



# Aplicación de Análisis Multivariante en Procesos Industriales

## *Application of Multivarial Analysis in Industrial Processes*

Nelson Aros-Oñate <sup>1</sup>

Carlos Muñoz-Poblete <sup>2</sup>

Sonia Salvo <sup>3</sup>

<sup>1</sup> Facultad de Ingeniería y Ciencias. Universidad de La Frontera.

<sup>2</sup> Facultad de Ingeniería y Ciencias. Universidad de La Frontera.

<sup>3</sup> Facultad de Ingeniería y Ciencias. Universidad de La Frontera.

### Resumen

En este trabajo se presenta la aplicación de dos métodos de análisis multivariante en la modelación del proceso de digestión continua en la industria de celulosa, Planta Santa Fe, Nacimiento - Chile. Específicamente, este estudio ha permitido comparar y evaluar los modelos de procesos generados por los métodos de regresión lineal y redes neuronales, proporcionando información cuantitativa de las relaciones entre las distintas variables de proceso y el número Kappa de salida del digestor. Particularmente, resulta interesante notar que los procesos del digestor no pueden ser modelados con el método de regresión lineal, por lo que se destaca a las redes neuronales como uno de los mejores métodos de predicción, si se trata de modelar procesos con alta correlación entre las variables y con características no lineales. Además, se presenta la aplicación de técnicas estadísticas para análisis de datos extraídos de un proceso industrial, específicamente del proceso de digestión continua en la industria de celulosa. A través de Software dedicado se obtiene información cuantitativa de las relaciones entre las distintas variables de proceso y el número Kappa de salida del digestor. Como resultado de este análisis se obtiene un conjunto de datos depurados que serán utilizados para modelar al proceso en forma empírica.

**Palabras clave:** análisis multivariable, modelación empírica, regresión lineal, redes neuronales, digestión continua.



### Abstract

*This paper presents the application of two methods of multivariate analysis in the modeling of the continuous digestion process in the cellulose industry, Santa Fe Plant, Nacimiento - Chile. Specifically, this study has allowed comparing and evaluating the process models generated by linear regression methods and neural networks, providing quantitative information of the relationships between the different process variables and the Kappa number of the digester output. Particularly, it is interesting to note that the digester processes cannot be modeled with the linear regression method, which is why neural networks stand out as one of the best prediction methods, if it is a matter of modeling processes with high correlation between variables and with non-linear characteristics. In addition, the application of statistical techniques for analyzing data extracted from an industrial process, specifically the continuous digestion process in the pulp industry is presented. Through dedicated software, quantitative information is obtained on the relationships between the different process variables and the Kappa number of the digester output. As a result of this analysis, we obtain a set of refined data that will be used to model the process empirically.*

**Keywords:** *multivariate analysis, empirical modeling, linear regression, neural networks, continuous digestion.*

## Introducción

El uso de computadoras de procesos para el seguimiento de las operaciones ha permitido obtener más y mejores mediciones en tiempo real de muchas variables. Dado que las capacidades de almacenamiento y las velocidades de adquisición continúan aumentando, hoy es posible disponer de grandes volúmenes de datos, los cuales se guardan en dispositivos electrónicos con alta velocidad de acceso. El uso apropiado de estos datos sobre la base de una elevada capacidad de cálculo se ha convertido en un componente crítico para el éxito de la operación a largo plazo.

Sin embargo, el problema de cómo utilizar eficientemente esta gran masa de números continúa siendo un desafío. Específicamente, cuando se trata de industrias de procesos químicos, se llega a afirmar que la mayoría de las plantas de producción son ricas en datos pero pobres en información (Marchetti, 2003). Los datos históricos recolectados rutinariamente en la mayoría de los procesos esconden características estructurales, las cuales puede resultar una valiosa fuente de información para mejorar su operabilidad y la calidad del producto. El problema de extraer información útil de estos datos se visualiza con mayor fuerza cuando, como ocurre en la mayoría de las operaciones industriales, las mediciones son obtenidas en un ambiente con ruido y de variables correlacionadas entre sí.

La primera etapa en el estudio de muchos procesos en la industria es el desarrollo de modelos fundamentales. Estos modelos, generalmente, son usados para el diseño de operaciones unitarias, optimización de las condiciones de operación y de pruebas de estrategias de control. Sin embargo, debido a la complejidad de estos modelos hace difícil que puedan incorporarse directamente en los esquemas de control.

