

MODELIZACION DE SERIES DE TIEMPO NO LINEALES DE SO₂ EN SALTA CAPITAL, ARGENTINA, MEDIANTE EL EMPLEO DE PERCEPTRONES MULTICAPA

O. Avila Blas¹, H. Musso de Dip²

1. Departamento de Matemática. Facultad de Cs. Exactas, Universidad Nacional de Salta
2. Departamento de Química, Facultad de Cs. Exactas, Universidad Nacional de Salta
Avenida Bolivia 5150, 4400 Salta, Argentina
Tel.: 54-387-4258612 – Fax: 54-387-4255449
E-mail: ojblas@yahoo.com.ar, haydeemusso@gmail.com

Recibido:5-11-10; Aceptado:30-11-10.

RESUMEN.- En este trabajo se realizó un estudio estadístico de variables físico químicas asociadas al fenómeno de contaminación ambiental, en particular concentración media mensual de SO₂, medidas en la ciudad Salta Capital, Argentina, simultáneamente a concentraciones de NO₂ y O₃. Las series bajo estudio presentaban comportamientos dinámicos no lineales, datos atípicos y cambios estructurales, lo que hizo imposible modelarlas con tipologías econométricas tradicionales (AR, MA, ARMA, ARIMA, entre otras). Una solución eficiente que se encontró, hace uso de la teoría de los perceptrones multicapa. Mediante el modelo estructural de series de tiempo, esta solución se presenta como un proceso matemático iterativo que permite obtener un modelado final el cual tiene una muy alta confiabilidad (95%) para realizar generación de valores sintéticos de la variable estudiada.

Palabras claves: contaminación ambiental, estadística, redes neuronales, modelado estructural, filtro de Kalman, pronósticos

NON LINEAR SO₂ TIME SERIES MODELLING IN SALTA CAPITAL, ARGENTINA, USING MULTILAYERED PERCEPTRONS

ABSTRACT.- In this paper a statistical study of physical chemistry variables associated to air pollution phenomena, specially monthly average concentration values of SO₂ measured in Salta Capital, Argentina, together with NO₂ and O₃ concentrations, is done. Series under consideration have non linear dynamic behavior, outliers and structural changes that made impossible to get a model using traditional econometric models (AR, MA, ARMA, among others). An efficient solution considering the multicape perceptrons theory, was found. With the structural time series modelling, the solution is presented as an iterative mathematical process that allows us to obtain a final model that has a very high confidence level (95%) in order to generate synthetic values of the variable under treatment.

Keywords: air pollution, statistics, neural networks, structural modelling, Kalman filter, forecasting

1. INTRODUCCION

Las tres herramientas principales para evaluar la calidad del aire propuestas en (OMS, 2000; OMS, 2004) son: monitoreo del ambiente, modelos e inventario de mediciones.

Los resultados de los modelos de dispersión son útiles para predecir los patrones de dispersión y de deposición de los contaminantes lo que contribuye a identificar las áreas de mayor impacto. Ningún programa de monitoreo puede aspirar a cuantificar en forma integral los patrones de contaminación del aire en espacio y tiempo para definir la exposición de una población en una ciudad. Por ello el monitoreo debe combinarse con otras técnicas de evaluación como la elaboración de modelos, la medición y elaboración de inventarios de emisiones, la interpolación, correlación y elaboración de mapas.

Un inventario completo de emisiones para una ciudad es una tarea titánica ya que requiere determinar las fuentes de emisiones puntuales, de áreas y móviles. El dióxido de azufre (SO₂) se produce generalmente por la quema de combustibles que contienen azufre y por la producción de energía en las plantas termoeléctricas, además de los vehículos automotores, por lo que las concentraciones más altas de este dióxido se presentan en las áreas de mayor actividad industrial y tránsito vehicular. Contribuye a formar la lluvia ácida y es un factor perjudicial en el hecho de que el dióxido de azufre es un precursor del ozono (O₃).

En altas concentraciones el dióxido de azufre puede ocasionar dificultad para respirar, humedad excesiva en las mucosas de las conjuntivas, irritación severa en vías respiratorias e incluso al interior de los pulmones por formación de partículas de ácido sulfúrico, ocasionando vulnerabilidad en las defensas.

El dióxido de azufre es causante de enfermedades respiratorias como broncoconstricción, bronquitis y traqueatitis, pudiendo llegar a causar broncoespasmos en personas sensibles como los asmáticos, agravamiento de enfermedades respiratorias y cardiovasculares existentes y la muerte; si bien los efectos señalados dependen en gran medida de la sensibilidad de cada individuo, los grupos de la población más sensibles al dióxido de azufre incluye a los niños y ancianos, a los asmáticos y a aquellos con enfermedades pulmonares crónicas como bronquitis y enfisema.

La combinación de óxidos de azufre y partículas suspendidas actúan sinérgicamente produciendo un efecto combinado mucho más nocivo que el efecto individual de cada uno de ellos por separado. En niveles bajos de exposición $< 50 \mu\text{g}/\text{m}^3$ (media anual), se han observado efectos como los informados por Kamat y Doshi (OMS, 2000), prevalencia de: resfriados frecuentes, desórdenes cardíacos, disnea, entre otros.

Se tienen como antecedentes referentes de trabajos con métodos estadísticos de generación de series de algunas variables asociadas a clima, tales como temperatura de superficie (Avila Blas, 1997), radiación horaria (Aguiar, 1992), utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA(p,d,q).

