

# Aproximaciones Simbólicas en un Proceso Industrial de Galvanizado

Beatriz Nava<sup>1</sup>, Mayra Herrera<sup>1</sup>, Pablo Guillén<sup>1</sup>, Francisco Martínez de Pisón<sup>2</sup>

<sup>1</sup>Universidad de Los Andes, Centro de Simulación y Modelos (CESIMO, Mérida, Venezuela,) pguillen@ula.ve

<sup>2</sup>Universidad de La Rioja, Grupo EDMANS, Logroño, España, <http://www.mineriadatos.com>

## RESUMEN

El uso de técnicas de aproximación simbólica que permitan cuantificar el comportamiento de las temperaturas reales de bobinas de acero cuando salen de un horno (dentro de un proceso continuo de galvanizado por inmersión) frente a las temperaturas de consigna (objetivo) resulta de suma importancia en la actualidad. Los índices estadísticos obtenidos a través de estas técnicas permiten identificar diferencias entre las temperaturas de consigna y reales revelando información inherente dentro de la zona de calentamiento del horno para una mejor toma de decisión, supervisión y control del proceso industrial.

Palabras claves: Aproximación Simbólica (SAX), Dinámica Simbólica, Minería de Datos, Series de Tiempo, Proceso Industrial de Galvanizado.

## *Symbolic Approximations in a Galvanizing Industrial Process*

## ABSTRACT

*The use of symbolic approximation techniques to quantify the behavior of measure temperatures regarding temperatures' set points during a continue galvanizing process results of great importance to obtain the desired product. The statistical index obtained with these techniques allows identify relevant information inside the heating zone that could be useful to monitoring and define strategies of supervision and control of the industrial process.*

*Key words: Symbolic Approximation, Symbolic Dynamics, Data Mining, Time Series, Industrial Process of Galvanize.*

## INTRODUCCIÓN

En la Minería de Datos se han propuesto muchas representaciones de series de tiempo de alto nivel. Sin embargo una de las representaciones que la minería de datos no ha examinado en detalle es la discretización de los datos que constituyen una serie temporal en cadenas simbólicas. En ella, se consigue una enorme riqueza de algoritmos y estructuras de datos que permiten la eficiente manipulación de representaciones simbólicas.

Avances recientes dentro de la minería de datos para la obtención de patrones, es el uso de técnicas de la dinámica no lineal como la dinámica simbólica (DS)<sup>1,2,3</sup>, cuyo objetivo consiste en obtener información oculta en series temporales para un mejor diagnóstico y toma de decisiones. Otra de las

técnicas para la obtención de patrones es el método de aproximaciones simbólicas (SAX)<sup>4,5</sup>, el cual permite que las variables medidas se definan en un espacio simbólico. La idea básica de estas técnicas consiste en dividir un conjunto de observaciones en el tiempo en un número finito de segmentos manteniendo la estructura de los datos. Cada segmento es asociado a un símbolo de manera que el proceso sea descrito por una secuencia de símbolos finitos que representan el comportamiento dinámico del sistema. La representación simbólica puede interpretarse como una versión gruesa de la serie temporal con la gran ventaja de reducir la información contenida en una señal a lo más esencial.

En los procesos industriales, específicamente en la industria del galvanizado de bobinas de acero, es común encontrar bases de datos que contienen series temporales. Normalmente, éstas

corresponden a medidas continuas o discretas que reflejan la dinámica real de las variables medidas que son capturadas a través de señales emitidas por sensores, que indican el estado de algún dispositivo a lo largo del tiempo, por ejemplo las temperaturas de la banda de acero a la salida de un horno para el ciclo de recocido de un proceso industrial de galvanizado de bobinas, ver figura 1.

Todo esto despierta gran interés en el estudio y análisis de las series temporales en la industria del galvanizado. El principal objetivo de este estudio es desarrollar e implantar dos algoritmos de aproximaciones simbólicas, para la extracción y caracterización de patrones, que permitan identificar información oculta de los registros históricos de las temperaturas de la banda monitoreadas y almacenadas en sensores en el sistema de control del horno de la línea de galvanizado continuo por inmersión. Los resultados obtenidos mediante estas aproximaciones simbólicas permitirán desarrollar estrategias de monitorización y supervisión para la mejora del proceso industrial.

## MATERIALES Y MÉTODOS

### Descripción de la línea de Galvanizado

El material de partida son bobinas de acero procedentes del tren de laminación. Las bobinas tienen ya el espesor requerido dentro del tren del laminado en frío. Para transformar las bobinas en una banda continua, éstas se desenrollan, despuntando la cabeza y la cola de la bobina, y se sueldan a solape. Las impurezas arrastradas de la laminación se eliminan mediante el calentamiento en atmósfera no oxidante.

Una vez limpia la banda, se somete a un ciclo de calentamiento y enfriamiento, conocido como ciclo de recocido (ver figura 2), que tiene como objetivo mejorar las propiedades del acero de la misma. A continuación, para mejorar la resistencia a la corrosión de la banda de acero, se sumerge en una piscina de zinc fundido quedando revestida de este metal. El control del revestimiento se realiza proyectando aire a presión sobre la banda recubierta. Para prevenir la oxidación de la banda, se somete a un tratamiento superficial de ácido crómico. Después de este proceso, la banda es aplanada obteniéndose el producto acabado bien en forma de bobinas o de chapas cortadas<sup>6</sup>.

La línea de galvanizado (ver figura 3) puede considerarse dividida en tres secciones:

Sección I o de Entrada: que comprende las unidades de desenrollado, enderezado, corte y soldadura.

Sección II o de Proceso: que comprende la unidad de tensión, horno del proceso, unidad de revestimiento, control de revestimiento, equipo de enfriamiento, unidad de tratamiento químico y unidades de aplanado.

Sección III o de Salida: que comprende el equipo de bobinado y de salida de bobinas, tijera, aplanadoras, inspección, apiladora y salida de chapas de rechazo, apilado y salida de chapas de primera calidad.

