSM	%Conf.	Diferencia
1	1C	14/08/2006	10:58	271	m+65	20,9062	21,8	95,725	-0,894	1	1C	14/08/2006	10:58	271	%SOL	39,5700	39,90	99,166	-0,330
2	1C	14/08/2006	11:05	271	m+65	20,2208	21,7	92,685	-1,479	2	1C	14/08/2006	11:05	271	%SOL	39,2540	39,90	98,354	-0,646
3	1C	14/08/2006	14:06	273	m+65	20,7679	21,4	96,956	-0,632	3	1C	14/08/2006	14:06	273	%SOL	37,6658	39,70	94,599	-2,034
4	1C	14/08/2006	14:09	274	m+65	19,3967	21,6	88,641	-2,203	4	1C	14/08/2006	14:09	274	%SOL	40,3211	39,80	98,708	0,521
5	1C	15/08/2006	9:38	275	m+65	20,9472	22,2	94,019	-1,253	5	1C	15/08/2006	9:38	275	%SOL	38,4894	39,50	97,374	-1,011
6	1C	15/08/2006	9:42	276	m+65	21,4085	22	97,237	-0,592	6	1C	15/08/2006	9:42	276	%SOL	38,2808	39,40	97,076	-1,119
7	1C	15/08/2006	10:20	280	m+65	21,5164	22,9	93,570	-1,384	7	1C	15/08/2006	10:20	280	%SOL	37,8963	39,50	95,768	-1,604
8	1C	15/08/2006	10:21	281	m+65	21,8406	22,6	96,523	-0,759	8	1C	15/08/2006	10:21	281	%SOL	37,0999	39,50	93,531	-2,400
9	1C	15/08/2006	13:17	291	m+65	23,3596	23,4	99,827	-0,040	9	1C	15/08/2006	13:17	291	%SOL	39,3051	39,80	98,741	-0,495
10	1C	15/08/2006	13:20	293	m+65	22,486	23,4	95,935	-0,914	10	1C	15/08/2006	13:20	293	%SOL	39,9134	39,90	99,966	0,013
11	1C	15/08/2006	14:09	282	m+65	23,429	22,7	96,888	0,729	11	1C	15/08/2006	14:09	282	%SOL	38,9733	39,70	98,135	-0,727
12	1C	15/08/2006	14:12	285	m+65	22,2636	22,6	98,489	-0,336	12	1C	15/08/2006	14:12	285	%SOL	38,8589	39,60	98,093	-0,741
13	1C	16/08/2006	9:24	275	m+65	22,6329	22,8	99,262	-0,167	13	1C	16/08/2006	9:24	275	%SOL	39,1294	39,80	98,286	-0,671
14	1C	16/08/2006	9:26	276	m+65	21,654	22,9	94,246	-1,246	14	1C	16/08/2006	9:26	276	%SOL	38,7540	39,90	97,043	-1,146
15	1C	16/08/2006	9:58	283	m+65	23,6335	23,2	98,166	0,434	15	1C	16/08/2006	9:58	283	%SOL	39,6897	40,10	98,966	-0,410
16	1C	16/08/2006	10:02	285	m+65	23,2728	23,6	98,594	-0,327	16	1C	16/08/2006	10:02	285	%SOL	39,8820	40,10	99,453	-0,218
17	1C	17/08/2006	9:42	294	m+65	21,8387	22,2	98,346	-0,361	17	1C	17/08/2006	9:42	294	%SOL	38,2486	39,90	95,682	-1,651
18	1C	17/08/2006	9:44	295	m+65	21,1337	22,1	95,428	-0,966	18	1C	17/08/2006	9:44	295	%SOL	38,2977	39,80	96,077	-1,502
19	1C	17/08/2006	10:31	295	m+65	21,1839	21,8	97,092	-0,616	19	1C	17/08/2006	10:31	295	%SOL	38,3294	39,70	96,424	-1,371
20	1C	17/08/2006	10:34	296	m+65	21,3437	22,1	96,457	-0,756	20	1C	17/08/2006	10:34	296	%SOL	38,1684	39,80	95,725	-1,632
21	1C	17/08/2006	13:12	300	m+65	21,632	22,1	97,837	-0,468	21	1C	17/08/2006	13:12	300	%SOL	38,3372	39,80	96,184	-1,463
22	1C	17/08/2006	13:14	301	m+65	20,6575	21,7	94,953	-1,043	22	1C	17/08/2006	13:14	301	%SOL	38,3482	39,70	96,475	-1,352
23	1C	17/08/2006	14:05	301	m+65	22,4981	22,3	99,119	0,198	23	1C	17/08/2006	14:05	301	%SOL	38,7653	39,70	97,589	-0,935
24	1C	17/08/2006	14:08	301	m+65	22,0549	22,3	98,889	-0,245	24	1C	17/08/2006	14:08	301	%SOL	38,6645	39,70	97,322	-1,035
25	1C	21/08/2006	11:02	232	m+65	19,2981	20,3	94,808	-1,002	25	1C	21/08/2006	11:02	232	%SOL	35,1282	39,20	88,409	-4,072
26	1C	21/08/2006	11:05	207	m+65	19,0679	20,2	94,063	-1,132	26	1C	21/08/2006	11:05	207	%SOL	34,6928	39,10	87,296	-4,407
27	1C	21/08/2006	11:38	207	m+65	14,3749	18,5	71,303	-4,125	27	1C	21/08/2006	11:38	207	%SOL	33,3368	39,00	83,012	-5,663
28	1C	21/08/2006	11:41	207	m+65	16,8876	18,9	88,084	-2,012	28	1C	21/08/2006	11:41	207	%SOL	35,1043	39,20	88,333	-4,096
29	1C	21/08/2006	14:55	257	m+65	24,414	21,4	87,655	3,014	29	1C	21/08/2006	14:55	257	%SOL	42,9146	40,40	94,141	2,515
30	1C	21/08/2006	14:58	258	m+65	16,7208	21,2	73,212	-4,479	30	1C	21/08/2006	14:58	258	%SOL	35,1162	40,30	85,238	-5,184
31	1C	21/08/2006	15:25	168	m+65	20,763	21,2	97,895	-0,437	31	1C	21/08/2006	15:25	168	%SOL	39,8275	40,20	99,065	-0,372
32	1C	21/08/2006	15:28	169	m+65	20,8302	20,6	98,895	0,230	32	1C	21/08/2006	15:28	169	%SOL	38,1655	40,00	95,193	-1,835
33	1C	22/08/2006	10:10	255	m+65	22,6744	22,1	97,467	0,574	33	1C	22/08/2006	10:10	0	%SOL	38,2845	39,90	95,780	-1,616
34	1C	22/08/2006	10:13	255	m+65	23,5081	22,3	94,861	1,208	34	1C	22/08/2006	10:13	0	%SOL	37,0276	40,00	91,972	-2,972
35	1C	22/08/2006	11:21	230	m+65	20,1592	20,6	97,813	-0,441	35	1C	22/08/2006	11:21	0	%SOL	35,8665	39,20	90,706	-3,333
36	1C	22/08/2006	11:24	225	m+65	19,6305	20,5	95,571	-0,869	36	1C	22/08/2006	11:24	0	%SOL	35,3177	39,20	89,007	-3,882
37	1C	29/08/2006	9:58	298	m+65	22,4383	23,4	95,714	-0,962	37	1C	29/08/2006	9:58	298	%SOL	38,5184	38,60	99,788	-0,082
38	1C	29/08/2006	10:01	297	m+65	22,4516	23,1	97,112	-0,648	38	1C	29/08/2006	10:01	297	%SOL	38,0352	38,60	98,515	-0,565
39	1C	29/08/2006	11:29	307	m+65	21,1949	23,4	89,596	-2,205	39	1C	29/08/2006	11:29	307	%SOL	38,6200	38,70	99,793	-0,080
40	1C	29/08/2006	11:32	302	m+65	21,8572	23,1	94,314	-1,243	40	1C	29/08/2006	11:32	302	%SOL	38,4008	38,80	98,960	-0,399
41	1C	31/08/2006	10:40	319	m+65	21,5908	22,2	97,178	-0,609	41	1C	31/08/2006	10:40	319	%SOL	39,1869	39,10	99,778	0,087
42	1C	31/08/2006	10:43	319	m+65	21,7466	23	94,236	-1,253	42	1C	31/08/2006	10:43	319	%SOL	38,8036	38,70	99,733	0,104
43	1C	31/08/2006	11:49	322	m+65	20,8879	23	89,888	-2,112	43	1C	31/08/2006	11:49	322	%SOL	38,9152	38,80	99,704	0,115
44	1C	31/08/2006	11:51	322	m+65	20,4242	22,5	89,837	-2,076	44	1C	31/08/2006	11:51	322	%SOL	38,5249	38,60	99,805	-0,075
<b>MEDIA</b>						<b>21,205</b>	<b>22,020</b>							<b>MEDIA</b>					
<b>DESV. STD (S)</b>						<b>1,887</b>	<b>1,151</b>							<b>DESV. STD (S)</b>					
<b>CV</b>						<b>8,899</b>	<b>5,227</b>							<b>CV</b>					

Para la elaboración de los modelos se utilizaron tres herramientas:

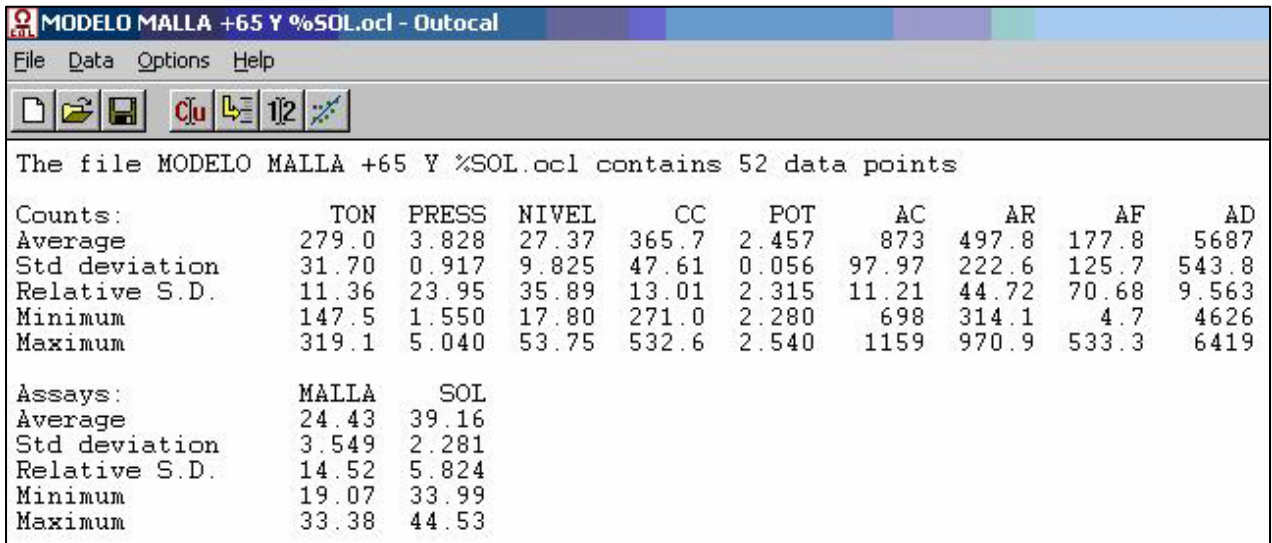
1. El software Outocal
2. El SCAN
3. El Excel

Los modelos obtenidos fueron los siguientes:

**Empleando el software Outocal:**

Se realizaron modelos tanto para el %Malla +65 y %Sólidos del O/F, ingresando un total de 52 puntos de muestreo.

A continuación se muestra la ventana de presentación en donde se



aprecia a las variables del modelo y sus principales características estadísticas:

**Figura 58**

MODELO %MALLA +65:

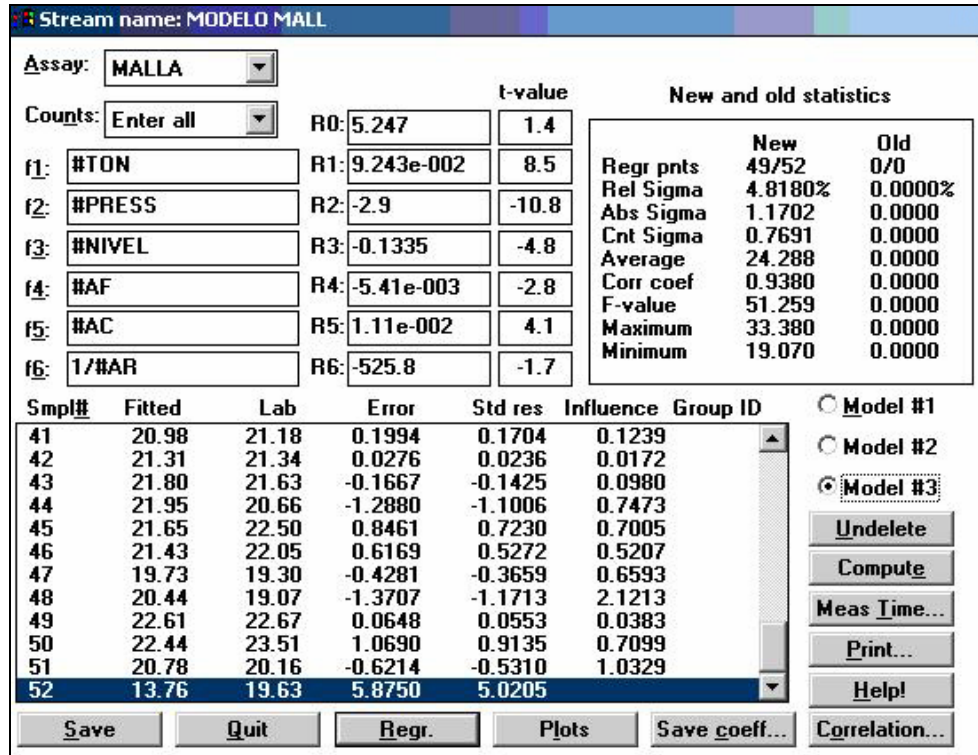


Figura 59

TABLA N° 20:

MODELO %MALLA +65 - OUTOCAL

VARIABLES DEL MODELO			Coeficientes	t-value
	Intercepto		5,247	1,43
WIC1043-1	Tonelaje	TMH	0,092	8,46
PI1173-01	Presión de Bomba de ciclones	psi	-2,900	-10,78
LIC1163-01	Nivel del cajón	%	-0,134	-4,84
FIC1103-2	Agua Fresca de Alimentación	LPM	-0,005	-2,76
FIC1163-3	Agua de Colectores de Alimentación	LPM	0,011	4,13
1/FIC1103-1	Agua Recuperada	LPM	-525,800	-1,75



**TABLA N°21:**  
**RESULTADOS ESTADÍSTICOS DEL MODELO %MALLA +65 - OUTOCAL**

<b>Puntos de Regresión</b>	49
<b>Coefficiente de Correlación</b>	0,94
<b>F-value</b>	51,26
<b>Promedio</b>	24,3
<b>Máximo</b>	33,4
<b>Mínimo</b>	19,1

Con lo que el modelo quedo de la siguiente forma:

$$\%Malla + 65 = +5.247 + 9.243E - 002 * WIC1043 - 1 + - 2.9 * PI1173 - 01 + - 0.1335 * LIC1163 - 01 - 5.41E - 003 * FIC1103 - 2 + 1.11E - 002 * FIC1103 - 3 + - 525.8 * (1 / FIC1103 - 1)$$

**MODELO %SÓLIDOS O/F:**

De la misma forma se realizó un modelo para el %Sólidos del Overflow, obteniendo el siguiente modelo:

**TABLA N° 22: MODELO %SÓLIDOS O/F - OUTOCAL**

<b>VARIABLES DEL MODELO</b>			<b>Coefficientes</b>	<b>t-value</b>
	<i>Intercepto</i>		36,254	12,56
<b>WIC1043-1</b>	<i>Tonelaje</i>	<i>MH</i>	0,070	9,59
<b>PI1173-01</b>	<i>Presión de Bomba de ciclones</i>	<i>psi</i>	-0,708	-1,98
<b>FIC1103-2</b>	<i>Agua Fresca de Alimentación</i>	<i>LPM</i>	-0,006	-3,82
<b>FIC1103-1</b>	<i>Agua Recuperada de Alimentación</i>	<i>LPM</i>	-0,002	-2,18
<b>FIC1163-05</b>	<i>Agua Recuperada de Dilución</i>	<i>LPM</i>	-0,002	-3,28

**TABLA N° 23: RESULTADOS ESTADÍSTICOS  
DEL MODELO % SOLÍDOS O/F - OUTOCAL**

<b>Puntos de Regresión</b>	50
<b>Coefficiente de Correlación</b>	0,90
<b>F-value</b>	48,00
<b>Promedio</b>	39,3
<b>Máximo</b>	44,5
<b>Mínimo</b>	34,7

El modelo obtenido es de la siguiente forma:

$$\%Sólidos\ O/F = +31.21 + 5.7E - 002 * 'WIC1043 - 1' + - 4.289E - 003 * 'FIC1103 - 2' + - 1.463 * 'PI1173 - 01' - 5.12E - 002 * 'LIC1163 - 01'$$

Ambas ecuaciones fueron ingresadas al PI mediante la opción Dataset, creando de esa manera los modelos predictivos de análisis multivariante que serían vistos en el PI Process Book.

### Empleando el EXCEL:

Para encontrar los modelos de %Sólidos O/F y %Malla +65, se utilizó la Herramienta de Análisis de Datos del Excel. Los modelos encontrados fueron los siguientes:

### MODELO %MALLA +65:

**TABLA N° 24: MODELO %MALLA +65 - EXCEL**

VARIABLES DEL MODELO			Coeficientes	t-value
	Intercepto		-81,625	-3,20
<b>WIC1043-1</b>	Tonelaje	TMH	0,103	4,14
<b>PI1173-01</b>	Presión de Bomba de ciclones	psi	-3,849	-7,47
<b>LIC1163-01</b>	Nivel de Cajón	%	0,053	1,44
<b>XY1103-1</b>	Carga Circulante	%	0,027	2,02
<b>J11103-1</b>	Potencia Motor	MW	28,787	2,99
<b>FIC1103-3</b>	Agua Colectores de Alimentación	LPM	0,013	2,46
<b>FIC1103-2</b>	Agua Fresca de Alimentación	LPM	-0,007	-1,94

**TABLA N° 25: RESULTADO ESTADÍSTICO  
DEL MODELO % MALLA +65- EXCEL**

<b>Puntos de Regresión</b>	57
<b>Coeficiente de Correlación</b>	0,79
<b>F-value</b>	11,53

El modelo obtenido fue:

$$\%Malla + 65 = -81.625 + 0.103*WIC1043-1 - 3.849*PI1173-01 + 0.053*LIC1163-01 + 0.027*XY1103-1 + 28.787*JI1103-1 + 0.013*FIC1103-3 - 0.007*FIC1103-2$$

**MODELO %SÓLIDOS O/F:**

**TABLA N° 26: MODELO %SÓLIDOS O/F- EXCEL**

VARIABLES DEL MODELO			Coefficientes	t-value
	Intercepto		36,254	12,56
<b>WIC1043-1</b>	Tonelaje	TMH	0,070	9,59
<b>PI1173-01</b>	Presión de Bomba de ciclones	psi	-0,708	-1,98
<b>FIC1103-2</b>	Agua Recuperada de Alimentación	LPM	-0,006	-3,82
<b>FIC1103-1</b>	Agua Fresca de Alimentación	LPM	-0,002	-2,18
<b>FIC1163-05</b>	Agua Recuperada de Dilución	LPM	-0,002	-3,28