Otros trabajos previos conocidos hasta el momento, de modelado de series de tiempo con la metodología del enfoque estructural introducido por estadísticos contemporáneos (Harvey y Shepard, 1993; Abril, 1997), se refieren a la variables radiación promedio diaria, temperatura de superficie promedio diaria y correlaciones entre ellas (O. Avila Blas *et al.*, 1999; O. Avila Blas *et al.*, 2000). En estos casos se logró diseñar estrategias estadísticas para trabajar en ausencia de normalidad en la distribución de las variables bajo estudio, con lo que consiguió una mayor robustez estadística a las modelizaciones, al mismo tiempo que una flexibilidad en su tratamiento fundamentalmente desde el punto de vista de la generación de valores sintéticos, con una alta confiabilidad (95%).

2. METODOLOGIA PARA LA OBTENCIÓN DE LOS DATOS

Una técnica simple para la determinación de SO_2 en la atmósfera usa el muestreador pasivo, basado en la difusión del gas en un tubo colector. Es una técnica económica y eficiente que no requiere mantenimiento alguno, ni utilización de fuentes de energía, dado que permite recoger el SO_2 a través de la difusión molecular.

De acuerdo al organismo internacional GEMS los muestreadores pasivos son especialmente aptos tanto para el estudio de nivel basal de contaminantes como para el monitoreo simultáneo a gran escala (UNEP - WHO, 1994). Se utilizaron tubos tipo Palmes de acrílico (Palmes *et al.*, 1976; Palmes y Tomczyk, 1979), construidos en nuestro laboratorio con tubos de acrílico de 7 cm de longitud y 12 mm de diámetro interno. En un extremo, sostenidos con una tapa de polietileno, se ubicaron dos discos de malla de acero inoxidable (0,2 x 0,2 mm) humedecidas con una mezcla trietanolamina (90% p/v), mientras que el otro extremo se selló con otra tapa similar.

El sistema muestreador, que consta de tres tubos idénticos, tal como se muestra en la Figura 1, se coloca a 2 m de altura para disminuir el riesgo de pérdida, teniendo en cuenta las recomendaciones en (OMS, 2000) para la ubicación de muestreadores pasivos. El fenómeno difusional es independiente de la fuerza de gravedad y la ubicación del muestreador con su extremo abierto ubicado hacia abajo, previniendo la contaminación con gotas de agua y partículas de polvo.

Todas las partes que componen el muestreador fueron limpiadas a conciencia usando un baño de ultrasonido con agua destilada sobre vidrio y mantenidos en ella hasta el momento de secarlas a 105°C previo a su armado.

Los tubos de acrílico y las mallas fueron reutilizadas, no así las tapas que contenían el reactivo absorbente. Con cada serie de muestreadores se prepararon dos blancos, que se mantuvieron en congelador durante todo el período de muestreo, para evitar su contaminación por la atmósfera del laboratorio, (Gair *et al.*, 1991).



Fig. 1: Juego de muestreadores por triplicado para NO_2 , SO_2 y O_3

El análisis del SO_2 absorbido se realizó colocando las tapas reactivas en recipientes de plástico, haciendo reaccionar el SO_2 con:

- $\text{H-CO-H} + \text{H}_2\text{O} \rightarrow (\text{H}_2\text{C}(\text{OH})_2)$
Metilenglicol
- $(\text{H}_2\text{C}(\text{OH})_2) + \text{SO}_2 \rightarrow \text{CH}_2\text{OH.O.SO}_2$
Sulfonmetilenglicol
- $\text{CH}_2\text{OH.O.SO}_2 + \text{p-rosanilina } 0,025\% \rightarrow$
complejo sulfónico de color púrpura que absorbe a 475 nm.

Para la calibración se usó un juego de patrones en un rango de 8 a 40×10^{-6} M de sulfito de sodio (preparado y valorado en el momento del análisis) en un espectrofotómetro GBC UV-VIS 918, alcanzándose un límite de detección de $4-6 \mu\text{g SO}_2/\text{m}^3$.

Los valores medidos corresponden a concentraciones $\mu\text{g SO}_2/\text{m}^3$ obtenidas en la zona céntrica de la ciudad de Salta, desde el año 2004 hasta el año 2009, observándose un incremento importante durante el año 2006 en algunos sitios de muestreo y con tendencia a disminuir. En general las

concentraciones medidas corresponden a las informadas en la mayoría de las ciudades del mundo.

3. AREA DE ESTUDIO

La ciudad de Salta, capital de la provincia del mismo nombre, se encuentra ubicada en la región noroeste de la República Argentina; presenta un trazado típico del plano colonial español, con un centro comercial donde se concentran las actividades administrativas, comerciales y educativas. Aun conserva el diseño arquitectónico colonial de calles angostas y edificación baja. Es una de las urbanizaciones más importantes del Noroeste Argentino. Está construida en el extremo norte de la depresión tectónica del Valle de Lerma, a 1100 m (snm) (Figura 2) y rodeada por elevados cordones montañosos que disminuyen de altura de Oeste a Este y de Sur a Norte. Los relieves positivos, que no sobrepasan los 2500 m s.n.m., definen el microclima de la región.

El Valle de Lerma presenta un clima tropical serrano con estación seca de abril a noviembre, con una media mínima (30 años) de 3 mm, concentrándose las lluvias, breves y torrenciales, en los meses de diciembre a marzo, sin sobrepasar los 500 mm anuales. Como consecuencia durante la mayor parte del año las precipitaciones no contribuyen a la eliminación de los contaminantes atmosféricos.

La temperatura media anual es de aproximadamente 17 °C, con una mínima media de 10 °C en el mes de julio, y una máxima media de 21 °C en diciembre. Las temperaturas diurnas y nocturnas difieren mucho, sobre todo en la época invernal, con una amplitud aproximada de 27°C.

El Valle de Lerma se caracteriza por tener aproximadamente 56 % de días al año de calma, con una preponderancia de los vientos del NE y N durante todo el año. Las velocidades medias anuales (para un período de 27 años) aun en las direcciones predominantes son muy bajas: 5,4 km/h para los vientos del NE y 1,2 km/h para los del N. Estos vientos débiles tienen baja capacidad de dilución de contaminantes.