## Modelo de Control del Horno

En este punto se describe el modelo matemático de la zona de calentamiento del horno de recocido en continuo de la línea de galvanizado para mejorar y optimizar el control del mismo.

El control de la zona de calentamiento del horno de recocido es fundamental, ya que tiene como objetivo recristalizar el metal endurecido que sale del laminado en frío y homogeneizar la estructura cristalina del acero. El estudio se centra en esta zona por ser la de mayor dificultad.

El modelo físico del horno se fundamenta en los mecanismos de transmisión de calor por conducción, convección y radiación. De estos mecanismos, es importante destacar que el flujo de calor intercambiado entre dos cuerpos debido al mecanismo de radiación térmica, depende de las temperaturas de cada uno de ellos y no solo de su diferencia, tal y como sucede en los procesos de convección y conducción. Esto indica que, estos dos últimos procesos son predominantes cuando las temperaturas son bajas, mientras que, para temperaturas elevadas, la radiación térmica es el mecanismo de transmisión de calor más importante. Debido a que la sección de calentamiento del horno se mueve entre los 750°C y 950°C, se considerará solamente el mecanismo de radiación para el cálculo de los modelos<sup>6</sup>. De esta forma, la primera aproximación considera la temperatura de la atmósfera del horno equivalente a la de los tubos radiantes y que la temperatura de las paredes del horno es del orden de la de la atmósfera del horno. Así se llega a la siguiente ecuación:

$$\sigma \cdot S \cdot RC \cdot (T_f^4 - T_s^4) = \rho \cdot d \cdot L_s \cdot W \cdot (QS_d - QS_e) \quad (1)$$

donde:

$\sigma$ : Constante de Stefan Boltzmann

(4,88x10<sup>-8</sup> Kcal/m<sup>2</sup>·h)

$S$ : Area de intercambio de calor (m<sup>2</sup>)

$RC$ : Coeficiente de emisividad

$T_f$ : Temperatura del horno (K)

$T_s$ : Temperatura de la banda (K)

$\rho$ : Densidad del acero (7.859 Kg/m<sup>3</sup>)

$d$ : Espesor de la banda (m)

$L_s$ : Velocidad de la banda (m/min)

$W$ : Ancho de banda (m)

$QS_d$ : Capacidad de calentamiento de la banda a la salida del horno (Kcal/Kg)

$QS_e$ : Capacidad de calentamiento de la banda a la entrada del horno (Kcal/Kg)

## Modelo Matemático del Calentamiento de la Banda

Dentro de las diferentes formas de resolver la ecuación del modelo físico, el proceso industrial que estamos analizando hace uso de la siguiente ecuación lineal interpolada<sup>7,8</sup>:

$$TF = A \cdot TSot + B \cdot (LS \cdot d) + C \cdot \frac{TSot}{RC} + D \cdot \frac{(LS \cdot d)}{RC} + E \quad 2)$$

donde:

*TF*: Temperatura de la banda

*TSot*: Temperatura de consigna para cada bobina

*d*: Espesor de la banda (m)

*LS*: Velocidad de la banda (m/min)

*RC*: Coeficiente de emisividad

*A*, *B*, *C*, *D*, *E*, y *RC*: Son coeficientes tomados de una tabla empírica dependiente de la temperatura de consigna, del producto *LS·d* y de los perfiles de calentamiento (*ST1* y *ST2*) que corresponden con las diferencias de temperaturas entre zonas de calentamiento. Una vez obtenidos los coeficientes, se calculan las temperaturas de consigna para cada una de las zonas mediante ecuaciones interpoladas.

### Calidad del Producto

Las bobinas de acero son convertidas en una banda continua que es limpiada mediante calentamiento no oxidante. Al finalizar la limpieza, a la banda se le dan las características mecánicas deseadas mediante un ciclo de recocido, que consta de un calentamiento y enfriamiento controlado en cuanto a temperaturas y a intervalos de tiempo. El control de estas temperaturas es indispensable para obtener unas bobinas con características adecuadas.

Después del ciclo de recocido, se realiza el galvanizado a través de la inmersión en un pote con zinc, recubriendo la banda por adherencia con una capa de espesor uniforme. El sistema empleado en líneas de galvanizado de este tipo instaladas en el mundo es el control por cuchillas de aire. Sobre la banda inciden dos chorros de aire que eliminan el zinc sobrante devolviéndolo al pote. Con el fin de evitar defectos superficiales, tras el enfriamiento de la banda, ésta es sometida a un tratamiento químico.

Resumiendo, la calidad del producto se puede focalizar en dos aspectos fundamentales:

- En cuanto a las propiedades del acero, depende fundamentalmente de: la composición del acero, el proceso de fundición, los procesos de laminación y el ciclo de recocido.
- En cuanto a las características anticorrosivas, viene marcada por el espesor del recubrimiento de zinc y por la uniformidad del mismo, y depende básicamente de: la preparación superficial del metal base, el control de la temperatura de recubrimiento y homogeneización de la misma, la composición del baño, el control de las cuchillas de aire y velocidad de la banda.

De esta forma, tal como se ha comentado anteriormente, las propiedades finales del acero y el grado de adherencia del espesor de la capa de zinc, van a estar muy influenciadas por el grado de efectividad con que se aplique el ciclo térmico para cada tipo de bobina. Este es, por lo tanto, uno de los objetivos de este trabajo: analizar, controlar y predecir de la mejor

manera posible, las señales de consigna para cada tipo de bobina dentro de la zona de calentamiento del horno.

### Base de Datos

La base de datos está compuesta por la información registrada en un sensor de temperatura ubicado a la salida del horno de la línea de galvanizado continuo por inmersión. Se disponen de registros de temperatura de consigna, *TMPP2C* (temperatura que debe tener el horno) y temperatura medida, *TMPP2M* (temperatura de la banda de acero). Estos registros serán considerados como las series temporales a analizar. Las series temporales están constituidas por 56.376 muestras cada una (ver figura 1). Hay que hacer la siguiente observación, las series temporales originales poseen datos espurios debido a errores de medición.