Por otro lado, existen variables que no siempre pueden ser medidas a través de sensores de tiempo real, por lo que se requiere un análisis de laboratorio que involucra un mayor tiempo para entregar el valor de una determinada variable, lo cual hace más complejo el control del proceso. Esto hace deseable desarrollar herramientas que permitan determinar la situación futura de un proceso, en particular, predecir el #Kappa. Los métodos de Análisis Estadístico Multivariante (RLM) pueden ser considerados como poderosas herramientas de predicción, ya que éstas se refieren a un conjunto de métodos, los cuales pueden



analizar relaciones simultáneas entre variables para conseguir una mejor comprensión de los complejos fenómenos presentes en los procesos industriales (Peña, 2002), (Hair, 1999), (Cuadras, 1996). Estos métodos se contrastan con métodos basados en redes neuronales artificiales con el objeto de analizar si la incorporación de modelos multivariantes no lineales mejora la calidad de la predicción (Coughlin, 1995), (Fern, 1999), (Musavi, 1999).

## Materiales y métodos o Metodología computacional

El proceso de cocción desarrollado en el digestor convierte astillas de madera en fibra de celulosa o pulpa, producto de la reacción entre la lignina presente en la madera y los reactivos químicos incorporados en el denominado licor blanco, que consiste en una solución de agua con hidróxido de sodio y sulfuro de sodio. El digestor continuo tiene medidas que bordean los 70 [m] de altura y 7 [m] de diámetro, y su misión es producir pulpa química con flujo constante. Este proceso se lleva a cabo en tres etapas, ver figura 1.

En la *zona de impregnación* - al comienzo del ciclo de digestión - el digestor es alimentado con astillas y licor blanco. Aquí el licor blanco penetra y se esparce dentro de las astillas de madera cuando fluye por esta zona. En la *zona de cocción* se agrega vapor, en volumen y temperatura de acuerdo a los niveles de operación establecidos, de modo de alcanzar los rangos de presión requeridos por el proceso, para que se produzca la cocción adecuadamente, es en esta zona donde ocurren la mayoría de las reacciones químicas que disuelve la lignina y que se remueve a través de las pantallas de extracción del licor; el licor libre está en flujo co-corriente o contra-corriente con respecto al flujo de astillas de madera. Finalmente, la *zona de lavado* - al final del digestor-, aquí un flujo contra-corriente de licor libre lava los productos degradados de la pulpa eliminando todo el reactivo residual; además este flujo ayuda a enfriar la pulpa así como también a detener la reacción de modo de reducir el daño a las fibras de celulosa provocados por la reacción química. La pulpa, retirada por la parte inferior de la vasija, se mide el grado de deslignificación mediante el #Kappa.