**TABLA N° 27: RESULTADO ESTADÍSTICO DEL TABLA N° 26: MODELO  
%SÓLIDOS O/F- EXCEL**

<b>Puntos de Regresión</b>	57
<b>Coefficiente de Correlación</b>	0,83
<b>F-value</b>	22,57

El modelo obtenido es de la siguiente forma:

$$\%Sólidos O/F = +36.254 + 0.07*WIC1043-1 - 0.708*PI1173-01 - 0.006*FIC1103-2 - 0.002*FIC1103-1 - 0.02*FIC1163-05$$

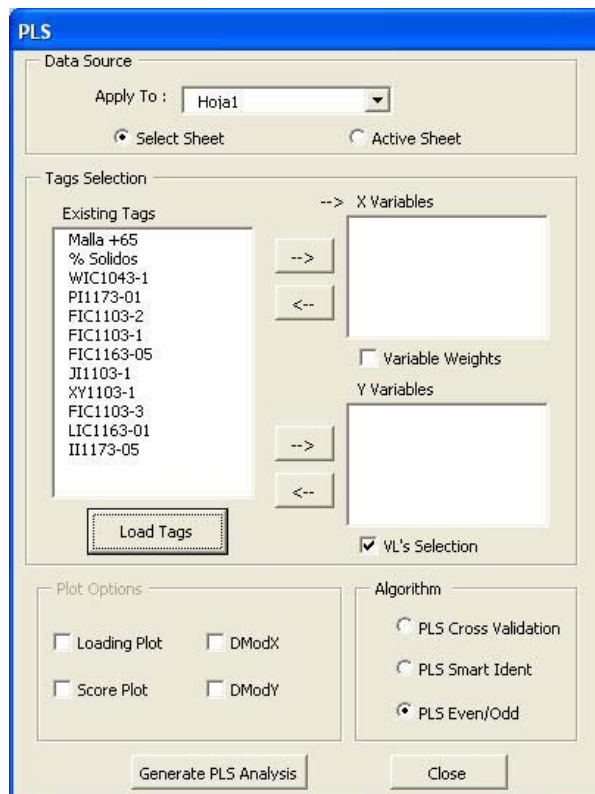
**Empleando el SCAN:**

La Aplicación SCAN es una de las herramientas de análisis multivariante que tiene su programación en el Visual Basic, y muestra sus resultados en la

Hoja de cálculo del Excel, lo que facilita su manejo y su aplicación. Los modelos predictivos multivariantes son generados aplicando la teoría de PLS, los cuales se muestran a continuación:

### MODELO %MALLA +65:

Primeramente se selecciona los Tags de las variables de operación, a continuación se distinguen cual es la variable de salida (Y variable) en este caso el %Malla+65, siendo las otras variables las variables de entrada (X variable). A continuación se presenta el formato de cómo se ingresan las variables de entrada y de salida.



*Figura 60*

El uso de información histórica de las variables permitirá predecir el comportamiento de la variable de salida en este caso el %Malla+65 para la creación de los Soft Sensor.

Para ello se crea automáticamente una planilla en el Excel que indica el peso de los coeficientes y las constantes del modelo PLS.

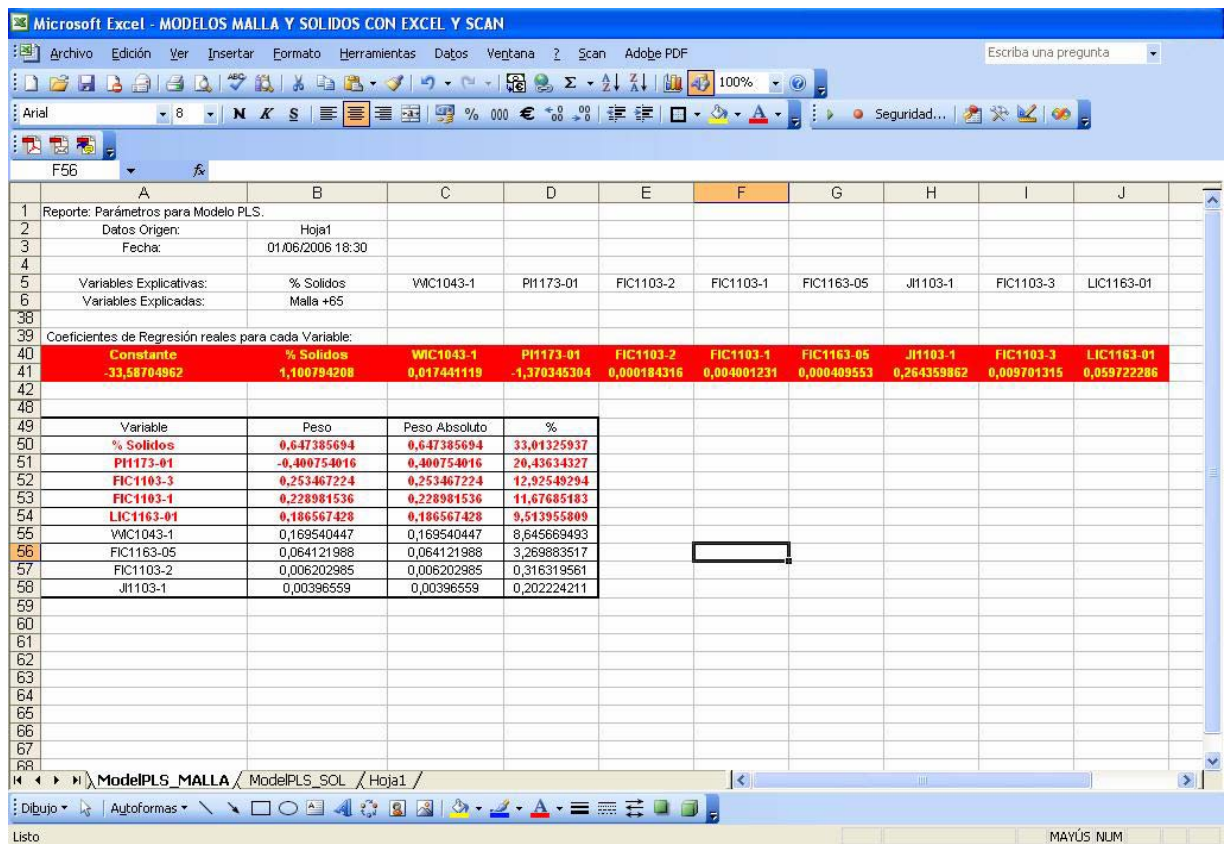
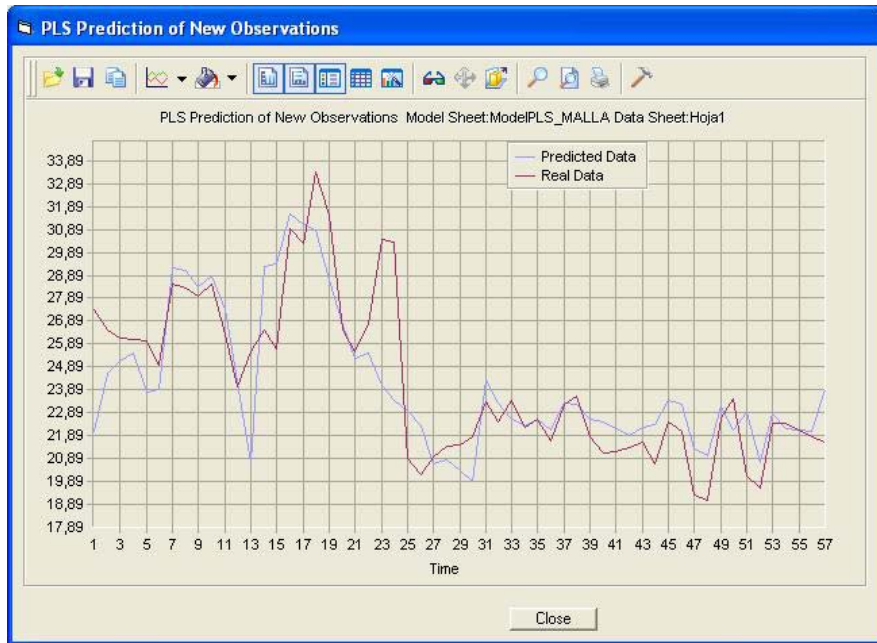


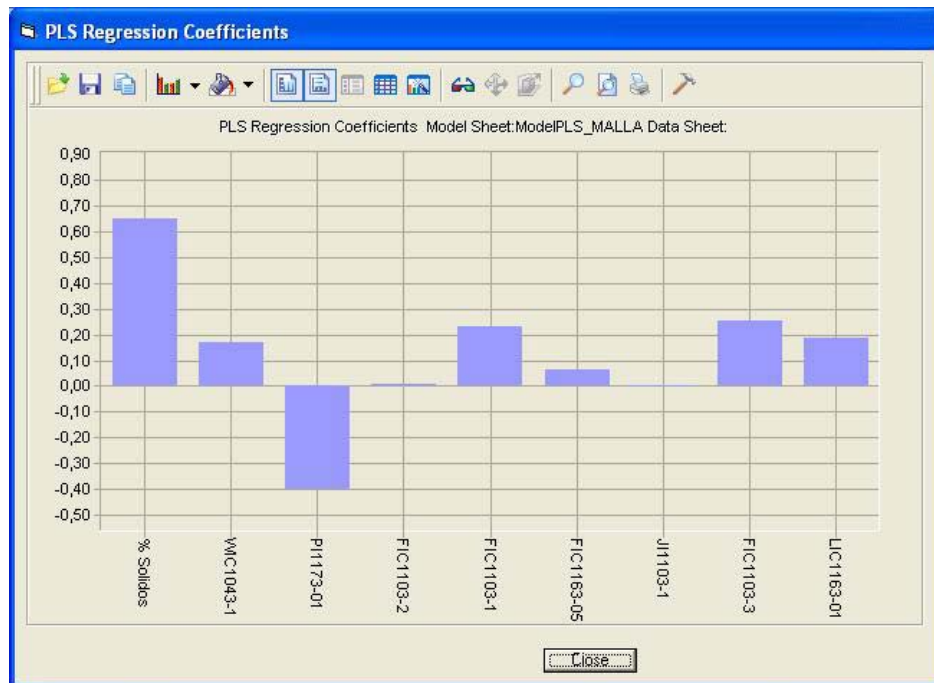
Figura 61

Con la creación de la plantilla mostrada en el Excel, los modelos PLS, permiten graficar la Predicción del modelo PLS frente a los datos observados como se aprecia a continuación:



*Figura 62*

Otro de los gráficos que son mucha importancia es el que nos muestra el peso de los coeficientes de Regresión:



*Figura 63*

Estos pesos indicados permiten examinar cual es la variable de mayor significancia en el modelo predictivo multivariante.