### Dinámica Simbólica

El principio básico de la dinámica simbólica es transformar una serie temporal en una secuencia de símbolos. Esto provee un modelo del sistema vía un espacio de sucesiones. Dado un conjunto de datos  $x(1), x(2), \dots, x(N)$  la secuencia de símbolos se obtiene cuantificando el conjunto de datos en cajas etiquetadas con un símbolo<sup>1,2</sup>. El cálculo de atributos de la secuencia de símbolos puede revelar características no lineales de la serie temporal original y del sistema dinámico examinado. Las secuencias de símbolos se determinan mediante el siguiente algoritmo: Una ventana de longitud *M* muestras se ajusta sobre el primer valor de la serie. Dentro de la ventana de *M* muestras se determinan las diferencias consecutivas muestra a muestra. Todas aquellas diferencias las cuales son menores que un cierto umbral *r*, el cual se expresa como la desviación estándar de las *M* muestras que constituyen la ventana, son cuantificadas. El resultado da un símbolo y la ventana se desplaza *M+1* valores de la serie, para la obtención del próximo símbolo.

El método de dinámica simbólica presentado en este estudio es:

$$S_j = \sum_{i=1}^{M-1} \left\{ \begin{array}{l} 0 \text{ si } |x_i - x_{i+1}| \geq r \\ 1 \text{ si } |x_i - x_{i+1}| < r \end{array} \right\} \quad j = 1, \dots, N/M \dots\dots\dots 3)$$

El valor de cada símbolo obtenido es un número del conjunto  $\{0, \dots, M-1\}$ . Valores bajos de los símbolos se obtienen cuando las diferencias consecutivas muestra a muestra de la ventana considerada *M* son mayores que el umbral definido, en este caso se cuantifica un comportamiento irregular de la señal y los valores más elevados de los símbolos se obtienen cuando las diferencias consecutivas muestra a muestra de la ventana considerada *M* son menores que el umbral definido, en este caso se cuantifica un comportamiento regular de la señal. Seguidamente, un histograma es construido de la secuencia de símbolos  $S_j$ . El número de ocurrencias de cada símbolo  $S_k$ ,  $k = 0, 1, \dots, M-1$ , es el parámetro propuesto para caracterizar y cuantificar el comportamiento de las series temporales de la temperatura de consigna y la temperatura medida.

## Aproximación Simbólica: SAX

SAX es un nuevo método de representación simbólica de series de tiempo. Este enfoque simbólico es una técnica que permite que una serie temporal de longitud  $n$ , se reduzca a una cadena de longitud  $w$ , (donde  $w < n$ ).

En la Tabla 1 se presentan las notaciones utilizadas en la metodología de este trabajo.

El procedimiento de discretización utiliza una representación intermedia entre la serie de tiempo original y la cadena de símbolos. Primero, se transforman los datos a una Aproximación Finita a Trozos (PAA) y luego la representación PAA se transforma a una nueva representación discreta de cadenas. Hay dos ventajas importantes al hacerlo: reducción de la dimensión y límites inferiores de las distancias medidas. Con más detalle el método SAX establece dos etapas:

### Representación PAA

Se define cuando los datos de una serie de tiempo de longitud  $n$  se dividen en segmentos de tamaños iguales de longitud  $w$ .

El  $i$ -ésimo elemento de  $\bar{C}$  es calculado por la siguiente ecuación:

$$\bar{c}_i = \frac{w}{n} \sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} c_j \quad 4)$$

Donde se calcula la media de los datos comprendidos en cada uno de los segmentos, siendo la representación PAA el promedio de cada segmento. La representación PAA puede ser visualizada como se muestra en la figura 4. En busca de la sencillez y la claridad se supone que  $n$  es divisible entre  $w$ .

### Discretización

La discretización consiste en aplicar una nueva transformación a los coeficientes obtenidos de la representación PAA con la finalidad de llevarlos a una palabra compuesta de símbolos. Con este propósito, es conveniente aplicar una técnica de discretización que produzca símbolos con igual probabilidad. En esta etapa se calculan los puntos de interrupción que originan áreas de un mismo tamaño bajo una curva de Gauss<sup>5</sup>, por lo que es necesaria la normalización previa de los datos. Antes de proseguir con el método hacemos las siguientes definiciones:

**Definición 1.** Puntos de Interrupción: Es una lista ordenada de números  $\beta = \beta_1, \dots, \beta_{a-1}$  tal que el área bajo una curva de

Gauss con distribución  $N(0,1)$  desde  $\beta_i$  a  $\beta_{i+1} = \frac{1}{a}$

Estos puntos de interrupción pueden determinarse de una tabla estadística, lo cual permite asociar cada valor real a un símbolo que corresponde a un alfabeto de tamaño  $a$  previamente definido por el usuario. Por ejemplo en la Tabla 2 se dan los puntos de interrupción para tamaños de alfabeto a de 3 a 10.

Una vez que los puntos de interrupción se han obtenido, se puede discretizar la serie de tiempo de la siguiente manera: Primeramente, se obtiene una representación PAA de la serie de tiempo, todo coeficiente PAA que se encuentre por debajo del punto de interrupción más pequeño es reemplazado por el símbolo "a", todo coeficiente superior o igual al de corte más pequeño y menor al segundo más pequeño es reemplazado por el símbolo "b", y así sucesivamente. La Figura 5 ilustra la idea.

**Definición 2.** Palabra: Agrupación de símbolos obtenidos de un alfabeto que permite resumir la información contenida en la serie de tiempo original.

Dada una secuencia  $C$  de longitud  $n$  puede ser representada a una palabra  $\hat{c} = \hat{c}_1 \dots \hat{c}_w$ <sup>5</sup>.