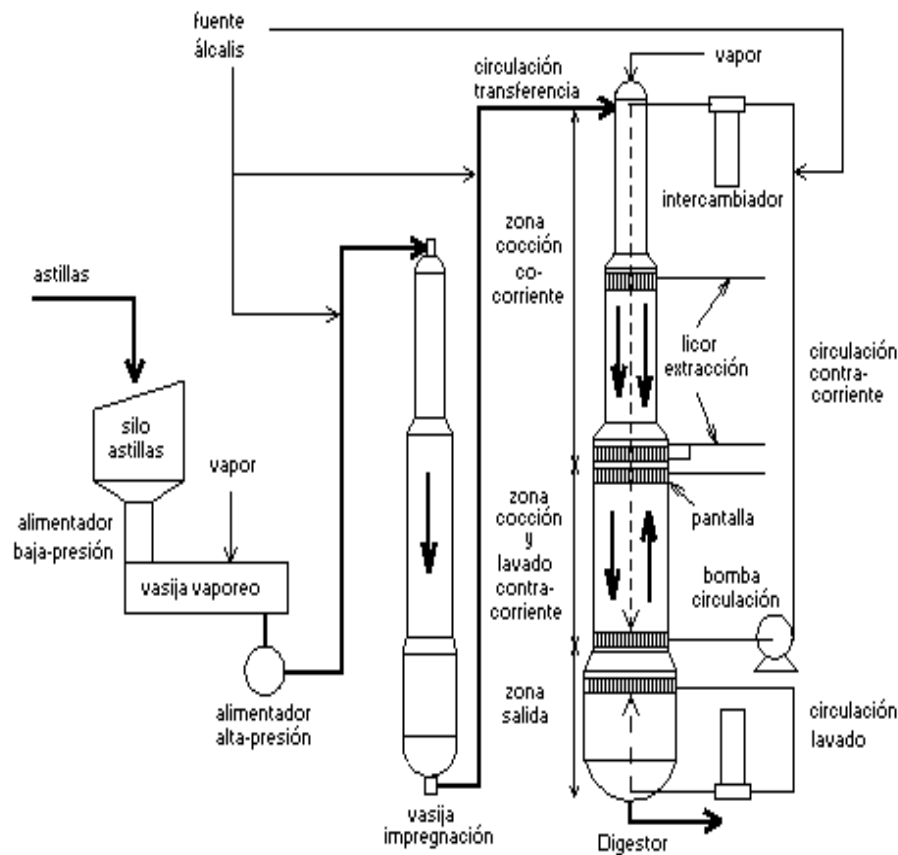


Figura 1. Esquema básico de un digester continuo

### Elección de variables de proceso

Los modelos matemáticos que caracterizan un proceso dinámico son descritos por la interrelación de las variables de entrada y salida de éste; sin embargo, para los modelos de predicción, toman mayor interés las variables de salida del proceso, con las cuales se puede inferir o predecir alguna variable del proceso interna o difícil de obtener en forma directa.

Del proceso de digestión continua de la Planta Santa Fe, es posible extraer una gran cantidad de datos, siendo de interés el conocer cómo se relacionan las variables medidas secundarias más relevantes del proceso en función de la variable objetivo ( $\#Kappa$ ). Las variables más importantes de los subprocesos se presentan en la Tabla 1, basándose en el criterio de los ingenieros de proceso y operadores del digester de dicha planta. Además, las principales variables medidas que caracterizan la calidad de la pulpa resultante son el  $\#Kappa$  y la viscosidad, ambas obtenidas por análisis de laboratorio en la mayoría de las fábricas. Sin embargo, para este trabajo sólo se considera la variable  $\#Kappa$  como variable dependiente en los modelos de predicción.

Luego de escoger las variables más relevantes en el proceso de cocción, se estiman los tiempos de retención de cada variable medida con respecto al inicio del proceso; se sabe que el proceso tiene un tiempo de residencia de cuatro horas y, además, las variables involucradas no influyen en el proceso con el mismo tiempo. Los tiempos de retención fueron estimados con la ayuda de los operadores de la sala de control centralizado a cargo de la operación digestor.

Tabla 1. Variables seleccionadas para representar modelo de proceso

Subproceso	#	Variable	Tag	Unidades	Retención [hrs]
Tolva	1	Humedad	ALI020	( % )	0
	2	Densidad	ALI017	(Kg/m <sup>3</sup> ssc)	0
	3	Temperatura Tolva Astillas	TRC026	( °C )	0
	4	Velocidad dosis. Astillas	SIC029	( rpm )	0
Impregnación	5	Temperatura Circ. Tope	TRA053	( °C )	0
	6	Temperatura Transferencia	TRC067	( °C )	0
	7	Alimentación Licor Blanco	FRC048	( L/s)	0
Cocción	8	Temp. Cocción Lo- solid	TIC208	( °C )	1
	9	Temperatura del Tope	TRC079	(°C )	1
	10	Álcali Pantalla Desc C-7	QI890	( g/L )	1
	11	Extracción Superior	QI540	( g/L )	1
	12	Alcali Residual	QRC068	( g/L )	1
	13	Licor Blanco Circ. Lo-solid	QRC087	( g/L )	1
	14	T. Sal. Lic. Negro Interfondo	TRC121	( °C )	2
	15	Licor Blanco Circ. Lavado	FIC505	( L/s )	2
	16	Kappa	ALPD01	( Kappa )	4

Dentro del período de operación de un proceso industrial se pueden identificar tres distintas etapas: Inicio del proceso, condiciones normales de operación y detención del proceso. Debido a las dinámicas específicas que afectan al proceso durante su inicio o detención (habitualmente más complejas y menos conocidas), sólo se considera el modo de operación normal del proceso.

### Análisis exploratorio de datos

La exploración de datos fue realizada en cuatro meses, específicamente diciembre - marzo. Se analizaron los cuatro meses en forma separada y comparándolos entre ellos, calculándose la media y desviación estándar para cada una de las variables involucradas en el proceso y correspondientes a cada mes, de tal manera de poder visualizar el comportamiento de éstas a través del tiempo.

Para verificar la validez de los datos, se realizaron los diagramas de caja y de dispersión a cada variable en cada mes. De estos diagramas se pudo observar que existen puntos atípicos (outliers), es decir, existe atipia en las observaciones, por lo cual se analizan los valores atípicos por variable y para cada mes en forma separada. A través de la presencia de observaciones atípicas es posible identificar situaciones anómalas del proceso; que lejos de ayudar a la creación del modelo, perjudican la obtención satisfactoria de éste. Como resultado de los diagramas de dispersión, se observa: i) el mes de diciembre presenta una gran cantidad de datos que están muy alejados de su valor de operación, y ii) los meses de enero, febrero y marzo también poseen valores atípicos, pero en muy menor cantidad.