Todas estas características facilitan la formación de una capa de inversión térmica sobre la ciudad en la cual los contaminantes están homogéneamente distribuidos. La formación y persistencia de este domo se puede observar fácilmente desde el cordón montañoso oriental, virtualmente en cualquier época del año. El fenómeno de inversión térmica se presenta con mayor frecuencia en la época invernal, aproximadamente entre los 1400 a 1600 m de altura s.n.m.

Estos factores geográficos y climáticos propician la acumulación de contaminantes en la baja atmósfera, entre ellos los óxidos de azufre. La principal fuente antropogénica de estos gases es el tránsito de vehículos, por tratarse de una región de 700000 habitantes con un parque automotor de 250000 unidades, de neto carácter agrícola y de escasa actividad fabril. La incineración de residuos municipales se discontinuó en 1994, aunque persiste la costumbre de quemar los pastos naturales de los cerros circundantes, causando a veces incendios incontrolables en los montes. Se usa energía hidroeléctrica, y la calefacción, que sólo abarca unos dos meses del año, utiliza gas natural, el combustible fósil menos contaminante.

Mediciones de flujo de vehículos realizadas en el año 2008 en el microcentro de la ciudad arrojan una media horaria de un 67 % de autos, 19 % de camionetas, 14 % de motos, 2 % ómnibus y 1% camiones (de pequeño o mediano porte). Los automóviles con encendido electrónico se comercializaron desde 1992, con catalizadores recién a partir de 1995, siendo un 70% accionados con gas natural comprimido. La combinación de las variables geográfico-climáticas y las provenientes de la actividad humana hacen difícil prever la magnitud de la contaminación por SO₂, lo que tornó necesario su estudio. El diseño de la toma de las muestras se basó en efectuar desde el año 2004 un monitoreo mensual, sistemático y representativo de los niveles de concentración de los óxidos de nitrógeno, ozono y dióxido de azufre en un área de 80 km², mediante 15 muestreadores pasivos difusionales, distribuidos sobre una transecta Norte-Sur y Este-Oeste con intersección en el centro comercial de ciudad, cubriendo 1 km² de la zona céntrica con seis sitios de muestreo (dentro de los cuales se encuentra el sitio seleccionado para este estudio) y los otros distribuidos en la zona urbana, suburbana y rural que rodea a la ciudad. La toma de datos se continúa en la actualidad, contando con medidas verificadas e incorporadas a la base de datos, hasta el año 2009.

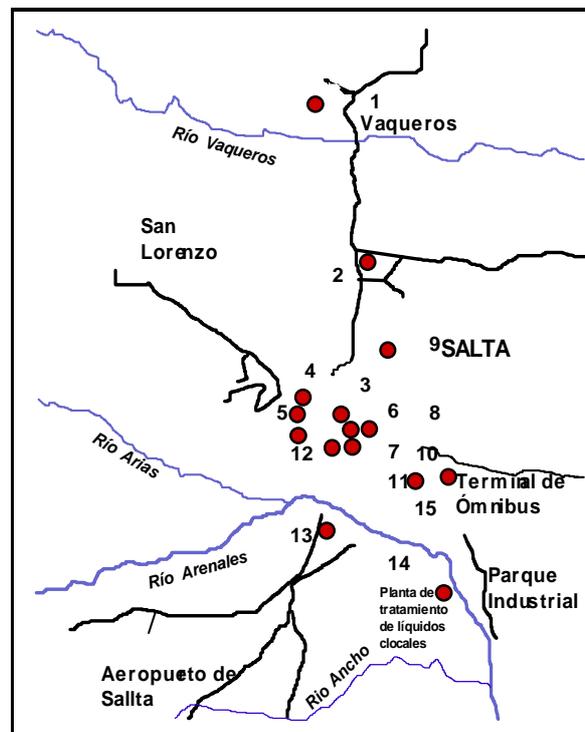


Fig. 2: Ubicación de los puntos de muestreo en el mapa de la ciudad de Salta.

4. TRATAMIENTO ESTADÍSTICO EMPLEADO

Para la simulación algunos tipos de sistemas dinámicos, como los inherentes a estudios de contaminación ambiental, son necesarias varias variables tanto químicas como meteorológicas (radiación solar, temperatura ambiente y vientos). Estas variables están correlacionadas entre sí y en el tiempo. Algunos científicos trabajan con modelos de tipo Box-Jenkins, pero éstos adolecen de algunas fallas tales como la arbitrariedad en la asignación de factores de peso a

los valores medios y la pérdida de algunas correlaciones. Este problema se agudiza cuando se trata de valores horarios (calculados a partir de datos diarios) y para series de días.

A través de los años se han propuesto métodos estadísticos de generación de series de algunas variables de interés: temperatura de superficie (Abril, 1999), radiación horaria (Avila Blas, 1997), utilizando modelos basados en el tratamiento de Box y Jenkins, es decir, modelos ARIMA (p,d,q). Los trabajos previos conocidos hasta el momento, de modelado de series de tiempo con la metodología del enfoque estructural introducido por estadísticos contemporáneos (Aguiar y Collares Pereira, 1992; Harvey y Shepard, 1992; Koopman, 1995), se refieren a la variables radiación promedio diaria, temperatura de superficie promedio diaria y correlaciones entre ellas (Avila Blas, 2001; Avila Blas y Lesino Garrido, 1987; Avila Blas *et al.*, 1999; Avila Blas *et al.*, 2000; Avila Blas y Grossi Gallegos, 2002; Avila Blas y Grossi Gallegos, 2003). En estos casos se logró "levantar" la condición de normalidad en la distribución de las variables de estudio, lo que le brindó una mayor robustez estadística al proceso estudiado, al mismo tiempo que una flexibilidad en su tratamiento fundamentalmente desde el punto de vista de efectuar pronósticos con una alta confiabilidad (95%).