El mapeo de una aproximación PAA a una aproximación palabra  $\hat{c}$  se obtiene de la siguiente manera:

$$\hat{c}_i = \text{Alpha}_j \text{ sii } \beta_{j-1} \leq \bar{c}_i \leq \beta_j \quad 5)$$

### Distancias medidas

Con las series de tiempo se pueden definir distancias. La más común entre ellas es la distancia Euclídea. Dadas dos series de tiempo  $Q$  y  $C$  se define su distancia Euclídea por la siguiente ecuación (ver figura6):

$$D(Q, C) \equiv \sqrt{\sum_{i=1}^n (q_i - c_i)^2} \quad 6)$$

Para la representación PAA de dos series  $\bar{Q}$  y  $\bar{C}$  se obtiene una aproximación menor de la distancia Euclídea (ver figura7):

$$DR(\bar{Q}, \bar{C}) \equiv \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (\bar{q}_i - \bar{c}_i)^2} \quad 7)$$

Al transformar los datos en la representación simbólica, se puede definir una función MINDIST que devuelve la mínima distancia de dos palabras que corresponden a las series de tiempo originales (ver figura8):

$$MINDIST = \sqrt{\frac{n}{w}} \sqrt{\sum_{i=1}^w (\text{dist}(\hat{q}_i - \hat{c}_i))^2} \quad 8)$$

La función se asemeja a la ecuación (6), excepto por el hecho de que la distancia entre los dos coeficientes PAA ha sido sustituida por la función  $\text{dist}()$ . Esta puede ser aplicada usando por ejemplo la Tabla 3 para 4 símbolos.

Las distancias entre dos símbolos se pueden leer mediante el examen de la correspondiente fila y columna. Por ejemplo,  $\text{dist}(a, b) = 0$  y  $\text{dist}(a, c) = 0,67$ <sup>5</sup>. El valor en la celda  $(i, j)$  para cualquier tabla se puede calcular por la siguiente expresión:

$$\text{Cell}_{ij} \begin{cases} 0 & \text{si } |i - j| \leq 1 \\ \beta_i - \beta_{j-1} & \text{otro valor} \end{cases} \quad 9)$$

## Reducción de la numerosidad

Se ha visto que el método SAX reduce de forma significativa la dimensionalidad de los datos en una serie de tiempo. Sumado a esto, también puede reducir la numerosidad de los datos correspondientes a algunas aplicaciones extrayendo sub cadenas de la serie de tiempo original mediante el uso de ventanas deslizantes, las cuales son almacenadas sin permitir repetición de sub cadenas que se encuentren seguidas. Para dar claridad a esto, imagine que al extraer sub cadenas de una serie de tiempo después de ser convertidas en símbolos: si la primera ocurrencia es **abcabc** y la segunda ocurrencia vuelve a ser **abcabc**, entonces esta segunda aparición no se incluye.

## RESULTADOS

**DS:** Como se discutió, la dinámica simbólica depende de dos parámetros: número de muestras de la ventana usada  $M$  y un umbral  $r$ . El umbral  $r$  se expresa como la desviación estándar ( $sd$ ) del número de muestras de la ventana usada, es decir,  $r=sd$ .

Para estimar los símbolos (patrones) de la dinámica simbólica se ha escogido un tamaño de ventana  $M = 20$  muestras. En la figura 9 se presenta la transformación en un conjunto de símbolos de la serie temporal de la temperatura de consigna y temperatura medida a la salida del horno, respectivamente, al aplicar la metodología de dinámica simbólica considerando  $M = 20$  muestras. Seguidamente se generan histogramas, ver figura 10, los cuales permitirán cuantificar el número de ocurrencias de cada símbolo y la extracción de patrones.

Como se puede observar existe una diferencia marcada entre la temperatura de consigna y temperaturas medidas en la salida del horno.

En las figuras 10(a) y 10(b) se muestran los histogramas de los símbolos obtenidos para la temperatura de consigna TMPP2C y temperatura medida TMPP2M, respectivamente, a la salida del horno. Como se puede observar el símbolo  $S_{19}$  (cuyo valor numérico es 19 posee un mayor número de ocurrencias en las series de consigna, a diferencia del número de ocurrencias de este mismo símbolo en las series medidas. La presencia del símbolo  $S_{19}$  con un mayor número de ocurrencias en la series de consigna cuantifica de acuerdo al método de dinámica simbólica que las temperaturas son semejantes, a diferencia del número de ocurrencias de este símbolo en las series medidas, donde se presenta con un menor número de ocurrencias, lo que significa que existen cambios de temperaturas en la banda. Además, se puede observar en el histograma de la figura 10(b) la presencia de otros símbolos los cuales no ocurren en la serie de consigna y nuevamente muestra la presencia de cambios de temperatura durante el proceso de galvanizado.

Esto demuestra que la serie medida caracteriza una información inherente al proceso la cual no es considerada por la serie de consigna. Esta caracterización de información manifiesta que durante el proceso de galvanizado ocurren fallos los cuales mediante el análisis y procesado de las series

temporales medidas permitirían una mejor toma de decisión y control del proceso.

Los resultados presentados anteriormente al aplicar la técnica de dinámica simbólica, se han realizado en la series temporales de temperaturas de consigna y medidas con la presencia de mediciones erróneas (datos espurios) atribuidas al proceso automatizado. Con el propósito de atenuar la presencia de estos datos espurios se aplicará una técnica de filtrado a cada serie temporal de las temperaturas de consigna y medida para eliminar el ruido y obtener unas series temporales que se aproximen a la data obtenida del modelo físico que describe el proceso de galvanizado continuo por inmersión.

Son diversas las funciones y los propósitos por las que son utilizadas las funciones de filtrado, podemos mencionar: filtrado de ruido, compresión de datos, eliminación de datos innecesarios, entre otros.

En este estudio se aplicará como técnica de filtrado una estandarización de los datos que consiste en restar a cada dato de la serie temporal la media y dividirla entre la desviación estándar.

Siguiendo con la aplicación de la técnica de dinámica simbólica a los datos estandarizados, en la figura 11 se muestra la ocurrencia de los símbolos obtenidos para la temperatura de consigna y la temperatura medida a la salida del horno.