Del análisis exploratorio se puede concluir que es necesario preparar y transformar los datos que van a ser analizados según el método estadístico a aplicar.

### Preprocesamiento de datos

Como el objetivo de este trabajo es procesar datos para luego realizar un identificación de modelos lineales y no lineales, se necesita eliminar todos aquellos datos que no representen las condiciones normales de operación, es decir, como por ejemplo cuando la planta se encuentra detenida o con errores de instrumentación, estos datos fueron identificados manualmente y eliminados de las muestras de trabajo. Para eliminar la mayor cantidad de datos atípicos, no se consideraron las observaciones fuera del rango común de  $\pm 3 \sigma$  respecto de un valor medio mensual.

Inicialmente se disponía de 2439 datos del proceso de digestión continua. De éstos, solo 1912 datos fueron considerados representativos del proceso en condiciones normales de operación, quedando una muestra con la cantidad de datos señaladas en la Tabla 2. El mes que presenta mayor cantidad de observaciones que se encuentran fuera de las condiciones normales de operación fue diciembre, donde el 58% de las observaciones tuvo que ser eliminado.

Tabla 2. Número de Datos antes y después del Filtro

	Antes	Representativos del procesos	Porcentaje Válido de Datos(%)
Diciembre (D)	723	305	42,185



Enero (E)	746	708	94,906
Febrero (F)	676	633	93,639
Marzo (M)	313	266	84,984
<b>Total</b>	<b>2439</b>	<b>1912</b>	<b>78,392</b>

Para trabajar con redes neuronales es muy aconsejable conseguir que los datos que se proporcionan a la red posean las siguientes cualidades: buena distribución (sigan una distribución estándar o uniforme), rango de valores parecidos para todas las entradas, rangos acotados dentro del intervalo de trabajo de la función de activación empleada en las capas ocultas y de salida de las redes neuronales. Para conseguir lo anterior se aplicó dos transformaciones consecutivas:

- Primero se realiza un proceso de *estandarización*, es decir, se transforman las variables de forma que presenten media cero y varianza unitaria. Para ello, se calcula la media ( $\bar{x}$ ) y la varianza ( $\sigma$ ) de cada variable, y se aplica la expresión.

$$x' = \frac{x - \bar{x}}{\sigma}$$

- A continuación, se escalan los nuevos valores al intervalo  $[-1, +1]$  mediante una transformación lineal:

$$y_i = \frac{2x'_i - (X'_{\max} + X'_{\min})}{(X'_{\max} - X'_{\min})} \quad \text{donde:} \quad \begin{aligned} X'_{\max} &= \text{máximo valor de } X' + 30\% \\ X'_{\min} &= \text{mínimo valor de } X' - 30\% \end{aligned}$$

Mediante el escalado se consigue que todas las entradas tengan la misma importancia al igualar su variabilidad y su rango de valores, sin importar el rango y distribución inicial.

Se realiza una descripción del proceso de cocción, donde se discute sobre las variables que intervienen en el proceso de digestión. El conocimiento obtenido del cómo funciona el proceso es fundamental para identificar y eliminar manualmente una gran cantidad de datos atípicos. Sin embargo, la tarea de discriminar los datos atípicos de aquellos que contienen información es compleja y no existen muchas experiencias adecuadamente informadas respecto de cuales métodos utilizar. En este trabajo se utilizaron métodos manuales basados en conocimiento del proceso y métodos automatizados basados principalmente en reglas estadísticas.

El análisis previo de datos reveló que sólo una fracción de los datos que registra un sistema de control distribuido entrega información útil para efectos de modelación. Se considera que el trabajo realizado es fundamental para evitar identificaciones de falsos modelos.