La idea básica de los modelos estructurales de series de tiempo es que ellos pueden ser puestos como modelos de regresión en donde las variables explicativas son funciones del tiempo, con coeficientes que pueden cambiar a través del tiempo. La estimación actual de los coeficientes ó filtrada, se logra poniendo al modelo en forma de espacio de estado y aplicándole luego el denominado Filtro de Kalman (Aguiar y Collares Pereira, 1992). Se emplean algoritmos específicos para hacer predicciones y para los suavizados. Esto último significa computar el mejor de los estimadores en todos los puntos de la muestra usando al conjunto de observaciones. La magnitud por la cual los parámetros pueden variar está gobernada por los llamados hiperparámetros. Estos pueden ser estimados por el método de máxima verosimilitud, construyendo la función específica a optimizar.

La metodología de trabajo en la selección para los modelos denominados estructurales, es diferente en muchos aspectos al tratamiento clásico de Box y Jenkins de los modelos ARIMA(p,d,q). Se pone menos énfasis en la observación del correlograma de diversas transformaciones de la serie con el objeto de obtener una especificación inicial. En lugar de ello, el énfasis está puesto en la formulación del modelo en términos de componentes cuya presencia está sugerida por el conocimiento del fenómeno bajo estudio, de sus aplicaciones o por una inspección del gráfico de la serie original. Una vez que el modelo ha sido estimado, el mismo tipo de tests de diagnóstico para los modelos ARIMA puede ser aplicado. Además el estudio se completa con tests de falta de normalidad y heterocedasticidad (varianza variable en el tiempo), tests para la calidad predictiva en períodos posteriores a la muestra y gráficos de los componentes suavizados. Por otra parte no es necesario que la serie sea estacionaria, es decir, con media y varianza constante en el tiempo, lo que sucede muy a menudo en variables asociadas a contaminación ambiental.

El tratamiento estadístico de los modelos estructurales de series de tiempo está basado en la forma de espacio de estado, el filtro de Kalman y el suavizador asociado.

La representación matemática de un modelo de espacio de estado vincula al vector de disturbios $\{\varepsilon_t\}$ con el vector de observaciones $\{y_t\}$ por medio de una caminata aleatoria Markoviana $\{\alpha_t\}$ definida por la ecuaciones:

$$\begin{aligned} y_t &= Z_t \alpha_t + G_t \varepsilon_t, & \varepsilon_t &\sim N(0, H_t) \\ \alpha_t &= T_t \alpha_{t-1} + H_t \eta_t, & \eta_t &\sim N(0, Q_t); \\ \alpha_0 &\sim N(a_0, P_0), & t &= 1, \dots, T \end{aligned} \quad (1)$$

donde α_t es el vector de estado, ε_t es un vector de disturbios o ruido blanco, y las matrices del sistema Z_t , T_t , G_t , y H_t contienen elementos que son funciones del tiempo. Los disturbios son ruido blanco mutuamente no correlacionados con medio cero y matriz de varianzas H_t . Bajo el supuesto de normalidad, los disturbios son independientes entre sí. Las matrices G_t y H_t pueden interpretarse como matrices de selección, lo que le brinda generalidad al modelo. El tratamiento de las matrices se complementa con el análisis del comportamiento asintótico del vector Ψ de hiperparámetros, el que es estimado por un método especial de máxima verosimilitud. Se emplean como estadísticos de prueba para medir la bondad de ajuste a: BS (de Bowman y Shenton) que emplea estimadores de la asimetría y de la kurtosis de los datos; el Q (de Box-Ljung) para la autocorrelación serial y el de DW (de Durbin y Watson), los que junto con otros estadísticos adicionales, califican según sus valores, al modelo como adecuado para ajustar a los datos observados.

El filtro de Kalman tiene como objetivo actualizar nuestro conocimiento del sistema cada vez que una nueva observación y_t es obtenida. Cuando el modelo ha sido puesto en forma de espacio de estado, se pueden aplicar una gran cantidad de algoritmos, siendo el punto central el mencionado filtro. Este filtro es un procedimiento recursivo que permite computar el estimador óptimo del vector de estado en el momento t , basado en la información disponible hasta el tiempo t , la que incluye a y_t . Cuando el supuesto de normalidad se deja de lado, no hay ninguna garantía de que el filtro produzca la media condicional del vector de estado.

De todos modos es todavía un estimador óptimo en el sentido de que minimiza el error medio cuadrático medio dentro de la clase de todos los estimadores lineales. El otro paso importante en este procedimiento es el denominado "suavizado", el que es una recursión hacia atrás en la que se emplean las inversas de las matrices relacionadas al proceso de filtrado. Las magnitudes suavizadas juegan el importante rol de pivotes en la construcción de tests de diagnóstico para observaciones atípicas y cambios estructurales. Cuando se trata de datos diarios, los modelos estructurales pueden ser extendidos de tal forma de permitir el manejo de los mismos introduciendo un componente diario. Este se modela en la misma forma que un componente estacional y puede permitírsele que varíe en el tiempo.

Otros componentes como la tendencia y la estacionalidad anual pueden ser introducidos como lo explicado en el modelo general. La falta de normalidad en los datos, que es el caso de nuestro serie, requiere de un tratamiento más particular que implica el uso de una función de verosimilitud que se maximiza por métodos iterativos, dado que las ecuaciones normales asociadas no son lineales. En este sentido se puede intentar tratar a los datos como provenientes de una familia exponencial; o bien, cuando los

datos tienen una distribución de "colas pesadas", se puede emplear combinaciones de distribuciones normales con la *t* de Student.

En este trabajo se manipularon series de promedios mensuales de la variable concentración de dióxido de azufre (SO₂). La base de datos analizada, contiene información desde el año 2004 hasta 2009 (con un total de T = 61 datos).