Como se puede observar en la figura 11 existe semejanza entre las temperaturas de consigna y las temperaturas medidas a la salida del horno. Esta similitud se debe a la reconciliación de la data que permite atenuar los errores de medición en las bandas. Al construir los histogramas presentados en la figura 12 se observa que existe una semejanza de los símbolos obtenidos, en este caso  $S_{17}$ ,  $S_{18}$ ,  $S_{19}$ . La presencia de los símbolos  $S_{17}$  y  $S_{17}$  manifiesta que aunque las series temporales hayan sido reconciliadas existe la presencia de cambios de temperatura durante el proceso.

**SAX:** Para estimar la distancia mínima que constituye la métrica que revela diferencias entre las series temporales de temperaturas de consigna y medidas, se ha seleccionado un tamaño para el número de segmentos  $W=20$  muestras. Al efectuar la previa normalización de los datos donde se está realizando un refinamiento de los mismos como lo exige el método y comprobarse una reducción de su dimensionalidad al convertir las 56.376 muestras que constituyen la base de datos en 2819 símbolos para  $W=20$ , se consiguen diferencias (distancias) lo que muestra un patrón que refiere que la data medida detecta anomalías en el proceso de galvanizado que la data de consigna no la captura.

En la Tabla 4 se presentan las distancias mínimas obtenidas al comparar las series temporales de temperaturas de consigna (TMPP2C) y temperaturas medidas (TMPP2M) a la salida del horno, considerando un tamaño de segmento  $W=20$  muestras.

En apoyo a lo anterior, se aplicó la representación SAX por medio del uso de ventanas deslizantes de tamaño  $n=10$ , a lo largo de las 56.736 muestras. Para este caso se obtiene  $(N-n+1)$  cantidad de ventanas, es decir 56.727, dando como

resultado valores de distancias nulos cuando se está en zonas donde la data es igual o similar y en otros casos valores distintos de ceros indicando que en la zona existe disimilaridad en los datos que componen las series de tiempo utilizadas en este trabajo, facilitando la ubicación del sitio donde ocurre la irregularidad, ver figura 13.

En busca de una mejor visualización, se tomó una muestra de 1000 valores de distancias al comparar la temperatura medida (TMPP2M) y temperatura de consigna (TMPP2C), considerando el tamaño de ventana ya definido  $n=10$ .

En la figura 14 con  $n=10$  el cambio se nota en 2 partes: primeramente, durante las ventanas: (491-507) y luego una diferencia más pequeña en las ventanas: 962, 964, 966, 968, 970 las cuales van disminuyendo para ventanas con mayor tamaño.

Esto confirma que se está en presencia de datos que difieren entre sí, sin embargo cabe destacar que el uso de ventana de tamaño  $n=10$  logra detectar cambios en la data comparada. En efecto, esta anomalía ocurre cuando la temperatura de consigna difiere bruscamente de la temperatura medida en una de las bobinas.

## CONCLUSIONES

En este trabajo se han presentado dos nuevas técnicas de la minería de datos: el método de la dinámica simbólica (DS) y el método de aproximaciones simbólicas (SAX), los cuales podrían ser potencialmente aplicados para la identificación de patrones en series temporales.

El análisis de las series temporales de las temperaturas de consigna y temperaturas medidas a la salida del horno, mediante el método DS, permitió la obtención de unos índices o símbolos, los cuales cuantifican diferencias entre las temperaturas durante el proceso industrial de galvanizado por inmersión.

Al realizar una reconciliación a las series temporales y comparando la dinámica simbólica entre las series temporales de consigna y medida se observó la presencia de unos símbolos los cuales cuantifican nuevamente diferencias entre las temperaturas.

Seguidamente, se logró verificar los dos aspectos fundamentales del método SAX: reducción de la dimensionalidad y reducción de la numerosidad de la data original en un nuevo campo de estudio como lo es la industria de galvanizado de bobinas. En un espacio simbólico de menor dimensión se encuentran diferencias en las series temporales que describen el proceso, permitiendo ubicar a través del número de ventana en que parte está ocurriendo el comportamiento anómalo originado por varias razones, tales como: fallas, microparadas, rupturas de bobinas, entre otros.

El grado de efectividad con que se aplique el ciclo térmico para cada bobina durante el proceso de galvanizado es de vital importancia para obtener óptimas propiedades finales del acero, por lo tanto, los métodos DS y SAX podrían ser utilizados como técnica de monitorización y supervisión que permitirían analizar, controlar y predecir de la mejor manera posible las señales de consigna para cada tipo de bobina dentro

de la zona de calentamiento del horno para una mejor toma de decisión y control del proceso industrial.

## REFERENCIAS

1. **GUILLÉN, P., JUGO, D., CARRASCO, H., ROJAS, R.:** "Dinámica no lineal para caracterizar la variabilidad del ritmo cardiaco en pacientes chagásicos", *CIENCIA* 11(4), 276-283, 2003.
2. **GUILLÉN P, VALLVERDÚ M., JUGO D., CARRASCO H., CAMINAL P.:** "Symbolic dynamics in heart rate variability". CIMENICS BI79-85, 2006.
3. **ESCALONA MORAN, M., COSENZA, M., GUILLÉN, P., COUTIN, P.** "Synchronization and clustering in electroencephalographic signals" *CHAOS, SOLITONS AND FRACTALS*, 31, 820-825, 2007.
4. **LIN, J., KEOGH, E., LONARDI, S. A** "Symbolic Representation of Time, with implications for Streaming Algorithms". *University of California*, 2003.
5. **LIN, J., KEOGH, E., LONARDI, S.** "Experiencing SAX: A Novel Symbolic Representation of Time Series". *DMKD Journal*, 2007.
6. **DREVER INTERNATIONAL S.A.** "Hot Dip Coating Line for ACERALIA". *Documentación Técnica del Modelo Matemático*, Spain. 030608222, Rev. 1.0, 1998.
7. **EDWARDS, W., CARLTON, A., HARVEY, G., EVANS, R., MCKERROW, P.** "Coating mass control system design for a continuous galvanizing line". Automatica Pergamon Press, Vol. 12, 225-235, 1976.
8. **MARTÍNEZ DE PISÓN, F.** "Optimización mediante técnicas de minería de datos del ciclo de recocido de una línea de Galvanizado". *Tesis de Dr. en Ingeniería Industrial*, Universidad de La Rioja, Logroño, España, 2003.