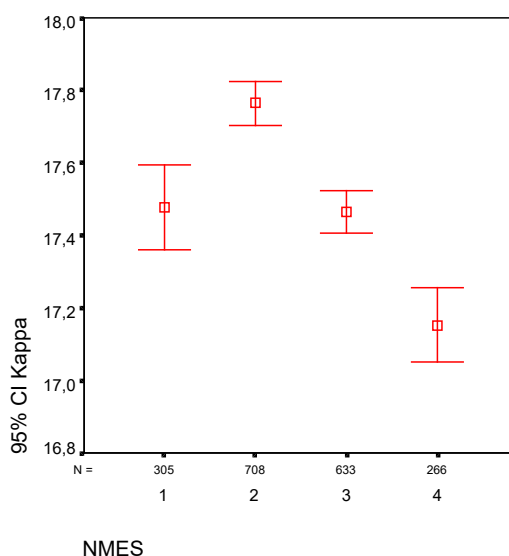
## Resultados y discusión

### Desarrollo de Modelos de Predicción - RLM

El análisis de la influencia de las variables de proceso seleccionadas en el #Kappa se realiza a través de la generación de varios modelos que caracterizan dicho proceso. Los modelos se desarrollan para cada mes, y combinación de meses, con selección de variables que mayor influencia tenían sobre la variable a predecir. Para poder decidir cuáles meses asociar, se desarrollaron diagramas de barra de error para la variable #Kappa, ver figura 2. Este tipo de diagrama es muy útil para ver si existe o no diferencia entre los grupos.

Para cada mes en cuestión se obtuvieron los estadísticos del #Kappa, ver tabla 3. Por diseño, la Planta trabaja un valor del #Kappa de 17. En la Tabla 4 se aprecian los resultados de cada uno de los modelos implementados ordenados de menor a mayor según  $R^2$  ajustado, con sus respectivas variables involucradas y su coeficiente de determinación múltiple ajustado.

Figura 2. Diagrama de Barras de Error para el Numero Kappa



En términos generales se puede apreciar que marzo posee el coeficiente de determinación múltiple ajustado más alto, lo que significa que el 54,4% de variación en el #Kappa está explicada por la relación lineal existente entre el conjunto de variables: Álcali Residual, Temp. Cocción Lo-solids, Temp. sal. Lic. Negro Interfondo, Licor blanco a circulación de Lavado, Temp. Tolva Astillas, Álcali Pantalla Desc C-7 y Licor Blanco circulación Lo-solid. Cabe destacar que el modelo generado con diciembre-enero-febrero-marzo indica que el 15,6% de la variación total en #Kappa ha sido explicada por la ecuación de regresión generada por este modelo.

Debido a lo anterior se trabajó con el modelo generado por marzo y con los cuatro meses juntos. A continuación se muestran los coeficientes estimados de la ecuación de regresión que dan cuenta de la con-



tribución de cada una de las variables de entrada a la variable de salida #Kappa. De la Tabla 6 se observa que los valores de los parámetros son significativamente diferentes de cero, ya que el valor p asociado a cada uno de ellos es menor que 0.01.

Tabla 3. Estadísticos del Numero Kappa

#Kappa				
		Promedio Intervalo Confianza 95% Confiabilidad		
#	Mes	Límite Inferior	Límite Superior	Media
1	Diciembre	17,362	17,593	17,477
2	Enero	17,704	17,825	17,764
3	Febrero	17,406	17,525	17,466
4	Marzo	17,052	17,254	17,153

Tabla 4. Modelos implementados

Modelo	Variables Predictorias Ver Tabla 5	R <sup>2</sup> Ajustado	Cantidad Observación
D	15-2	5,8	261
E	11-10-9-7-3	15,5	660
D-E-F-M	10-4-6-5-14-8-2-7-15	15,6	1674
D-F	2-5-6-12-11-10-1-15-14	17,8	819
E-F	10-4-7-14-1-5-2	18,4	1231
F	10-2-5-14-15-12-9-11-3	25,6	597
M	12-8-14-15-3-10-13	54,4	243

Tabla 5. Variables seleccionadas para representar el modelo

Subproceso	#	Variable
Tolva	1	Humedad
	2	Densidad
	3	Temperatura Tolva Astillas
	4	Velocidad dosis. Astillas
Impregnación	5	Temperatura Circ. tope
	6	Temperatura Transferencia
	7	Alimentación Licor Blanco
Cocción	8	Temp. Cocción Lo-solid
	9	Temperatura del Tope
	10	Álcali Pantalla Desc C-7
	11	Extracción Superior
	12	Alcali Residual
	13	Licor Blanco Circ. Lo-solid
	14	T. Sal. Lic. Negro Interfondo
	15	Licor Blanco Circ. Lavado
	16	Kappa