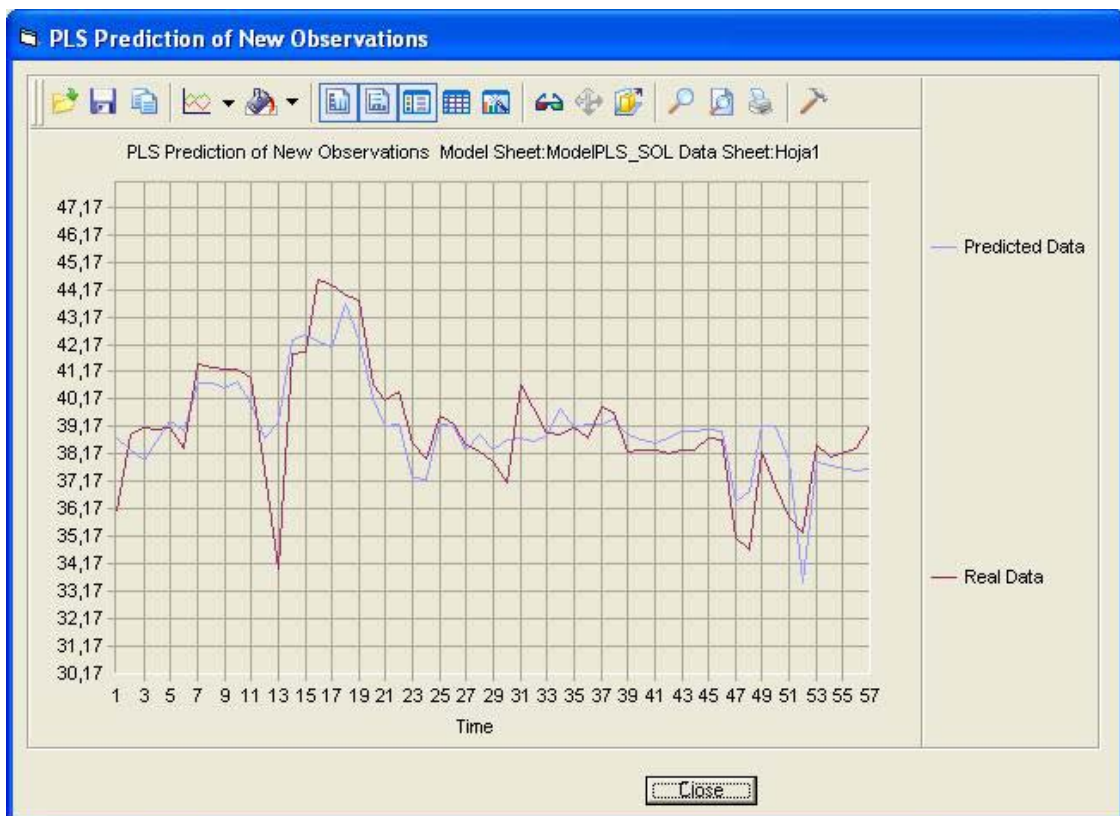
El modelo encontrado con el SCAN para el %Malla+65 fue el siguiente:

$$\begin{aligned} \%Malla + 65 = & -33.587 + 1.1007 * 'DI1103 - 35' + 0.0174 * 'WIC1043 - 1' - 1.3703 * 'PI1173 - 01' \\ & + 0.0597 * 'LIC1163 - 01' + 0.2643 * 'JI1106 - 1' + 0.0097 * 'FIC1103 - 3' + 0.004 * 'FIC1103 - 1' \\ & + 0.0004095 * 'FIC1163 - 05' + 0.000184 * 'FIC1103 - 2' \end{aligned}$$

### MODELO %SÓLIDOS O/F:

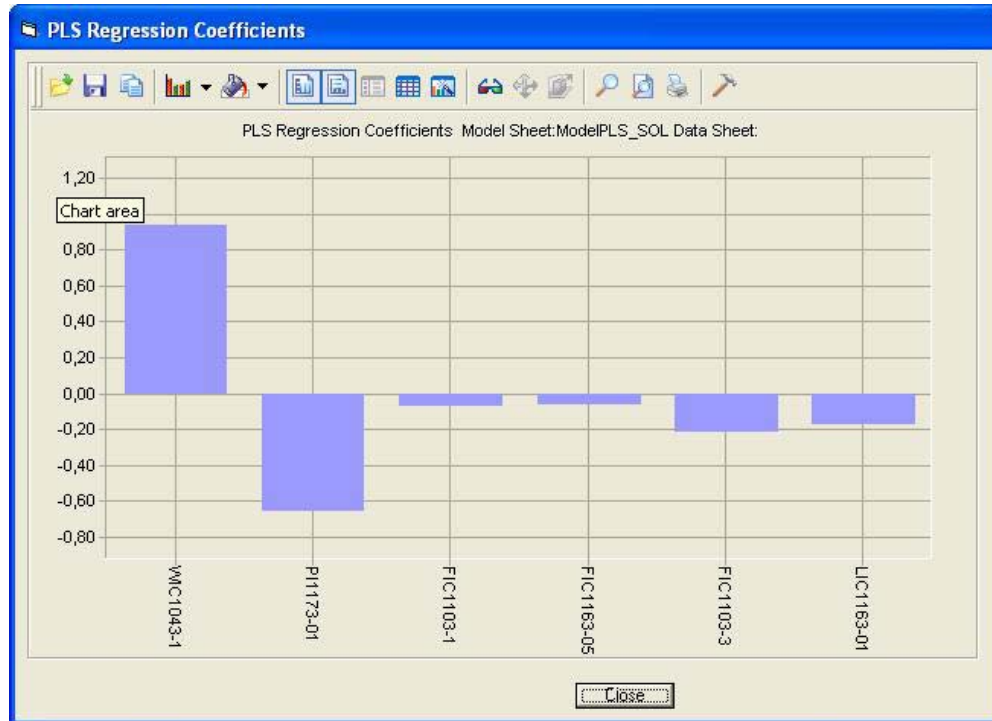
Análogamente se realizó un modelo para el %Sólidos del Overflow, siguiendo los mismos pasos indicados anteriormente.

El modelo PLS generado por el SCAN que nos indica el Modelo Predictivo versus la data real se muestra a continuación:



*Figura 64*

Similarmente el modelo PLS que nos indica el peso de los coeficientes de regresión:



*Figura 65*

El modelo encontrado con el SCAN para el %Sólidos del O/F fue el siguiente:

$$\%Sólidos\ O/F = +35.36 + 0.0566 * WIC1043 - 1 - 1.32 * PI1173 - 01 - 0.0317 * LIC1163 - 01 - 0.00489 * FIC1103 - 3 - 0.000712 * FIC1103 - 1 - 0.000238 * FIC1163 - 05$$

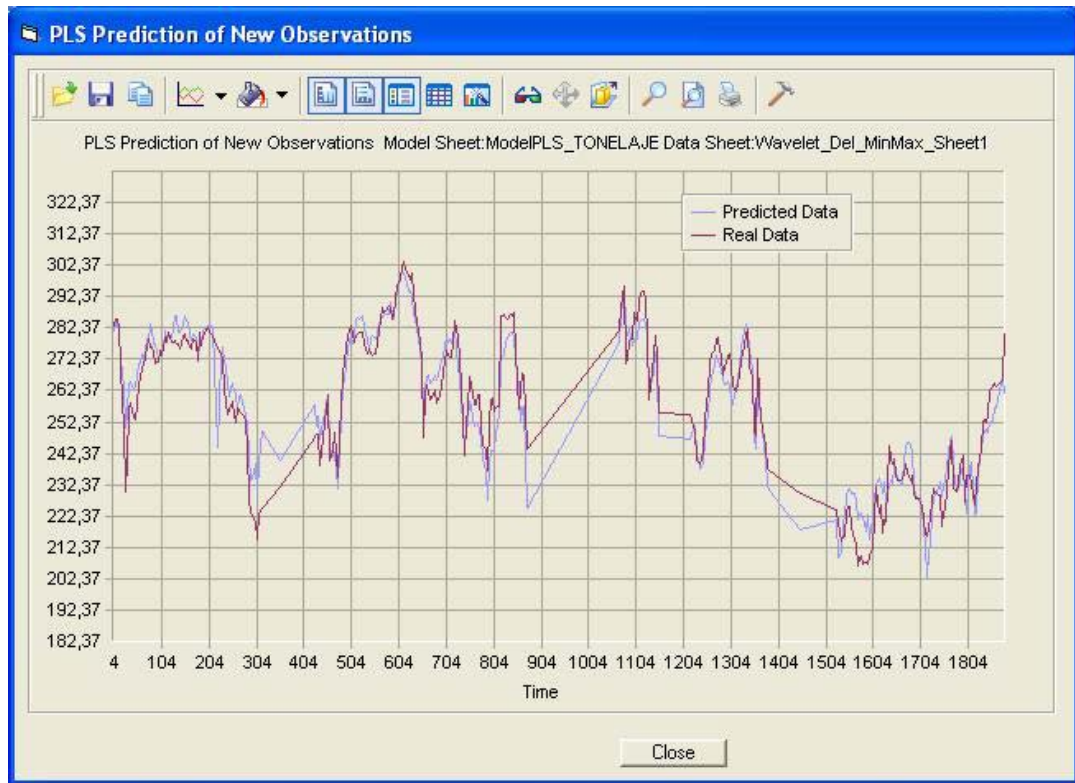
### 3.3.4. MODELO PREDICTIVO TONELAJE:

Los modelos generados para el %Malla+65 y para el %Sólidos del O/F, son modelos que nos permiten predecir el comportamiento de dicha variable cuando el PSM falle o entre en mantenimiento. A dichos modelos se les conoce como Soft Sensor o Sensores Virtuales.