## AUTORES

**Beatriz Nava.** Ingeniero de Sistemas, Estudiante de la Maestría en Modelado y Simulación de Sistemas, Universidad de Los Andes, Centro de Simulación y Modelos, CESIMO, Mérida, Venezuela.

**Mayra Herrera.** Ingeniero de Sistemas, Universidad de Los Andes, Centro de Simulación y Modelos, CESIMO, Mérida, Venezuela.

**Pablo Guillén Rondón.** Licenciado en Matemática, M.Sc. en Matemática Aplicada, DEA en Ingeniería Biomédica, Dr. en Ingeniería. Prof. Agregado, Universidad de Los Andes, Centro de Simulación y Modelos, CESIMO, Mérida, Venezuela.

**Francisco Martínez de Pisón Ascacíbar.** Ingeniero Industrial, Dr. en Ingeniería Industrial. Prof. Titular, Universidad de La Rioja, Logroño, España.

Temperatura Objetivo de la Banda en el Pirómetro 2 para cada bobina

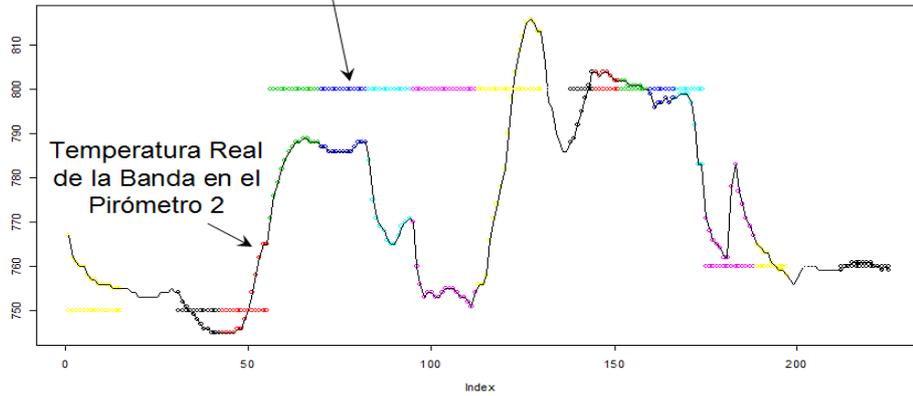


Figura 1. Series temporales de la temperatura de consigna objetivo (TMPP2C) y temperatura real (TMPP2M) de bobinas de acero a la salida de un horno de galvanizado

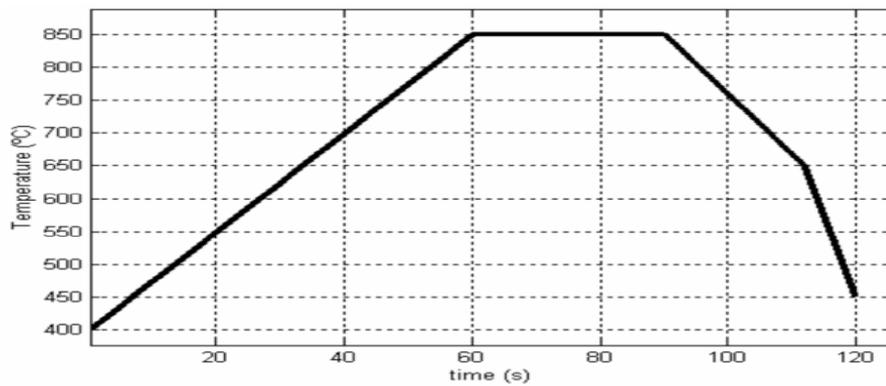


Figura 2. Ejemplo de ciclo de recocido aplicado a un acero para la mejora de sus propiedades.

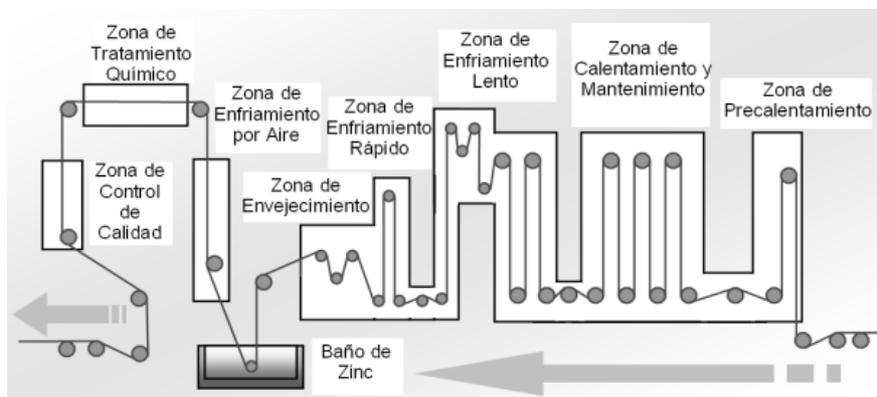


Figura 3. Esquema de una Línea de Galvanizado

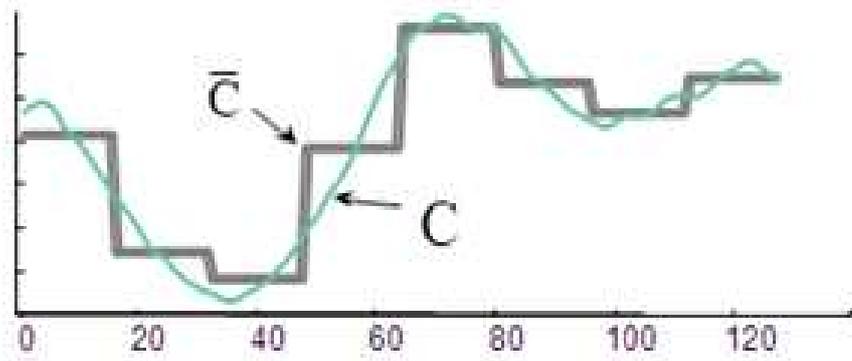


Fig. 4: Visualización de la representación PAA. En este caso una serie de tiempo de longitud  $n=128$  es reducida a  $w=8$  segmentos