Tabla 6. Modelos implementados

Modelo	Coeficientes no Estandarizados				Estadístico de Colinealidad
	B	Error Std.	t	Sig.(p)	VIF
<b>D-E-F-M Constante</b>	33,567	4,732	7,113	0	
Var 10	0,1	0,009	-10,87	0	1,495
Var 4	0,287	0,058	4,987	0	5,056
Var 6	0,13	0,014	9,152	0	1,957
Var 5	-0,174	0,034	-5,08	0	1,587
Var 14	-0,197	0,028	-6,93	0	3,892
Var 8	0,106	0,028	3,809	0	3,639
Var 2	-0,084	0,002	-3,621	0	1,150
Var 7	-0,131	0,042	-3,158	0,002	4,746
Var 15	0,521	0,176	2,955	0,003	2,556
<b>Marzo Constante</b>	48,173	11,615	4,147	0	
Var 12	0,383	0,09	4,274	0	1,607
Var 8	0,37	0,059	6,301	0	5,900
Var 14	-0,585	0,084	-6,973	0	7,740
Var 15	2,472	0,536	4,612	0	5,688
Var 3	-0,127	0,04	-3,179	0,002	1,109
Var 10	-0,332	0,096	-3,463	0,001	3,366
Var 13	0,291	0,105	2,765	0,006	3,154

En la Tabla 7 se observa la desviación estándar y la media del residuo entre el valor observado y el valor estimado de #Kappa. Se puede apreciar que la menor distancia al punto ( $u = 0, \sigma = 0$ ) es la correspondiente al modelo generado con los datos de marzo.



Tabla 7. Estadísticos Descriptivos de los Residuales

Residuales no Estandarizados					
	Promedio Intervalo Confianza 95% Confiabilidad				
Modelo	Límite Inferior	Límite Superior	Media	Desviación Std	Distancia
D-E- F-M	0.526	0.566	0.546	0.418	0.687
Marzo	0.403	0.484	0.444	0.319	0.546

Luego de plantear el modelo de trabajo, se debe confirmar su validez mediante el test F correspondiente (directamente de la tabla de análisis de varianza, ANOVA), estableciendo si el modelo es lineal en los parámetros de la ecuación. Se puede observar de la tabla ANOVA (Tabla 8) por medio del test de hipótesis que por lo menos uno de los parámetros  $\beta_i$  posee una relación lineal diferente de cero para ambos modelos, ya que existe evidencia muy fuerte para rechazar la hipótesis nula (de que no existe relación lineal), por lo tanto el modelo es adecuado. Sin embargo las variables predictoras explican muy poco de la variabilidad de  $y$  ( $R^2$  ajustado bajo).

Tabla 8. Estadísticos Descriptivos de los Residuales

	Función de Var.	Suma Cuadr.	Grados de Libertad	Cuadrado Medios	F	Valor p
D-E- F-M	Modelo	151,548	9	16,839	35,45	0,0
	Residual	790,463	1664	0,475		
	Total	942,011	1673			
M	Modelo	91,147	7	13,021	42,25	0,0
	Residual	72,423	235	0,308		
	Total	163,57	242			

A través del análisis de varianza ANOVA se pudo observar que existe una relación lineal entre, a lo menos, una de las variables predictoras y la variable dependiente. Sin embargo, a través del coeficiente de correlación ajustado ( $R^2$ ), se puede observar que las variables predictoras explican muy poco de la variabilidad de  $y$  para ambos modelos, por lo tanto, se puede concluir que si bien existe una relación lineal entre las variables predictoras y el  $\#Kappa$ , ésta es muy baja. Esto quiere decir que existe la presencia de relaciones no lineales entre las variables de proceso.

Por lo anterior, queda demostrado que para el proceso de digestión continua para la planta analizada se genera un modelo predictivo -a través del método de regresión lineal múltiple- que presenta un nivel

muy bajo de confianza. Es posible atribuir la baja calidad de la estimación a la no linealidad del proceso, ya que los datos de este proceso además de ser modelados con Regresión Lineal Múltiple fueron modelados a modo de prueba con Análisis de Factores y los supuestos aún no se seguían cumpliendo, es decir, el modelo no era válido, lo que se traduce en una mala calidad de las predicciones.

### Desarrollo de Modelos de Predicción - RNA

Después de haber concluido que el proceso de digestión continua corresponde a un problema del tipo no lineal, se prestó mayor interés en diseñar las redes neuronales de tal forma que permitieran resolver este tipo de problemas, por lo que se seleccionó la función de activación *Tansig*, ya que sus características no lineales se transmiten a la red. Para generar un modelo de red neuronal se decidió utilizar el *perceptrón multicapa*, puesto que posee gran rapidez de procesamiento, genera buenas representaciones internas de las características de los datos de entrada.

