

# Estado del arte en predicción de series temporales

J.J. Asensio, P.A. Castillo

Departamento de Arquitectura y Tecnología de Computadores  
ETSIIT, CITIC-UGR  
Universidad de Granada, España  
{asensio, pacv}@ugr.es

**Resumen** bonito abstract test

## 1. Introducción

Una serie temporal es una secuencia de observaciones dispuestas a intervalos regulares de aparición. Existen multitud de ejemplos donde se registran datos de esta forma, por citar algunos: precios de acciones en bolsa, temperatura mínima y máxima de cada día, accidentes domésticos... etc.

Aquí se destacan los trabajos de investigación publicados en revistas patrocinadas por el Instituto Internacional de Predictores (IIF), aunque también se cubren publicaciones clave de otras revistas. Se ofrece así una guía selectiva para la literatura en predicción de series temporales, cubriendo el periodo 1982-2005 y resumiendo cerca de 940 artículos que incluyen aproximadamente 340 artículos del IIF. También se revisan artículos y libros clave publicados en otros sitios que han sido altamente influyentes para varios desarrollos en este campo. Los trabajos referidos comprenden 380 artículos de revista y 20 libros y monografías.

Los artículos se han clasificado primero de acuerdo a los modelos introducidos en la literatura de series temporales, en lugar de hacerlo respecto a los métodos. Por ejemplo, los métodos bayesianos en general pueden aplicarse a todos los modelos. Los artículos que no conciernen con un modelo particular fueron entonces clasificados atendiendo a los diferentes problemas que abordan (medidas de precisión, combinación).

Puede haber artículos que no aparezcan o que sean citados por otros autores, aunque esto no significa que no sean importantes. Esta revisión es un breve recorrido histórico de los principales desarrollos.

## 2. Alisado exponencial

### 2.1. Preámbulo

En 1981, los métodos de alisado exponencial a menudo eran considerados como un conjunto de técnicas especiales para extrapolar varios tipos de series

temporales univariantes. Si bien los métodos de alisado exponencial eran ampliamente utilizados en la industria y los negocios, habían recibido poca atención por parte de los estadísticos y no tenían un fundamento estadístico bien desarrollado. Estos métodos se originaron en las décadas 50 y 60 con el trabajo de Brown (1959, 1963) [?][?], Holt (1957, reimpreso en 2004)[?], y Winters (1960) [?]. Pegels (1969) [?] proporcionó una clasificación sencilla pero útil de la tendencia y los patrones estacionales dependiendo de si eran aditivos (lineales) o multiplicativos (no lineales).

Muth (1960) [?] fue el primero en sugerir una base estadística para el alisado exponencial simple (SES) demostrando que proporcionaba la predicción óptima para el camino aleatorio con ruido. Los siguientes pasos para poner en un marco estadístico el alisado exponencial fueron proporcionados por Box y Jenkins (1970), Roberts (1982)[?], y Abraham y Ledolter (1983, 1986) [?][?], quienes mostraron que algunas predicciones de alisado exponencial lineales resultan como casos especiales de los modelos ARIMA. Sin embargo, estos resultados no se extienden a cualquier método de alisado exponencial no lineal.

Los métodos de alisado exponencial recibieron un impulso debido a dos artículos publicados en 1985, los cuales establecieron la base para gran parte del trabajo posterior en el área. En primer lugar, Gardner (1985) [?] realizó una revisión exhaustiva y una síntesis del trabajo en alisado exponencial hasta la fecha y extendió la clasificación de Pegels para incluir tendencia amortiguada. Este artículo reunió gran cantidad del trabajo existente lo que estimuló el uso de estos métodos y promovió más trabajos de investigación al respecto. Posteriormente ese mismo año, Snyder (1985) [?] demostró que SES podría ser considerado como una forma resultante de una innovación del modelo de espacio de estados (modelo con una única fuente de error). Aunque esta revelación fue en gran parte inadvertida entonces, en años recientes ha proporcionado la base para una gran cantidad de trabajos en modelos de espacio de estados en los que subyacen los métodos de alisado exponencial.

La mayor parte del trabajo desde 1980 ha incluido el estudio de las propiedades empíricas de los métodos (Bartolomei & Sweet, 1989 [?]; Makridakis & Hibon, 1991 [?]), propuestas para nuevos métodos de estimación o inicialización (McClain, 1988 [?]; Sweet & Wilson, 1988 [?]), o ha estado relacionado con modelos estadísticos en los que se considera que subyacen los métodos (McKenzie, 1984 [?]).

Los métodos de alisado multiplicativo de Taylor (2003)[?] proporcionan los únicos métodos de alisado exponencial genuinamente nuevos. Durante este periodo. Por supuesto, hay habido numerosos estudios que aplican métodos de alisado exponencial en varios contextos incluyendo componentes de computadores (Gardner, 1993 [?]), pasajeros de vuelo (Grubb & Masa, 2001 [?]), y planificación de producción (Miller & Liberatore, 1993 [?]).