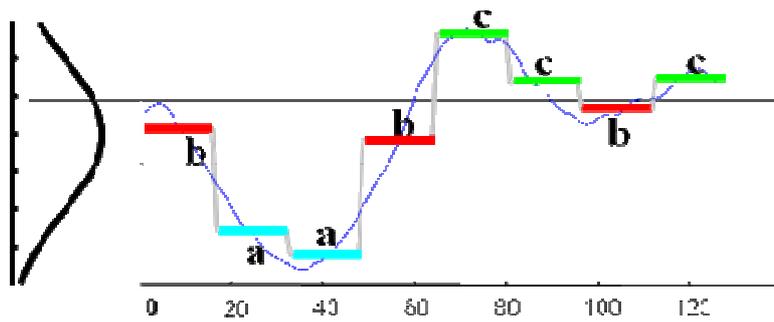


Fig. 5: Visualización de la discretización de una serie temporal con  $n = 128$ ,  $w = 8$  y  $a = 3$ , formando la palabra baabccbc.

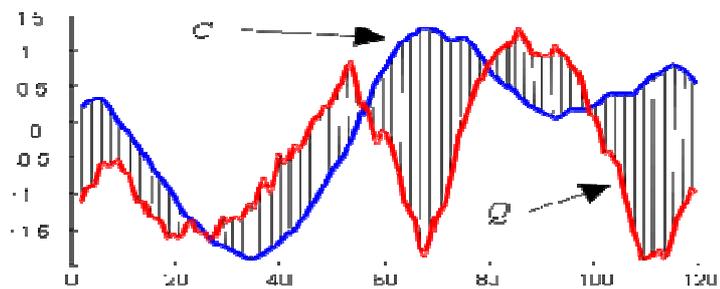


Fig. 6: Visualización de la distancia Euclídea entre dos series de tiempo  $Q$  y  $C$

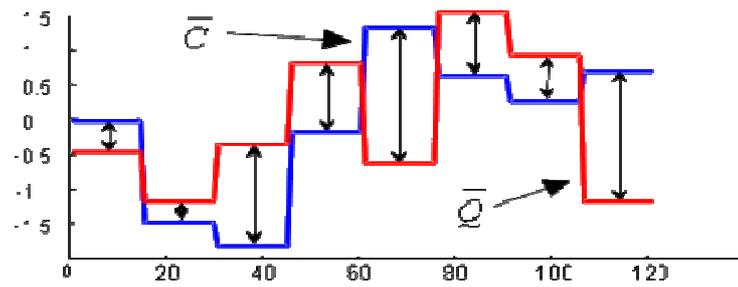


Fig. 7: Visualización de la distancia medida para la aproximación PAA.

$$\begin{array}{l}
 \hat{C}^r = \mathbf{baabcabc} \\
 \updownarrow \updownarrow \updownarrow \updownarrow \updownarrow \updownarrow \updownarrow \updownarrow \\
 \hat{Q} = \mathbf{babacca}
 \end{array}$$

Fig. 8: Visualización de la distancia entre dos representaciones SAX (distancia mínima entre dos series de tiempo Q y C)

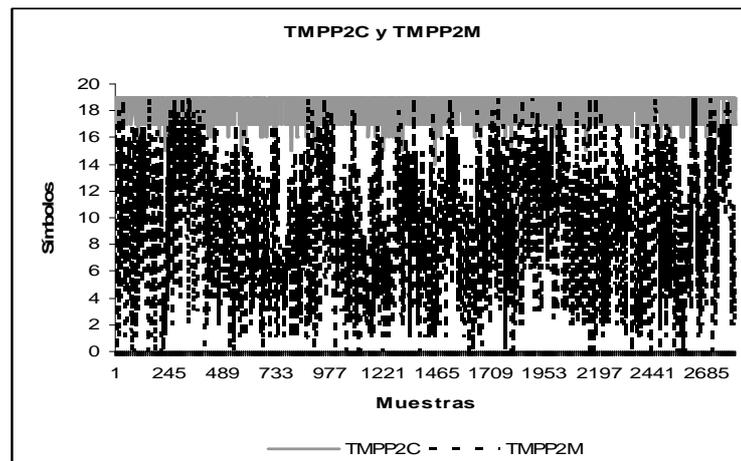


Fig. 9: Símbolos obtenidos considerando M=20 muestras para TMPP2C y TMPP2M

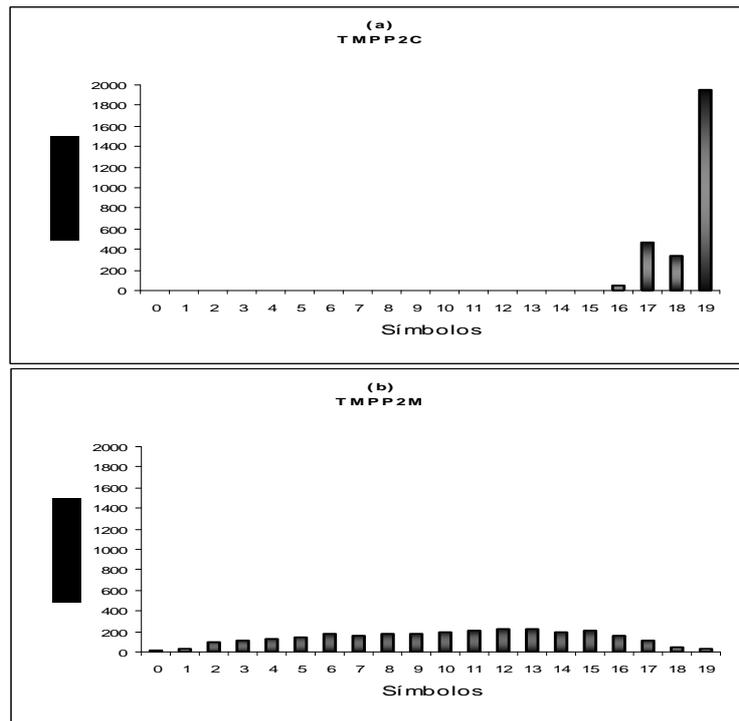


Fig. 10: Histograma de los símbolos obtenidos considerando  $M = 20$  muestras.