Existen meses en particular en los que la experiencia resultó fallida, y en consecuencia no se cuenta con el dato medido, y en otros casos, el dato medido resulta estar fuera del intervalo de valoración usual para el SO₂.

Este tipo de comportamientos muy particular, puede verse en la Figura 3, la que corresponde al punto de muestreo número 8 en el mapa de muestreo. Este punto es el que presenta el mayor historial estadístico en cuanto a la recolección de datos. Resultados análogos se obtienen en el modelado para los puntos de muestreo 6, 7, 10 y 11, correspondientes a sitios del micro y macrocentro de la ciudad, con alta densidad de automotores.

En esta figura se observan datos atípicos o "outliers", los que posiblemente correspondan a situaciones físicas reales observadas (como por ejemplo, bacheo de calle con alquitrán fundido, situación observada). Para su tratamiento estadístico la variable en cuestión se denomina $y_t = \text{"Svar1"}$.

Nota: Cuando en una gráfica no se indiquen unidades es porque se trata de un índice o estadístico adimensional.

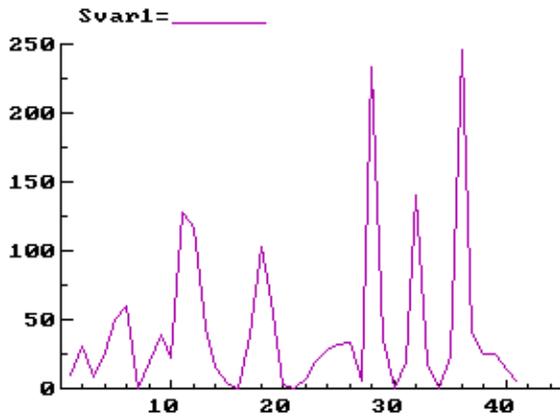


Fig. 3: Serie original de valores de concentración promedio mensual de SO₂ (µg/m³) en el punto de muestreo 6, (eje x en meses)

La figura 4, correspondiente al comportamiento de la función de autocorrelación, muestra una moderada inestabilidad durante los primeros 8 meses, al caer los valores de la misma, fuera del intervalo de confianza máximo tolerado, y se aprecia también la presencia de un ciclo subyacente. Es importante destacar que este ciclo sólo se pudo detectar con el devenir de los años, y el modelo estadístico final que permite la generación sintética de datos, que pudo describirlo mejor, es el que resulta óptimo en este trabajo.

Por otro lado, cuando se contaba con menos años de mediciones (4 años y medio aproximadamente), el modelado estructural de la serie fue arrojando diferentes patrones de comportamiento, no siempre sencillos de definir.

Aquí, uno de los puntos clave a tener en cuenta, no solo es brindar un modelado robusto desde el punto de vista de la obtención de estimadores máximo verosímiles y con buen comportamiento asintótico, sino que puedan ser controlados en su variabilidad temporal, para evitar producir valores pronosticados que no sean confiables, o bien, sean poco estables. De todos modos, como éste tipo de procesos corresponde a situaciones dinámicas, es de esperarse que con el incremento del número de datos medidos, los modelos puedan perfeccionarse y simplificarse tal vez, para que su uso práctico sea más fácil.

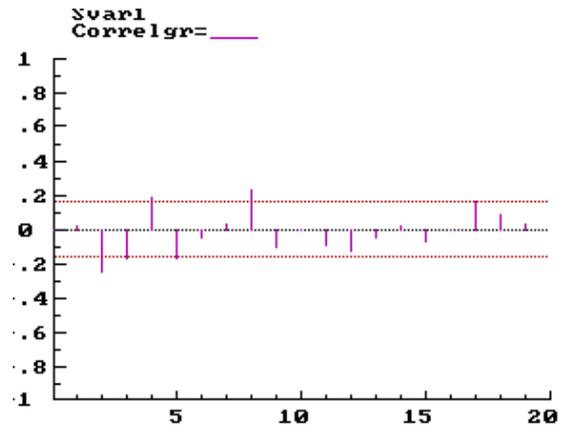


Fig.4: Correlograma muestral de la serie bajo estudio (eje y adimensional, eje x en fracciones de meses)

Las observaciones anteriores sobre la serie bajo modelado estadístico, se complementan con el estudio del espectrograma y la densidad espectral estimada, como se ve en la figura 5. En ella, aparecen picos bien definidos en los meses correspondientes a períodos secos (en general entre Mayo y Octubre de cada año), en los que aumenta considerablemente la densidad de polvo en suspensión en el aire, un mayor número de automotores que circulan por la ciudad y otros factores, entre lo que podemos citar el denominado de la capa de inversión térmica, que se produce a una altura promedio de 150 m sobre la ciudad de Salta.

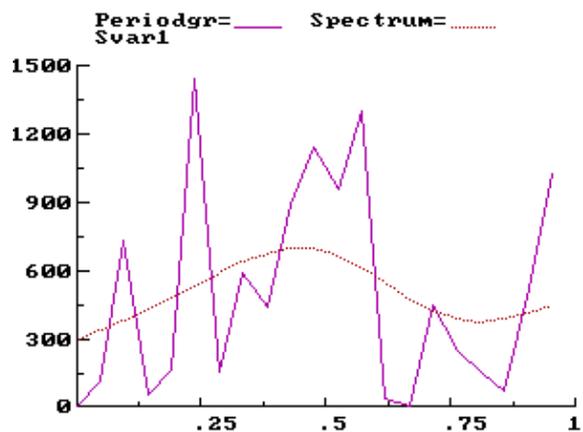


Fig.5: Periodograma y densidad espectral estimada de la serie bajo estudio (eje x en unidades enteras de $2\pi/T$)

Un hecho importante a destacar que, según se puede ver en la figura 6, los datos no presentan distribución normal, por el contrario, es muy asimétrica por la derecha. Este hecho hace que el proceso Markoviano definido en (1), deba ser redefinido y transformado matemáticamente, para poder

realizar las diferentes etapas del modelado estructural. Se destaca que en estos casos, la ausencia de normalidad en los datos, afecta notablemente el procedimiento de obtención de modelos robustos mediante el empleo del filtro de Kalman. Por ello se hace imperioso linealizar las ecuaciones a fin de poder aproximarlas a un caso cuasi lineal, aplicar el Filtro, y buscar las transformaciones inversas correspondientes para poder obtener el generador de valores sintéticos de la variable.