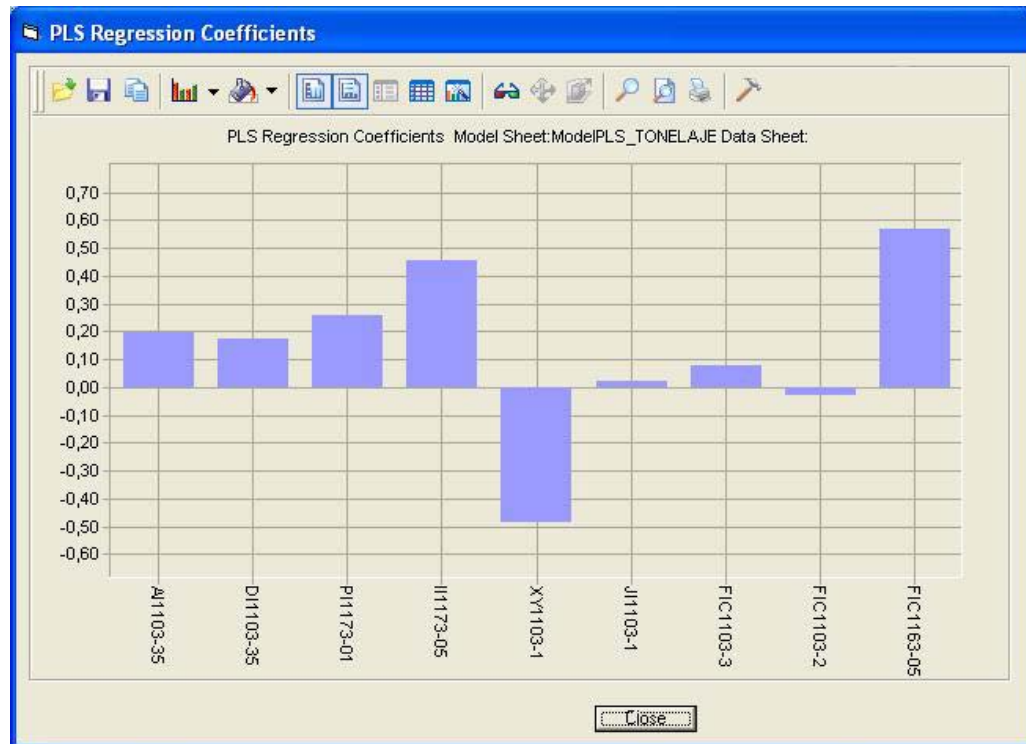


El modelo generado para el Tonelaje es un modelo predictivo que en función al comportamiento de las variables de operación de la molienda permiten predecir cual sería el Tonelaje óptimo a tratar.

El modelo obtenido con el SCAN para el Tonelaje se muestra a continuación.



*Figura 66*



*Figura 67*

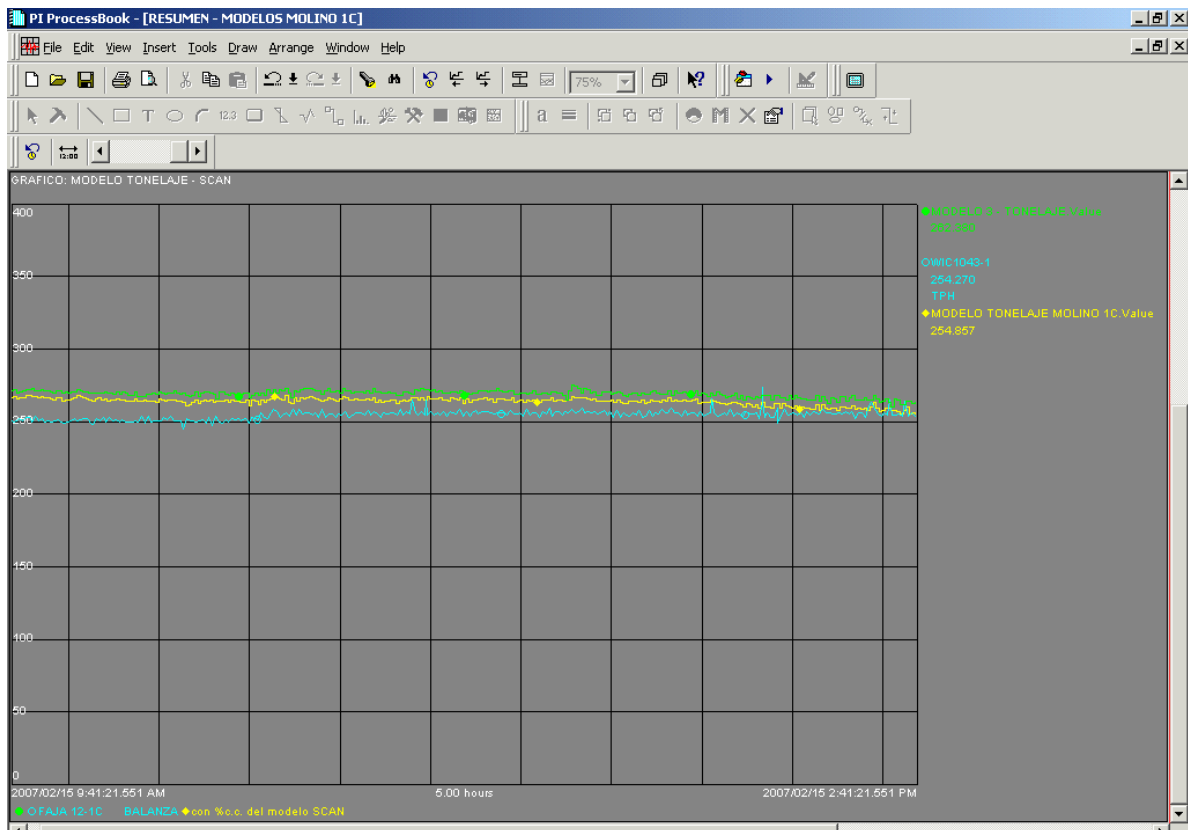
El modelo es el siguiente:

$$\text{Tonelaje TMH} = -145.67638 + 0.9052418 * \text{AI1103-35}' + 2.711986 * \text{DI1103-35}' + 8.871823 * \text{PI1173-01}' + 2.73878 * \text{II1173-05}' - 0.2573 * \text{XY1103-1}' + 10.257077 * \text{JI1103-1}' + 0.0229 * \text{FIC1103-3}' - 0.004816 * \text{FIC1103-2}' + 0.017769 * \text{FIC1163-05}'$$

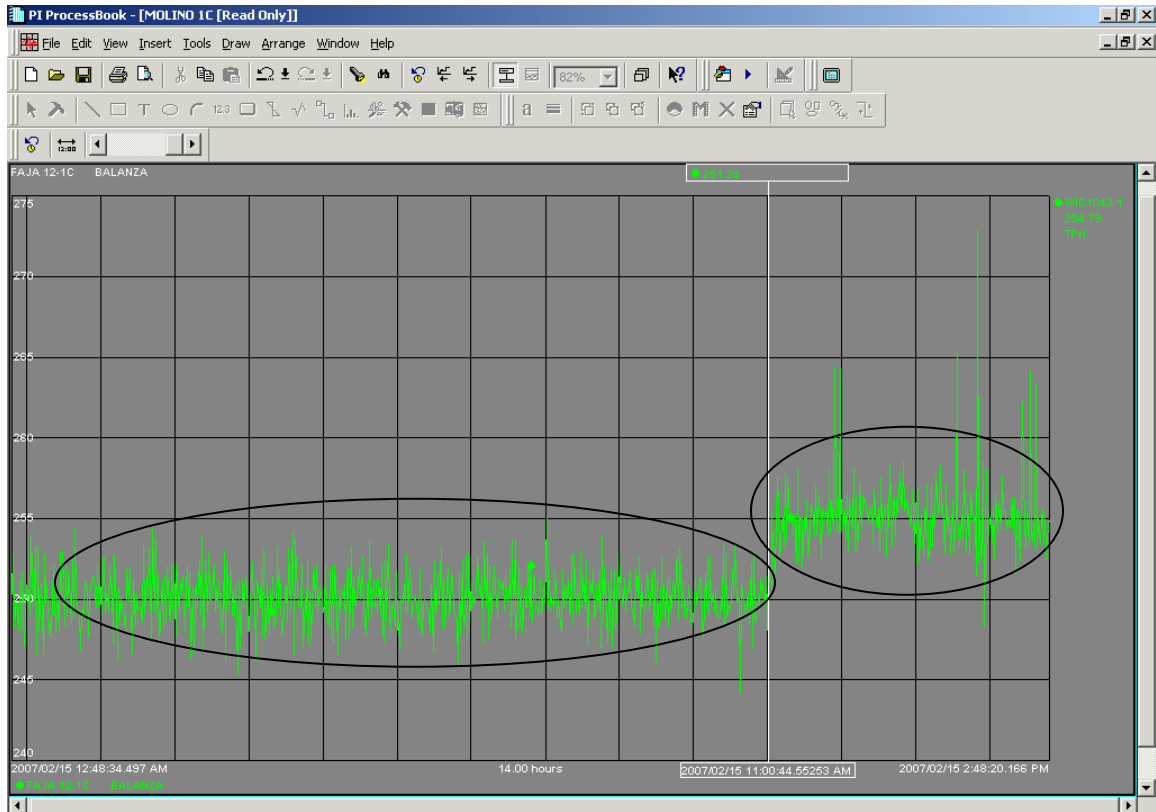
## ANALISIS DE INCREMENTO DE TONELAJE MOLINO 1C CON APLICACIÓN DEL SCAN

Los modelos predictivos generados con el SCAN fueron grabados en el Sistema para poder ser vistos mediante el Process Book. La aplicación de los modelos son una herramienta que permiten incrementar el tonelaje cuando el operador no se da cuenta que es posible poder lograrlo.