### Elección de la Red

**Número de Capas:** Para la mayoría de los problemas prácticos basta con utilizar una sola capa oculta (Funahashi, 1989; Hornik, et al., 1989). Por lo tanto el número de capas necesarias para formar la RNA es de dos, es decir la Capa Intermedia y la Capa de Salida. **Número de Neuronas en la Capa Intermedia:** Debido a que no existe un método suficientemente seguro para determinar este número, se recurrió a tres posibles candidatos: 20, 10 y 5 neuronas. Manteniendo fijas las demás características se determinó que la red que posee 5 neuronas en la capa intermedia permite generar un error promedio de simulación menor que las demás, por lo que se procedió a implementar dos nuevas redes, esta vez con tres y dos neuronas en la capa intermedia. Los errores de simulación generados por dichas redes permitieron concluir que la red que posee tres neuronas en la capa intermedia es la que presentó un mejor comportamiento al momento de realizar predicciones. **Funciones de Activación:** Debido a que la red debe presentar características no lineales, ésta debe tener funciones de activación no lineal. Se evaluó el comportamiento de las combinaciones *Tansig-Purelin* y *Tansig-Tansig* en las distintas capas de la red neuronal. **Conjunto de entrenamiento:** El conjunto de entrenamiento debe contener una cantidad considerable de observaciones para que la red tenga mayor número de situaciones para “aprender” y evitar tener que sobre-entrenar la red. Por esta razón se tomó desde un mes hacia arriba para entrenar la red y quince días o un mes para predecir. El mes de diciembre no fue considerado dentro de este conjunto por presentar un comportamiento atípico respecto de los demás meses. **Número de Entradas:** Anteriormente se había mencionado que el número de variables a considerar para la modelación del proceso era de 15 neuronas. Considerando esto se probó la incorporación de una nueva variable: El valor de Kappa desplazado en una hora. Se pensó que su efecto podía ser importante al momento de mejorar las predicciones. Es por esto que se probaron, en este sentido, dos tipos de redes: un tipo con 15 entradas y otro con 16. **Épocas de Entrenamiento:** Este es otro de los puntos que se tuvo que determinar con prueba y error. Matlab para el diseño de las Redes Neuronales presentaba un número predeterminado de 100 épocas de entrenamiento y se trabajó con este valor (Demuth, 2001).



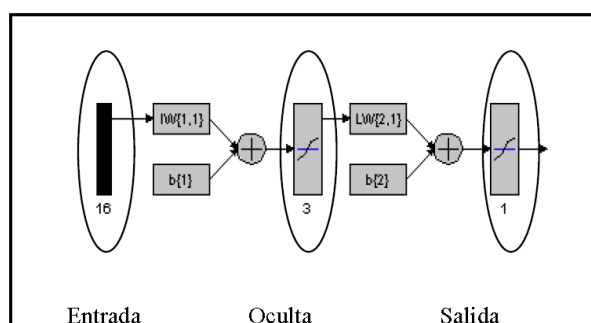


Figura 3. Red mejor evaluada – FFB19

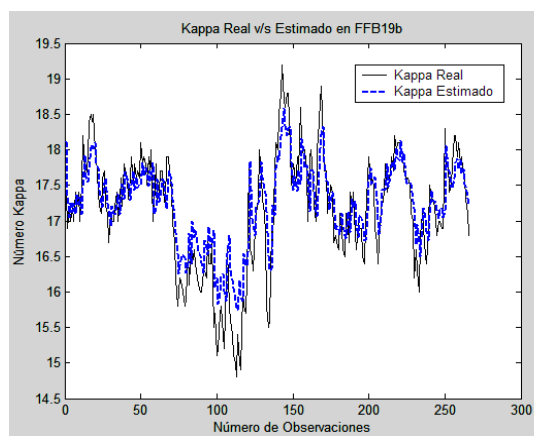


Figura 4. Simulación de la Red FF19 (mes de marzo)

De un conjunto de 28 tipos de redes se seleccionó la mejor red de acuerdo al siguiente criterio: Se determinó el error de predicción (simulación) de las redes implementadas, del cual se extrajeron estadísticos como la media, la desviación estándar y los límites inferiores y superiores del intervalo de confianza. Además se calculó un vector distancia  $d = \sqrt{\mu^2 + \sigma^2}$ , considerado el indicador más importante al momento de evaluar el desempeño de la red. Mientras menor sea la distancia a la coordenada (0,0), mayor será la calidad de predicción del modelo dado por la red.

De acuerdo a este criterio de evaluación se logró determinar que la mejor red posee las siguientes características: dos capas: una oculta y una de salida, tres neuronas en la capa oculta y una neurona en la capa de salida, 16 variables de entrada, de las cuales, la 16ª variable corresponde al #Kappa desplazado en una hora y la función de activación es *Tansig*, en ambas capas, ver figura 3.