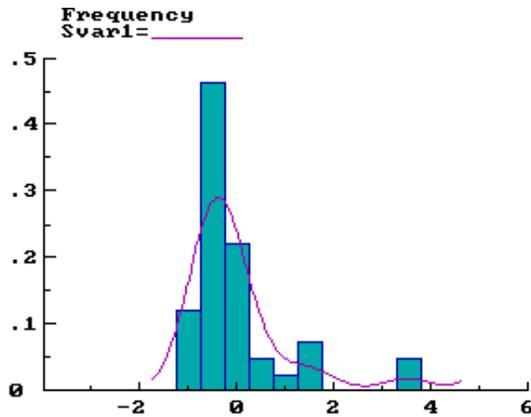


Fig. 6: Distribución empírica de la serie bajo estudio (eje y en $\mu\text{g SO}_2/\text{m}^3$, eje x en escala de tiempo estandarizada)

En la serie original de datos, se observa como un hecho destacado que, a partir de la observación número 50, aparece un importante cambio estructural en la serie bajo estudio dado por un brusco aumento relativo del valor medio en el tiempo.

Este fenómeno que se presenta en todas las series correspondientes a cada uno de los 15 sitios de muestreo, es el punto inicial para pensar en la aplicación de una metodología de modelado, en principio, diferente a la que se venía aplicando para los primeros años de trabajo (Avila Blas *et al.*, 2003; Avila Blas *et al.*, 2004; Avila *et al.*, 2007; Avila Blas y Musso, 2008).

En una primera etapa, se modeló la serie bajo estudio, en cada uno de los sitios estudiados de muestreo, mediante las tipologías econométricas tradiciones de Box-Jenkins (AR, MA, ARMA, ARIMA, GARCH, y otros).

Este procedimiento no fue fructífero debido básicamente a las siguientes causas:

- 1 La modelización lineal con la formulación de Box-Jenkins no fue eficiente
- 2 Se obtuvieron varianzas residuales muy grandes
- 3 No se logró realizar una identificación adecuada de cambios de nivel estructural

En pos de salvar estas deficiencias, se buscó una solución mediante métodos que usan fuertemente la propiedad de no linealidad de las series (como es el caso de nuestra variable de estudio, valores medios de SO_2).

El modelo óptimo logrado es uno que emplea la modelización de series de tiempo, en forma estructural, utilizando una transformación no lineal de tipo logístico, para la cual previamente se consigue encontrar el representante maximal, en el sentido de minimizar la

varianza del disturbio o ruido blanco. En este sentido, se consiguió formular el siguiente:

Teorema:

Un perceptrón multicapa con una capa oculta, usando la función de transferencia lineal para activar la capa de salida, se puede representar mediante el siguiente proceso aleatorio iterativo de orden pq implícito y dependiente de los disturbios o ruido blanco asociados a un proceso markoviano estructural, como sigue:

$$Z_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} x_t^{(i)} \right] + \varepsilon_t \quad (2)$$

$$g(u) = \frac{2}{1 + \exp(-u)} - 1 \quad (3)$$

La función g , que es de la familia logística permite estimar un modelo no lineal, para el que, el correspondiente modelo lineal asociado tiene regresores que son tomados como entradas a la red, generando así un proceso recursivo. Los esquemas que representan las tres formas básicas de redes neuronales que se emplearon son los mostrados en la figura 7.

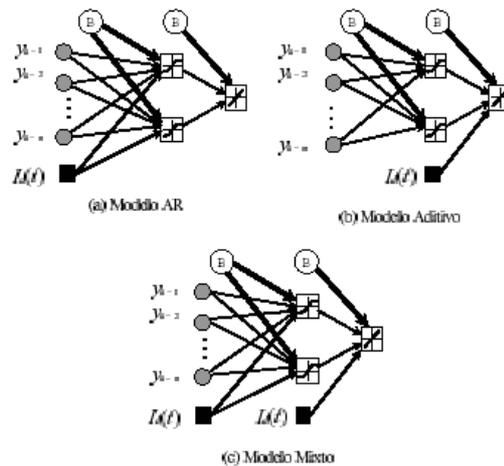


Fig. 7: Modelos básicos de redes neuronales empleados

El proceso se detiene cuando al realizar las correspondientes pruebas no paramétricas (estimación de parámetros e hiperparámetros), el MSC es mínimo en el sentido de mínimos cuadrados.

Se demostró que para un perceptrón multicapa con 2 neuronas en la capa oculta asociada a la entrada, la modelización de la serie bajo estudio es altamente confiable (95%).