El siguiente gráfico muestra un ejemplo de cómo los modelos indicaban un mayor tonelaje al que realmente se estaba pasando en ese momento; a raíz de ello se realizó la prueba cuidando siempre el %+malla 65 y la caída de la potencia del motor.



*Fig 68. Se tiene dos modelos generados por el SCAN (línea amarilla y verde) los cuales indicaban que se podía pasar un mayor tonelaje al real (línea azul), fue así que se procedió a incrementar el tonelaje de 250TMH a 255TMH*



*Fig 69. El tonelaje se incrementó de 250TMH a 255TMH, haciendo uso del modelo generado en el SCAN*

## CAPITULO IV

### DISCUSION - ANALISIS DE LOS RESULTADOS

---

La evaluación del Control Estadístico Multivariante que se realizó al Molino 1C, mediante los modelos predictivos generados, permitió evaluar el incremento del tonelaje por guardia de 12h, durante un periodo de 9 días como se muestra en la Tabla a continuación. Los tonelajes mostrados son la sumatoria del incremento del tonelaje por hora, durante dicha guardia.

**TABLA N° 29: RESULTADOS DEL INCREMENTO DEL TONELAJE**

<b>INCREMENTO DEL TONELAJE</b>		
<b>FECHA</b>	<b>TM -12h</b>	<b>TMD</b>
1-sep-2006	49,4	98,9
2-sep-2006	93,2	186,3
3-sep-2006	113,0	225,9
4-sep-2006	23,1	46,2
5-sep-2006	122,6	245,2
6-sep-2006	101,1	202,2
7-sep-2006	39,3	78,5
8-sep-2006	72,8	145,6
9-sep-2006	21,6	43,1
<b>PROMEDIO</b>	<b>70,7</b>	<b>141,3</b>
<b>Toneladas Mínimo</b>	<b>21,6</b>	<b>43,1</b>
<b>Toneladas Máximo</b>	<b>122,6</b>	<b>245,2</b>

Estos resultados nos indican que el tonelaje promedio **incrementado por día** en los 9 días de la prueba fue de **141.3 TM** (equivalente a un incremento de 22% de Cobre Total por día), lo cual significaría un incremento de 4239.6 TM húmedas en un mes de 30 días solo para el molino 1C.

A continuación se realizó un Balance Metalúrgico para evaluar el incremento de Contenido Metálico y de esta manera el beneficio económico expresado en miles de US\$, tanto por el Molino 1C como por los 8 Molinos de 16' x 20' considerando el mismo incremento de tonelaje del Molino 1C.

**TABLA Nº 30: BENEFICIO ECÓNÓMICO DEL INCREMENTO DEL TONELAJE**

	<b>TM Húmedas</b>	<b>TMS</b>	<b>Ley Cabeza</b>	<b>Contenido</b>	<b>Recuperación</b>	<b>Contenido</b>	<b>Precio</b>	<b>Beneficio</b>
	<b>Mes</b>	<b>Mes</b>	<b>%Cu</b>	<b>Metálico</b>	<b>%Cu</b>	<b>Metálico Adicional™</b>	<b>Cu US\$/lb</b>	<b>miles US\$</b>
<b>MOLINO 1C 16,5' x 20'</b>	4239	4133	0,72	29.76	85	25,29	3	\$ 167.142
<b>8 MOLINOS 16,5' x 20'</b>	33912	33064	0,72	238.06	85	202,35	3	\$1,337.134

## CAPITULO V

### CONCLUSIONES y RECOMENDACIONES

---

#### 5.1. CONCLUSIONES

- Las primeras pruebas que sirvieron de hipótesis a la tesis se realizaron en los molinos de bolas 1C y 1D de 16.5' x 20'. El incremento obtenido en estas pruebas fue de 93.86 TM en un tiempo de 8h, utilizando para ello los modelos multivariantes PLS de la Potencia y el Tonelaje.
- Para lograr el objetivo de la Hipótesis de la Tesis: Incrementar el Tonelaje en la Molienda, fue necesario adaptar, integrar y conciliar diversos componentes de la tecnología disponible en la Planta Concentradora Cuajone. El logro del incremento del tonelaje, requirió en principio el desarrollo de un Sensor Virtual del %malla +65 como un variable de control de calidad y del % de Sólidos en el OverFlow. La información recolectada de las variables de operación del circuito de molienda clasificación, permitió elaborar modelos multivariantes con el uso de la Aplicación SCAN, los cuales pudieron ser ingresados de dos formas al Sistema PI: mediante los Data Set, en forma manual y mediante el PI ACE, en forma automática.
- Los modelos generados fueron el %Carga Circulante, Potencia, %Malla +65, %Sólidos en el OverFlow y el Tonelaje. Dichos modelos permitieron realizar correcciones de las desviaciones de las condiciones de la operación, con los cuales se logró del incremento del tonelaje.

- Para el estudio se hizo análisis de toda una data extraída, los cuales cumplían las condiciones de Normalidad, Homocedasticidad y Linealidad, requisitos para la generación de los análisis multivariantes. Además de ello se hizo uso de los límites de control, eliminando de esta forma datos que ocasionaban un bajo resultado en la capacidad del proceso.
- El conocimiento del procesos y lazos de control existentes en la planta cobró especial relevancia a la hora de estudiar la relación causa - efecto. De esta forma, se pudo establecer un conjunto de variables que permitió “explicar” el comportamiento de otra. Las relaciones matemáticas que vinculan a la variable en estudio con su set de variables explicativas se conoce como “modelo”.
- El incremento de tonelaje en la molienda aplicando el control estadístico multivariante se realizó enfocando a los 8 molinos de 16.5' x 20', el cual aumentaría si se generasen los modelos predictivos para los molinos de mayor tamaño de 20' x 33.5'. Los resultados indican que el tonelaje promedio incrementado por día en el periodo de la prueba fue de 141.3 TM, lo cual significaría un incremento de 4239.6 TM húmedas en un mes de 30 días solo para el molino 1C y de 33912 TM si se aplicase a los 8 molinos de 16.5' x 20', significando un beneficio económico de 44'571 US\$/día.
- El uso del SCAN es solo uno de los tantos software que existe, pero es quizás el más factible de utilizar gracias a que tiene desarrollado su aplicación en una hoja de cálculo como es el Excel.



## 5.2. RECOMENDACIONES

- Es tarea de todo metalurgista contribuir al desarrollo del país siendo para ello indispensable tener contacto con la tecnología que se aplica en las grandes empresas, como es el caso de Cuajone.
- Incursionar en temas que anteriormente fueron solo aplicados a la medicina o a otras ramas que no fuesen ingeniería metalúrgica como es el caso de la teoría de Control de Procesos, Análisis Multivariante, específicamente estos dos temas que fueron la base para el desarrollo de la presente tesis.
- La aplicación y el desarrollo de la tecnología en la Planta Concentradora Cuajone es un modelo para continuar con la investigación y desarrollar teorías como es el caso de la Estadística Multivariante, la cual se pudo aplicar para el desarrollo de sensores virtuales y modelos predictivos, los cuales son una herramienta alternativa de optimización para el incremento del tonelaje en la molienda.

## BIBLIOGRAFIA

---

- ANALISIS MULTIVARIANTE - OSEPH HAIR, ROLPH ANDERSON, RONALD TATHAM  
5TA EDICION - 1999
- ANALISIS DE DATOS MULTIVARIANTES - DANIEL PEÑA  
1RA EDICION EN ESPAÑOL -2000
- CONTROL ESTADISTICO DE LA CALIDAD - DOUGLAS MONTGOMERY  
3RA EDICION – 2005
- MANUAL DE OPERACIONES DE MOLIENDA DE LA CONCENTRADORA CUAJONE  
DE SOUTHERN PERU COPPER CORPORATION  
PERFORMANCE ASSOCIATES INTERNATIONAL  
1999
- ANALISIS MULTIVARIABLES Y SUPERVISION DE PROCESOS CON  
HERRAMIENTAS SCAN - CONTAC INGENIEROS LTDA – OSISOFT  
2006
- CONTROL DE PROCESOS EN LA CONCENTRADORA CUAJONE: PROYECTO DE  
IMPLEMENTACION DE UN SISTEMA DE GESTION OPERACIONAL EN PLANTA  
GARY JARAMA - UNI  
2005
- METODOS MULTIVARIANTES EN CONTROL ESTADISTICO DE LA CALIDAD  
WILLIAM RICHARD ANCHIRAICO AGUDO  
FACULTAD DE CIENCIAS MATEMATICAS – UNMSM  
2003
- REVIEW MANUAL A DATA PROCESSING HANDBOOK  
KENNISTON W. LORD - THIRD EDITION

## GLOSARIO

---

AE	Elementos de análisis
AI	Indicador de análisis
AIC	Controlador indicador de análisis
AIT	Transmisor indicador de análisis
AV	Válvula de análisis
AY	Computadora de análisis
FCV	Válvula de control de flujo
FE	Flujo metro (elemento de medición)
FIC	Controlador indicador de flujo
FFIC	Controlador indicador de relación de caudal
FIT	Transmisor indicador de flujo
FQI	Totalizador indicador de flujo
FQIT	Transmisor totalizador indicador de flujo
FV	Válvula de flujo
FY	Computadora o transductor de flujo (ver también I/P)
HC	Controlador manual
HIC	Controlador indicador manual
HS	Switch manual
HV	Válvula manual
IAH	Alarma de amperaje alta
II	Indicador de amperaje
IIT	Transmisor indicador de amperaje
JAH	Alarma de energía alta
JI	Indicador de energía
JIT	Transmisor indicador de energía
JQI	Indicador totalizador de energía
JT	Transmisor de energía
LAH	Alarma de nivel alta