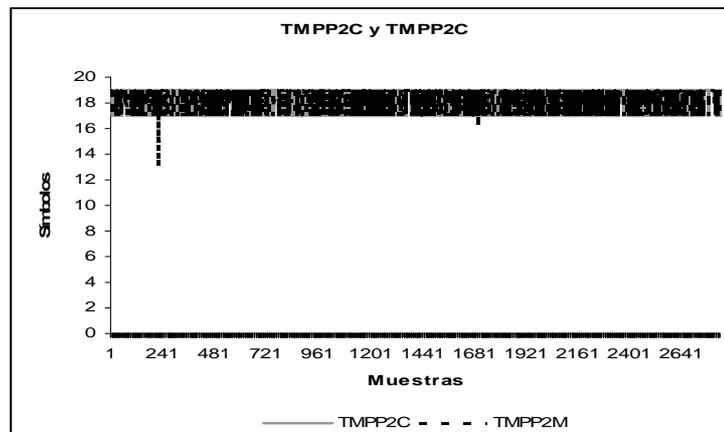


Fig. 11: Símbolos obtenidos considerando  $M = 20$  muestras para TMPP2C y TMPP2M.

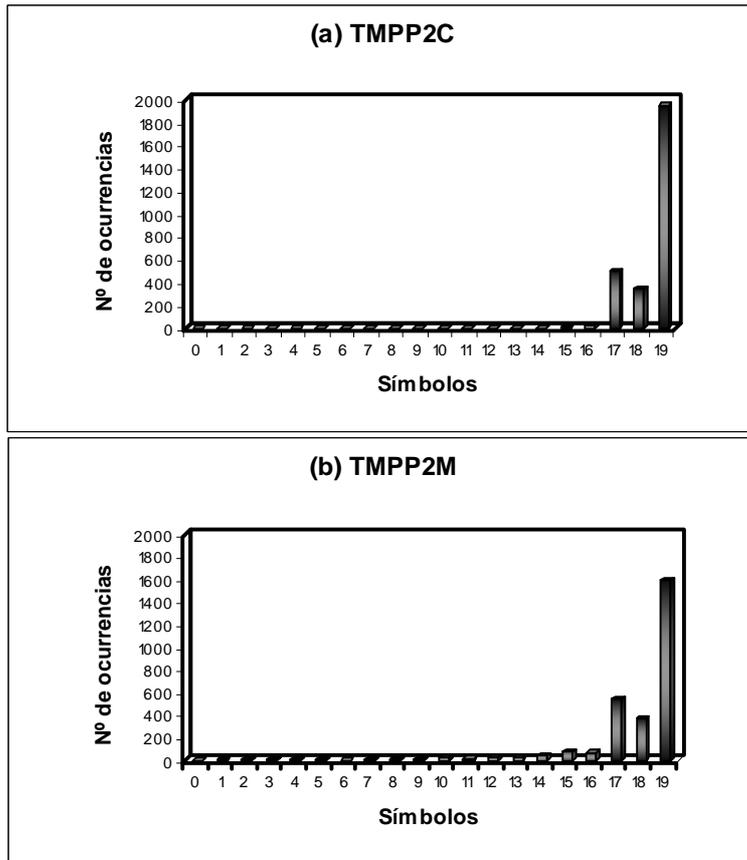


Figura 12. Histograma de los símbolos obtenidos considerando M=20 muestras.

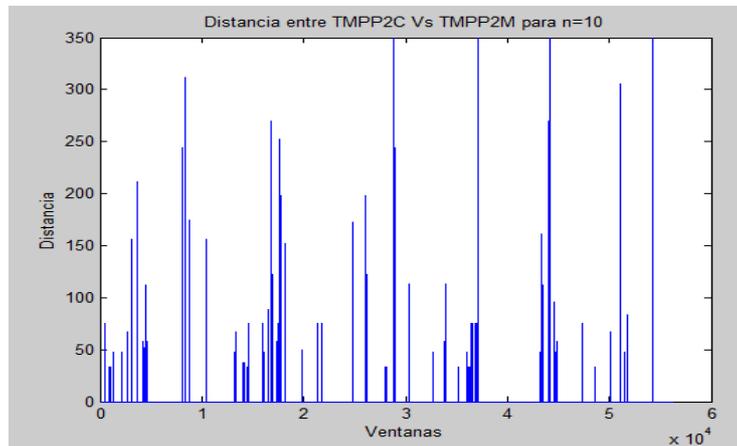


Fig. 13: Gráfica que corresponde a las distancias de N=56.727 muestras entre TMPP2C y TMPP2M



Tabla 3. Tabla de búsqueda utilizada por la función MINDIST. Corresponde a un alfabeto de tamaño 4, es decir,  $a = 4$ .

	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
<b>a</b>	0	0	0.67	1.34
<b>b</b>	0	0	0	0.67
<b>c</b>	0.67	0	0	0
<b>d</b>	1.34	0.67	0	0

Tabla 4. Distancias mínimas con  $a=8$

DISTANCIA MÍNIMA TMPP2C Vs TMPP2M	
$W=20$	26.2