Las ecuaciones no lineales asociadas a cada tipo de modelo presentado en la Figura 7, se pueden caracterizar mediante las siguientes regresiones dinámicas (con coeficientes que varían en el tiempo) respectivamente:

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-1} + \phi_1 \cdot I_d(t) \right] + \varepsilon_t$$

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-1} \right] + \phi_1 \cdot I_t(t) + \varepsilon_t$$

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j g \left[\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} + \phi_1 I_d(t) \right] \phi_2 I_d(t) + \varepsilon_t \quad (4)$$

Las etapas realizadas para proceder a estimar los parámetros e hiperparámetros de los modelos propuestos, fueron las siguientes (en base a las normas ARIMA de Box-Jenkins):

- 1 Construcción del modelo no lineal mediante la aplicación recursiva de la formulación del Teorema
- 2 Análisis residual

Para la construcción de la serie no lineal asociada al modelo, se formuló la serie auxiliar:

$$Z_t = (\phi_{1,0} + \phi_{1,1} \cdot Z_{t-1})(1 - G(Z_{t-1}; \gamma, c)) + (\phi_{2,0} + \phi_{2,1} \cdot Z_{t-1})(1 - G(Z_{t-1}; \gamma, c)) + \varepsilon_t \quad (5)$$

con

$$G(s_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma \cdot (s_t - c))} \quad (6)$$

A fin de minimizar la varianza residual, se tomaron varios juegos de valores particulares de los parámetros $\phi_{i,j}$, siendo en particular, los óptimos, los que se muestran en la tabla 1.

Tabla 1: Parámetros seleccionados para el Modelo No Lineal

$\phi_{1,0}$	$\phi_{1,1}$	$\phi_{2,0}$	$\phi_{2,1}$	γ	c
1.5	-0.5	-1.5	0.5	1.5	0.5

Para éstos, se obtuvieron los siguientes valores del error cuadrático medio (MSE), que confirman que la serie tratada con un modelo neuronal mixto, es la que tiene una mayor confiabilidad en cuanto a los pronósticos que se pueden hacer con ella (menor MSE). Los valores se muestran en la tabla 2.

Tabla 2: Valores del MSE de los modelos propuestos para la serie No Lineal simulada con Cambio de Nivel.

	Sin Cambio	Con Cambio
Modelo NN-AR	0.07821	0.05230
Modelo NN-Aditivo	0.07821	0.04871
Modelo NN-Mixto	0.07821	0.04851

Se observa que cuando se trabaja con modelos que no tienen en cuenta cambios estructurales, el MSE permanece constante, indicando que no es detectado por el modelo.

Con el tratamiento estructural, este cambio es detectado por los tres modelos propuestos, obteniéndose el menor MSE para el modelo mixto.

5. INTERPRETACION DEL MODELO

A partir de la realización del análisis residual asociado a la serie trabajada con el modelo óptimo, se obtienen las siguientes cuatro gráficas fundamentales, presentadas en la figura 8. Se destaca que en el histograma residual, las magnitudes de los ejes están expresados en forma adimensional, pues responden a procesos de estandarización.

A partir de los gráficos de la figura anterior, podemos extraer los siguientes resultados:

- La serie de los residuos tiene un comportamiento estacional, y fluctúa alrededor del cero de manera aleatoria, cumpliendo la hipótesis fundamental de los modelos estructurales de series de tiempo.
- Análoga conclusión puede extraerse de la gráfica de los cuadrados de residuos, en la que los picos guardan una alta concordancia con los meses del año correspondientes a los meses de mayor concentración de SO_2
- La función de autocorrelación muestra un comportamiento muy estable desde el punto de vista de la regularidad estadística, y corrobora que el modelado obtenido es altamente confiable para poder realizar la generación de valores sintéticos de la variable bajo estudio.
- En cuanto a la última gráfica, obsérvese el excelente comportamiento de la distribución asociada a los datos simulados con el modelo estructural presentado en este trabajo (curva en trazo más oscuro), el cual junto a los resultados arrojados por el estadístico BS, presenta un comportamiento cuasi-normal, que facilita el proceso de estimación, y de reducción de la varianza asociada al ruido blanco.

6. CONCLUSIONES

Mediante la introducción de perceptrones multicapa, se logró modelar una serie de promedios mensuales de la variable concentración de dióxido de azufre (SO_2) no lineal con cambios de nivel, en particular sobre su componente AR(1).

El modelado es óptimo cuando se toma una red neuronal mixta, obteniendo la serie que permite generar valores de tipo sintético ó pronosticados a partir del modelo final de la variable bajo estudio, con una confiabilidad muy alta (95%)

AGRADECIMIENTOS

Este trabajo se realizó dentro del marco del Proyecto N° 1457 finalizado, y el N° 1797, del Consejo de Investigación de la Universidad Nacional de Salta (C.I.U.N.Sa). Se agradece al Sr. Carlos Moya, Jefe del Servicio de Meteorología del Aeropuerto Internacional "Martín Miguel de Güemes", de la ciudad de Salta, República Argentina.

REFERENCIAS

- Abril, Juan Carlos. (1999). Análisis Estadístico de Series de Tiempo Basado en Modelos de Espacio de Estado. E.U.D.E.B.A.
- Aguiar R. y Collares Pereira M. (1992). Tag : A time dependent, autorregressive, gaussian model for generating synthetic hourly radiation. *Solar Energy*, **49**, 3, 167-174.