LAL	Alarma de nivel baja
LCV	Válvula de control de nivel
IE	Sensor de nivel (elemento de medición)
LIC	Controlador indicador de nivel
LIT	Transmisor indicador de nivel
LT	Transmisor de nivel
LY	Transductor de nivel
PCV	Válvula de control de presión
PDAH	Alarma de diferencial de presión alta
PI	Indicador de presión
PIC	Controlador indicador de presión
PT	Transmisor de presión
PY	Transductor de presión
RIC	Controlador indicador de relación (ver también FFIC)
SE	Sensor de velocidad
SI	Indicador de velocidad
SIC	Controlador indicador de velocidad
SSL	Switch de velocidad bajo
ST	Transmisor de velocidad
TCV	Válvula de control de temperatura
TE	Sensor de temperatura (elemento de medición)
TI	Indicador de temperatura
TIC	Controlador indicador de temperatura
WE	Sensor de peso (elemento de medición)
WFY	Calculador de relación de peso
WIC	Controlador indicador de peso
WQI	Indicador totalizado de peso
WT	Transmisor de peso

# ANEXOS

ANEXO A  
FACTORES PARA CONSTRUIR CARTAS DE CONTROL PARA VARIABLES

Observaciones en la muestra, <i>n</i>	Carta de promedios			Carta para desviaciones estándar								Carta para rangos				
	Factores para los límites de control			Factores para la línea central		Factores para los límites de control				Factores para la línea central		Factores para los límites de control				
	<i>A</i>	<i>A</i> <sub>2</sub>	<i>A</i> <sub>3</sub>	<i>c</i> <sub>1</sub>	1/ <i>c</i> <sub>4</sub>	<i>B</i> <sub>3</sub>	<i>B</i> <sub>4</sub>	<i>B</i> <sub>5</sub>	<i>B</i> <sub>6</sub>	<i>d</i> <sub>2</sub>	1/ <i>d</i> <sub>2</sub>	<i>d</i> <sub>3</sub>	<i>D</i> <sub>1</sub>	<i>D</i> <sub>2</sub>	<i>D</i> <sub>3</sub>	<i>D</i> <sub>4</sub>
2	2.121	1.880	2.659	0.7979	1.2533	0	3.267	0	2.606	1.128	0.8865	0.853	0	3.686	0	3.267
3	1.732	1.023	1.954	0.8862	1.1284	0	2.568	0	2.276	1.693	0.5907	0.888	0	4.358	0	2.575
4	1.500	0.729	1.628	0.9213	1.0854	0	2.266	0	2.088	2.059	0.4857	0.880	0	4.698	0	2.282
5	1.342	0.577	1.427	0.9400	1.0638	0	2.089	0	1.964	2.326	0.4299	0.864	0	4.918	0	2.115
6	1.225	0.483	1.287	0.9515	1.0510	0.030	1.970	0.029	1.874	2.534	0.3946	0.848	0	5.078	0	2.004
7	1.134	0.419	1.182	0.9594	1.0423	0.118	1.882	0.113	1.806	2.704	0.3698	0.833	0.204	5.204	0.076	1.924
8	1.061	0.373	1.099	0.9650	1.0363	0.185	1.815	0.179	1.751	2.847	0.3512	0.820	0.388	5.306	0.136	1.864
9	1.000	0.337	1.032	0.9693	1.0317	0.239	1.761	0.232	1.707	2.970	0.3367	0.808	0.547	5.393	0.184	1.816
10	0.949	0.308	0.975	0.9727	1.0281	0.284	1.716	0.276	1.669	3.078	0.3249	0.797	0.687	5.469	0.223	1.777
11	0.905	0.285	0.927	0.9754	1.0252	0.321	1.679	0.313	1.637	3.173	0.3152	0.787	0.811	5.535	0.256	1.744
12	0.866	0.266	0.886	0.9776	1.0229	0.354	1.646	0.346	1.610	3.258	0.3069	0.778	0.922	5.594	0.283	1.717
13	0.832	0.249	0.850	0.9794	1.0210	0.382	1.618	0.374	1.585	3.336	0.2998	0.770	1.025	5.647	0.307	1.693
14	0.802	0.235	0.817	0.9810	1.0194	0.406	1.594	0.399	1.563	3.407	0.2935	0.763	1.118	5.696	0.328	1.672
15	0.775	0.223	0.789	0.9823	1.0180	0.428	1.572	0.421	1.544	3.472	0.2880	0.756	1.203	5.741	0.347	1.653
16	0.750	0.212	0.763	0.9835	1.0168	0.448	1.552	0.440	1.526	3.532	0.2831	0.750	1.282	5.782	0.363	1.637
17	0.728	0.203	0.739	0.9845	1.0157	0.466	1.534	0.458	1.511	3.588	0.2787	0.744	1.356	5.820	0.378	1.622
18	0.707	0.194	0.718	0.9854	1.0148	0.482	1.518	0.475	1.496	3.640	0.2747	0.739	1.424	5.856	0.391	1.608
19	0.688	0.187	0.698	0.9862	1.0140	0.497	1.503	0.490	1.483	3.689	0.2711	0.734	1.487	5.891	0.403	1.597
20	0.671	0.180	0.680	0.9869	1.0133	0.510	1.490	0.504	1.470	3.735	0.2677	0.729	1.549	5.921	0.415	1.585
21	0.655	0.173	0.663	0.9876	1.0126	0.523	1.477	0.516	1.459	3.778	0.2647	0.724	1.605	5.951	0.425	1.575
22	0.640	0.167	0.647	0.9882	1.0119	0.534	1.466	0.528	1.448	3.819	0.2618	0.720	1.659	5.979	0.434	1.566
23	0.626	0.162	0.633	0.9887	1.0114	0.545	1.455	0.539	1.438	3.858	0.2592	0.716	1.710	6.006	0.443	1.557
24	0.612	0.157	0.619	0.9892	1.0109	0.555	1.445	0.549	1.429	3.895	0.2567	0.712	1.759	6.031	0.451	1.548
25	0.600	0.153	0.606	0.9896	1.0105	0.565	1.435	0.559	1.420	3.931	0.2544	0.708	1.806	6.056	0.459	1.541

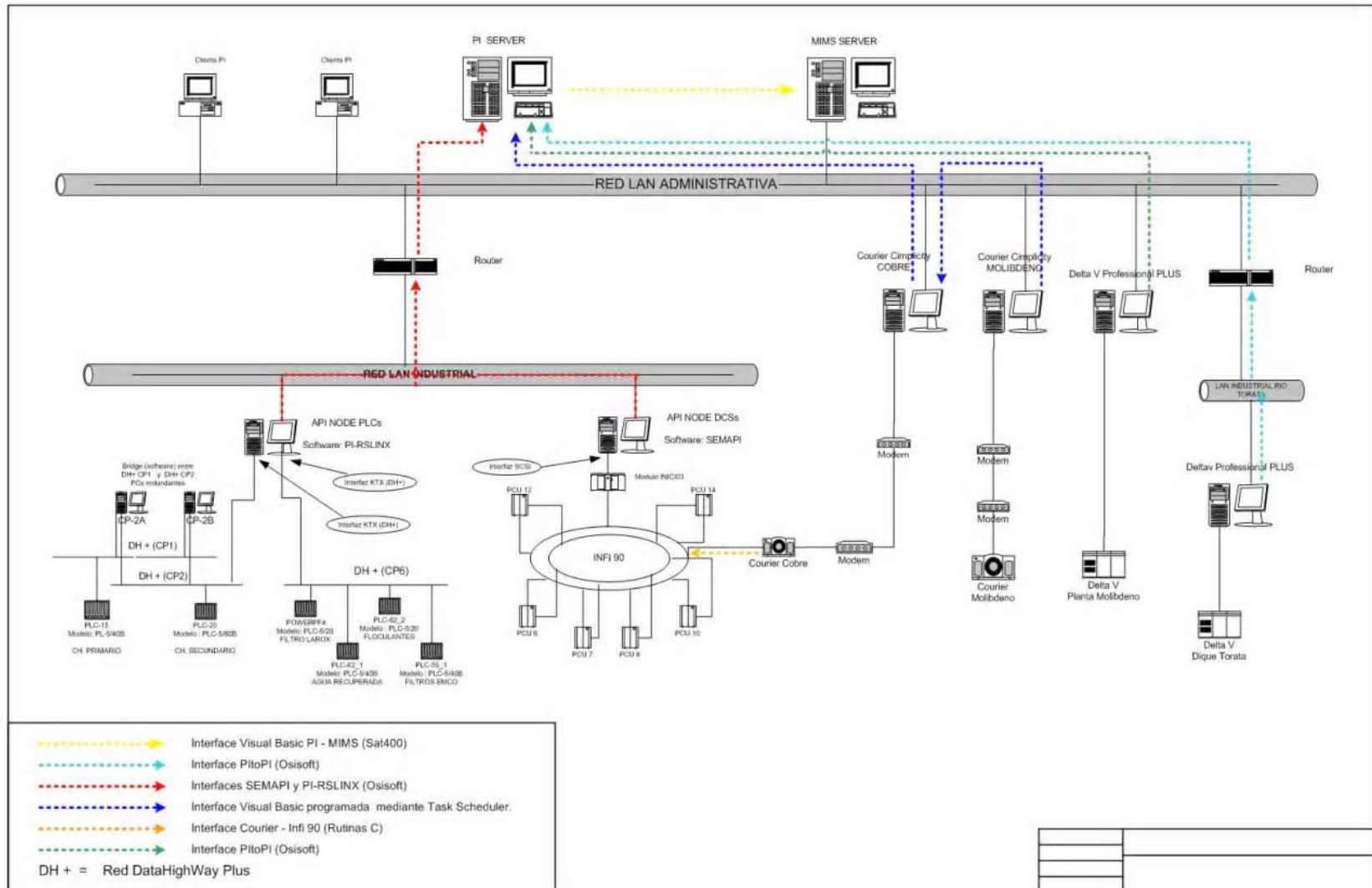
Para *n* > 25.