Con la estructura antes mencionada es posible obtener excelentes predicciones, como se puede apreciar en la figura 4 con los estadísticos señalados en la Tabla 9. Cabe destacar que los valores señalados en dicho gráfico y tabla corresponden al #Kappa en su escala original, en este caso la red fue entrenada con los meses de Enero y Febrero con la intención de hacer la predicción del #Kappa con un horizonte de 15 días (1º al 15 de Marzo del 2017, con 266 observaciones).

Tabla 9. Estadísticos Descriptivos del error generado por la simulación – mes de marzo

Red	Media	Intervalo de Confianza al 95%			Desv	Dist
		Límite Inf.	Límite Sup.			
FFB19	0,3559	0,322	0,389	0,2786	0,4520	

Una de las mayores complicaciones de la implementación de las Redes Neuronales radica en el hecho de que no existen procedimientos que aseguren la implementación más adecuada del modelo de Redes Neuronales, es decir, no existe un procedimiento que asegure la cantidad de capas ni de nodos en la capa oculta para obtener los mejores resultados.

La gran ventaja de las Redes Neuronales está dada por la casi inexistencia de supuestos que debe cumplir para lograr buenos resultados. De aquí la denominación de modelo libre. Las únicas restricciones están dadas por la cantidad de observaciones necesarias para obtener un conjunto aceptable de información. Pero como se mencionó anteriormente, esto resulta ser un arma de doble filo ya que el modelo no puede ser validado.

## Conclusiones

Este trabajo permitió desarrollar capacidades provenientes de diferentes áreas de conocimiento en los estudiantes de ingeniería que enfrentaron este desarrollo. Capacidades de manejo de métodos de estadística, un entendimiento profundo de la operación de la planta, un manejo de los sistemas informáticos industriales y del sistema de control distribuido, todo desarrollado completamente en la planta industrial hacen de este trabajo una experiencia de formación integral de ingenieros civiles industriales.

Se muestra un procedimiento para la obtención de modelos estadísticos de procesos industriales y se detallan los resultados de múltiples experimentos de análisis para el proceso de digestión continua en la planta Santa Fe de CMPC, utilizando como plataforma de trabajo los software SPSS, que permite identificar modelos multivariados usando Regresión Lineal Múltiple (RLM), y Matlab, que permite entrenar Redes Neuronales Artificiales (RNA) usando una amplia variedad de configuraciones de red y funciones de activación. Por medio de estas herramientas, se analizó el ajuste de modelos de inferencia y, por último, la validación de éstos, con el objetivo de comparar el desempeño de la aplicación de estas al proceso de digestión continua.

## Referencias

- Coughlin D.R., Musavi M.T., Qiau M, (1995). Prediction of Wood Pulp K with Radial Base Function Neuronal Networks. *International Symposium on Circuits and Systems*, 3, 1716-1719.
- Cuadras Carles, (1996). *Métodos de Análisis Multivariante*. 3º Edición, Editorial Universitaria de Barcelona, Barcelona, España.
- Demuth Howard, Beale Mark, (2001). *Neural Network Toolbox for Use with Matlab*. Version 4, The MathWorks.
- Fern A., Miranda J., Musavi M.T, Coughlin D.R., (1999). *Pulp Digester Level Prediction Using Multiresolution Networks of Locally Active Units*. International Conference a Neuronal Networks, 2, 633-638.
- Funahashi, K., (1989). *On the approximate realization of continuous mapping by neural networks*. Neural Networks, Vol 2, pp 183-192.
- Hair J.F., Anderson R.E., Tatham R.L., Black W.C. (1999). *Análisis Multivariante*. 5ª Ed., Prentice Hall, España.
- Hornik, K., Stinchcombe, M. y White, H. (1989). *Multilayer feedforward networks are universal approximators*. Neural Networks, Vol 2, pp 359-366.
- Marchetti Jacinto L., (2003). *Métodos de Análisis Multivariante aplicado al Control de procesos Industriales*. Instituto de Desarrollo Tecnológico para la industria Química, Universidad Nacional del Litoral. Recuperado de <http://www.intec.unl.edu.ar>.



*Ciencias informáticas: investigación, innovación y desarrollo*

Musavi M.T., Domnisoru C., Smith G., (1999). *A Neuro-Fuzzy System for Prediction of Pulp Digester K-Number*. International Conference a Neuronal Networks, **6**, 4253-4258

Peña Daniel, (2002). *Análisis de Datos Multivariantes*. Ed. Mc Graw Hill.



Este contenido se publica bajo licencia CC-BY 4.0