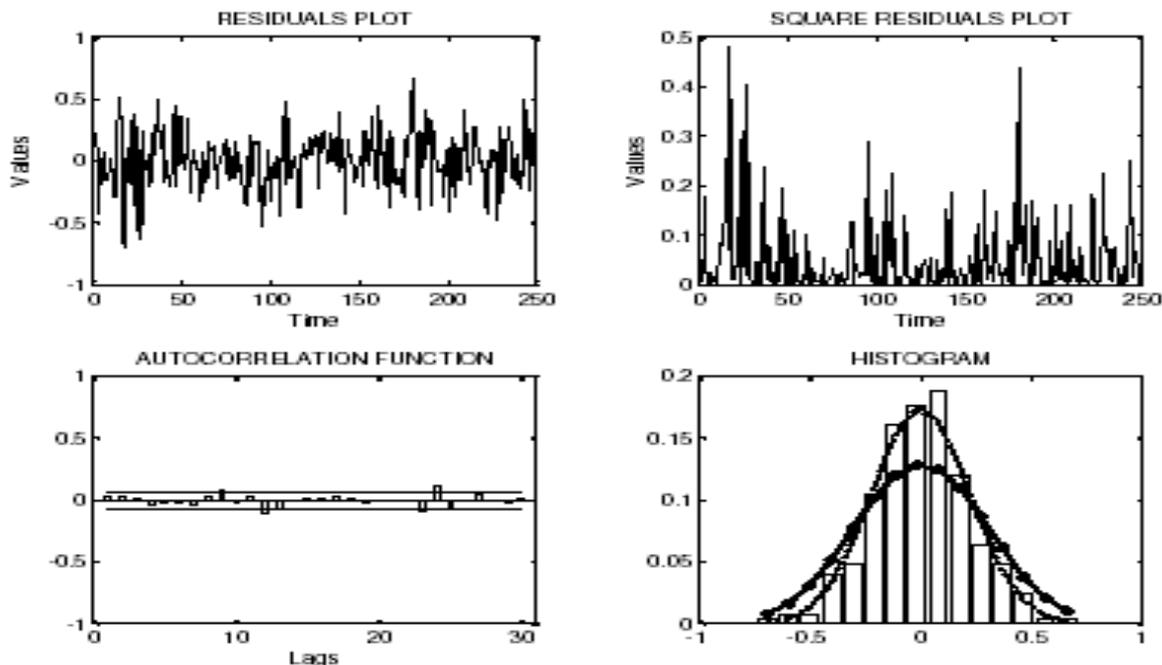


Fig. 8: Gráficos más representativos obtenidos luego del análisis residual

- Avila Blas O.J., Musso H., Avila G. y Farfán R. (2003). Evaluación de Concentración de dióxido de nitrógeno en Salta Capital: Un análisis estadístico estructural. *Avances en Energías Renovables y Medio Ambiente* Vol. 7, Nº 1, pp. 01.17 a 01.22.
- Avila Blas O.J., Musso H. y Avila G. (2004). Concentración de dióxido de nitrógeno en Salta Capital: Análisis estadístico para validar modelos estructurales previos. *Avances en Energías Renovables y Medio Ambiente*, Vol. 8, Nº 2, 2004, Tomo II, pp. 01.05 a 05.06.
- Avila Blas O. J. y Grossi Gallegos H. (2002). Modelos estadísticos estructurales de series de irradiación solar global diaria para Córdoba, Marcos Juárez y Paraná. *Avances en Energías Renovables y Medio Ambiente* Vol. 6, Nº 2, pp. 11.07 a 11.11.
- Avila Blas O.J., Abril J. C. y Lesino Garrido G. (2000) Radiación y temperatura diarias: Un modelo de Correlación Estructural. *Avances en Energías Renovables y Medio Ambiente*, Vol. 4, Nº 2, pp. 11.31 a 11.36.
- Avila Blas, J.O. (1997). Análisis Espectral de Series de Temperatura de Superficie. *Revista FACENA, Univ. Nac. Nordeste*, **13**, 79-99.
- Avila Blas, O. y Musso, H.E. (2008). *Independencia Estadística de Sitios de Muestreo de SO₂ en el Centro de la Ciudad de Salta*. XXVII Congreso Argentino de Química. Tucumán, del 17 al 19 de septiembre.
- Avila Blas, O.J. (2001). Análisis Estadístico de Series Climatológicas para su uso en simulación de edificios solares. *Tesis Doctoral en Ciencias*, Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de Salta. Defendida y aprobada el 07/12/01. ISBN Nº 987-9381-51-3.
- Avila Blas, O.J. y Lesino Garrido G. (1987). Tratamiento estadístico de datos meteorológicos del NOA para su uso en simulación. *Actas ASADES*, Buenos Aires.
- Avila Blas, O.J., Abril J.C. y Lesino Garrido G. (1999). Análisis estadístico estructural de Series de radiación diaria, *Avances en Energías Renovables y Medio Ambiente*, Vol. 3, Nº 2, pp. 11.17 a 11.20.
- Avila, G. N., Avila Blas, O. J. y Musso, H.E. (2007).. Modelado estructural de una serie de valores medios de ozono en Salta Capital. II Congreso Iberoamericano, IV Congreso Argentino de Química Analítica. 27 al 30 de Agosto. Buenos Aires.
- Gair A.J., Penkett S.A. and Oyolola, P., (1991). Development of a simple passive technique for the determination of nitrogen dioxide in remote continental locations. *Atmospheric Environment* 25A, 9: 1927-1939.
- Harvey A. C. y Shepard N. (1993). Estructural Time Series Models. *Handbook of Statistics*, Elsevier Science Publishers B.V (Editores), **11**, 261-302.
- Koopman S. J. (1995). STAMP 5.0, *Structural Time Series Analyser, Modeller and Predictor*.
- OMS, 2000. *Guidelines for Air Quality*, (WHO/SDE/OEH/00.02)
- OPS/CEPIS/PUB/04.110. GUÍAS PARA LA CALIDAD DEL AIRE. (2004). Traducción realizada por el Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente (CEPIS/OPS), agencia especializada de la Organización Panamericana de la Salud (OPS/OMS). Lima, Perú.
- Palmes E.D. and Tomczyk C., 1979. *Personal sampler for NO_x*. American Industrial Hygiene Association Journal 40: 588-591.
- Palmes E.D., Gunnison, A.F., Di Mattio J. and Tomczyk C., 1976. Personal samplers for nitrogen dioxide. *American Industrial Hygiene Association Journal* 37:570-577.
- Passive and Active Sampling Methodologies for Measurement of Air Quality. Nairobi. UNEP - WHO (ed.), 1994. GEMS/AIR Methodology Review Handbook Series. Vol. 4