$$A = \frac{3}{\sqrt{n}} \quad A_3 = \frac{3}{c_4 \sqrt{n}} \quad c_4 = \frac{4(n-1)}{4n-3}$$

$$B_3 = 1 - \frac{3}{c_4 \sqrt{2(n-1)}} \quad B_4 = 1 + \frac{3}{c_4 \sqrt{2(n-1)}}$$

$$B_5 = c_4 - \frac{3}{\sqrt{2(n-1)}} \quad B_6 = c_4 + \frac{3}{\sqrt{2(n-1)}}$$

ANEXO B  
ARQUITECTURA SISTEMA DE INFORMACION DE PLANTA - PI



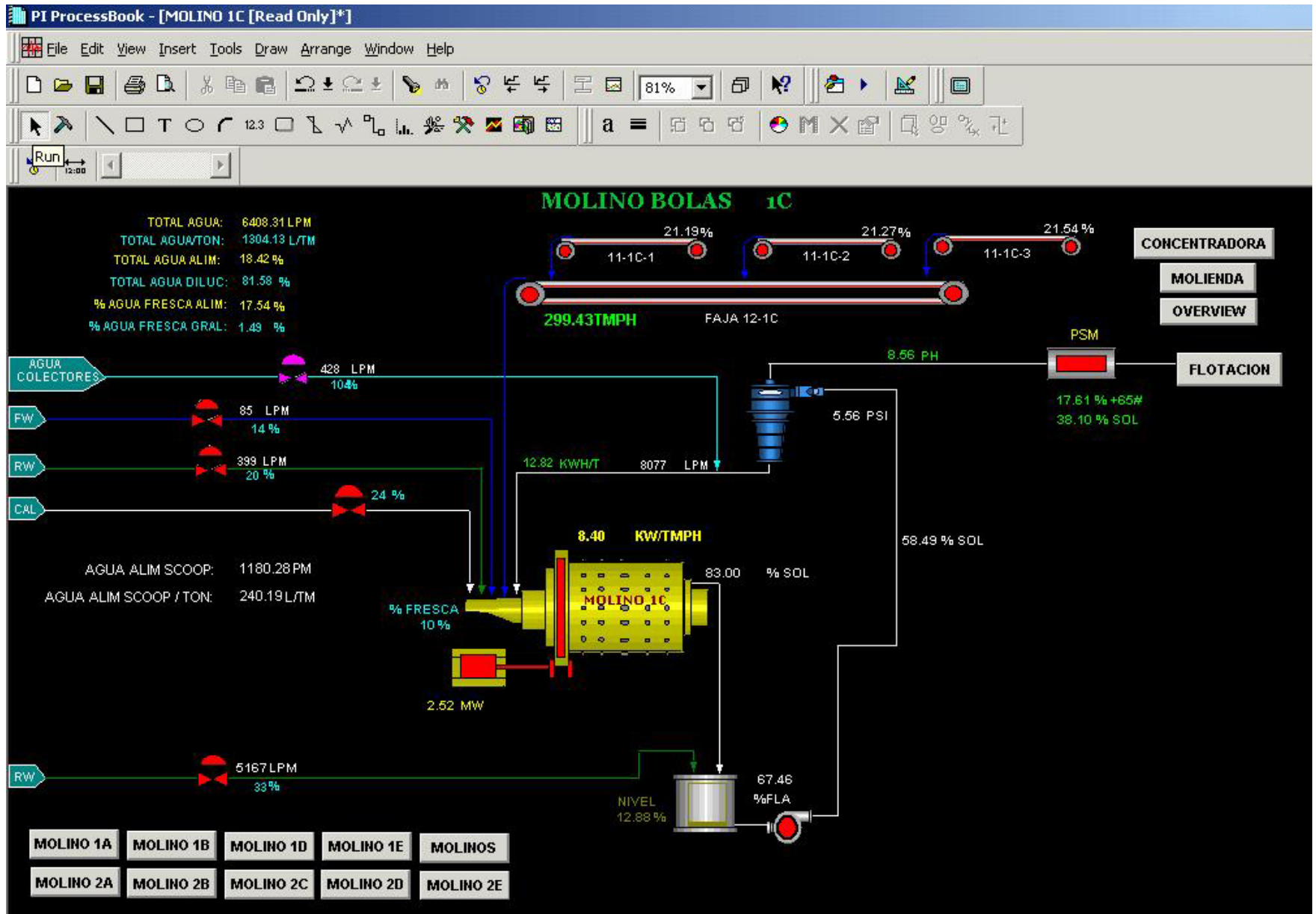
ANEXO C  
VARIABLES DE CONTROL EN LA ETAPA DE MOLIENDA

Sistema del Proceso	Variable del Proceso	Rango de Operación	Método de Control	Impacto en el Proceso
Recuperación de mineral fino	Tasa de alimentación a los Molinos de bolas 1C a 2E	300 a 350 toneladas húmedas por hora o un nivel de 70 a 80 por ciento en el cajón de alimentación a ciclones del molino de bolas.	<p>El operador puede seleccionar controlar la tasa de alimentación según la tasa de peso o según el nivel en el cajón de alimentación a ciclones del molino de bolas.</p> <p>Cuando se requiere una tasa de peso constante al molino, el operador usa un controlador indicador de peso para controlar la tasa de alimentación al molino. Cuando se requiere un tamaño de partículas constante de partículas del producto, el operador emplea un controlador indicador de nivel para controlar la tasa de alimentación al molino.</p> <p>En el modo <i>Automático</i>, el operador ingresa el set point de tonelaje en el controlador indicador de peso y la salida del controlador indicador de peso controla el peso total del mineral entregado por los dos alimentadores exteriores, equipados con transmisiones de velocidad variable.</p> <p>En el modo <i>Cascada</i>, el operador ingresa el set point en el controlador indicador de nivel del cajón de alimentación a ciclones. Este set point se transforma en el set point remoto para el controlador indicador de peso de la tasa de alimentación de mineral.</p>	<p>Una tasa de alimentación muy alta sobrecarga el circuito de molienda, lo que produce un producto grueso y una pérdida de recuperación de mineral valioso. Una tasa de alimentación baja produce una producción de concentrado baja.</p> <p>Un nivel de sumidero alto reduce la caída total que debe producir la bomba. Esto ahorra energía y reduce el desgaste de la bomba. Un nivel de sumidero alto también aumenta la vulnerabilidad del sumidero a las variaciones repentinas que provocan el overflow. Un nivel de sumidero bajo aumenta el consumo de energía. Un nivel de sumidero muy bajo causa cavitación de la bomba y desgaste excesivo de la bomba.</p>
Molienda Control de Densidad de Pulpa	Densidad de la pulpa del molino de bolas (porcentaje de sólidos en la descarga del molino)	68 a 72 por ciento de sólidos según peso (a determinar por la práctica de la planta).	<p>Al chute de alimentación al molino se agrega agua de recuperación y agua fresca en proporción a la tasa de alimentación de mineral nuevo. El operador establece la proporción de agua fresca en relación al agua de recuperación. También controla la densidad de la pulpa en el molino, ajustando el set point de la relación de flujo. El set point representa las proporciones relativas de mineral y agua.</p>	<p>La eficiencia de la molienda disminuye si la densidad es superior o inferior al rango de operación. Una relación de agua alta (baja densidad) ayuda a lavar con abundante agua un molino sobrecargado, pero produce una fuga a través de los pernos del revestimiento del casco. Una relación de agua demasiado baja (alta densidad) puede producir descargas de bolas desde el molino</p>



Molienda Control de pH	pH del overflow del ciclón	8 a 11	<p>El operador ingresa el set point del pH en el controlador indicador de análisis que controla el flujo de la lechada de cal en el chute de alimentación al molino de bolas. Existe un controlador de pH por circuito de molienda de molino de bolas/ciclón.</p> <p>El ajuste principal para la demanda de energía es la adición de bolas de molienda. El operador puede hacer pequeños cambios a la demanda de energía, ajustando la densidad de descarga del molino de bolas, tal como se describió anteriormente.</p>	<p>La lechada de cal agregada al molino de bolas se emplea para ajustar el pH antes de la flotación. La adición excesiva de lechada de cal desperdicia este reactivo y puede impedir una flotación eficiente. La adición insuficiente de lechada de cal en este punto puede tener un efecto adverso en la flotación de las partículas de mineral valiosas y puede permitir que minerales de sulfuro no deseados floten con los minerales de cobre en los circuitos aguas abajo.</p>
Molienda Consumo de Energía	Demanda de energía del molino de bolas.	3.000 hp para los Molinos de bolas 1C a 2E (máx.)	<p>Para las baterías de ciclones de los Molinos de bolas 1A y 1B, el operador agrega o saca de servicio uno o más ciclones para mantener la presión deseada en el distribuidor de ciclones. Ara los Molinos de bolas 1C a 2E, el operador puede ajustar el flujo de volumen a los ciclones al ajustar la adición de agua al cajón de alimentación a ciclones.</p>	<p>Mientras más alta es la energía, más efectiva es la tasa de reducción del tamaño de las partículas. En el molino de bolas, el uso de energía adicional puede arrojar un producto más fino en el mismo tonelaje o moler más mineral en el mismo tamaño de producto. Una carga de bolas demasiado alta puede producir descargas de bolas desde el molino. La energía baja indica una carga baja de bolas. La operación del molino a una carga de bolas menor a la normal puede resultar en una producción total baja en el tamaño de partículas de operación o un tamaño de partículas grueso en la producción de operación. Ambos casos provocan una disminución de la tasa de producción.</p>
Molienda Control de la Presión	Presión de alimentación al ciclón	3 a 7		<p>La presión baja causa un molido más grueso. La presión alta aumenta el desgaste del ciclón y puede llevar a que los ciclones hagan “soga”.</p>

ANEXO D  
MOLINO DE BOLAS 1C - PI PROCESS BOOK



ANEXO E

# Flow Sheet Concentradora Cuajone