La taxonomía de Hyndman, Koehler, Snyder, y Grose (2002 [?] extendida por Taylor, 2003 [?]) proporciona una categorización útil para describir la variedad de métodos. Cada método se componen de una de cinco tipos de tendencia (ninguna, aditiva, aditiva alisada, multiplicativa y multiplicativa alisada) y uno

de tres tipos de estacionalidad (ninguna, aditiva y multiplicativa). Por tanto, hay 15 métodos diferentes, donde entre los más conocidos está SES (sin tendencia ni estacionalidad), el método lineal de Holt (tendencia aditiva, sin estacionalidad), el método aditivo de Holt-Winters (tendencia aditiva y estacionalidad aditiva), y el método no lineal de Holt-Winters (tendencia aditiva y estacionalidad multiplicativa).

## 2.2. Variantes

Se han propuesto numerosas variaciones de los métodos originales. Por ejemplo, Carreno y Madinaveitia (1990) [?] y Williams y Miller (1999) [?] propusieron modificaciones para tratar discontinuidades, y Rosas y Guerrero (1994) [?] consideraron predicciones con alisado exponencial sujetas a una o más restricciones. También hay variantes por cómo y cuándo las componentes estacionales deberían ser normalizadas. Lawton (1998) [?] abogó por una renormalización de los índices estacionales para cada periodo de tiempo, puesto que elimina sesgo en las estimaciones de los componentes estacionales y de nivel. Roberts (1982) [?] y McKenzie (1986) [?] dieron esquemas ligeramente diferentes de normalización. Archibald y Koehler (2003) [?] desarrollaron nuevas ecuaciones de normalización que eran más sencillas de usar dando el mismo grado de predicción que los métodos originales. Una variante útil a mitad de camino entre SES y el método de Holt, es SES con desviación. Es equivalente al método de Holt con el parámetro de tendencia a cero. Hyndman y Billah (2003) [?] mostraron que este método también era equivalente al método Theta de Assimakopoulos y Nikolopoulos (2000) [?], cuando el parámetro de desviación es igual a la mitad de la pendiente de una tendencia lineal ajustada a los datos. El método Theta dio un rendimiento extremadamente bueno en la competición M3, aunque el por qué esta particular elección de modelo y parámetros era buena todavía no se ha determinado. Ha habido extraordinariamente poco trabajo en desarrollar versiones multivariadas de los métodos de alisado exponencial para predicción. Una excepción notable es Pfeiffermann y Allon (1989) [?] quienes consideraron datos de turismo israelí. El SES Multivariado se usa para gráficos de control de procesos (Pan, 2005) [?], donde es llamado “medias móviles multivariadas ponderadas exponencialmente”, pero aquí la cuestión no es predicción.

## 2.3. Modelos de espacio de estados

Ord, Koehler y Synder (1997) [?] a partir del trabajo de Snyder (1985) [?] propusieron una clase de modelos de espacio de estado de innovaciones en la que se puede considerar que subyacen algunos métodos de alisado exponencial. Hyndman et al (2002) [?] y Taylor (2003) [?] amplían esto para incluir los 15 métodos de alisado exponencial. De hecho, Hyndman et al (2002) [?] propusieron dos modelos de espacio de estado para cada método, correspondientes a los casos de error aditivo y multiplicativo. Estos modelos no son los únicos, y otros modelos relacionados para los métodos de alisado exponencial se presentaron en Koehler, Snyder y Ord (2001) [?] y Chatfield, Koehler, Ord y Snyder (2001) [?]. Desde hace

tiempo se conoce que algunos modelos ARIMA dan predicciones equivalentes a los métodos lineales de alisado exponencial. La importancia del trabajo reciente en modelos de espacio de estado de innovaciones es que los métodos de alisado exponencial no lineales también pueden derivarse de modelos estadísticos.

#### 2.4. Selección del modelo

Gardner y McKenzie (1988) [?] proporcionaron algunas reglas sencillas basadas en las varianzas de series temporales diferenciadas para elegir un método de alisado exponencial apropiado. Tashman y Kruk (1996) [?] compararon estos resultados con otros propuestos por Collopy y Armstrong (1992) [?] y un enfoque basado en el criterio de información bayesiano (BIC). Hyndman et al. (2002) [?] también propuso un enfoque basado en criterio de información, pero usando los modelos de espacio de estado subyacentes.

#### 2.5. Robustez

El buen rendimiento de la predicción de los métodos de alisado exponencial ha sido estudiado por varios autores. Satchell y Timmermann (1995) [?] y Chatfield et al. (2001) [?] mostraron que SES es óptimo para un rango amplio de procesos generadores de datos. En un pequeño estudio de simulación, Hyndman (2001) mostró que SES rendía mejor los modelos ARIMA de primer orden porque no está tan sujeto a problemas de selección de modelo, especialmente cuando los datos no siguen una distribución normal.

#### 2.6. Intervalos de predicción

Una de las críticas a los métodos de alisado exponencial a principios de los 80 era que no había forma de producir intervalos de predicción para las predicciones. La primera aproximación analítica a este problema fue asumir que las series eran generadas mediante funciones deterministas del tiempo más ruido blanco (Brown, 1963 [?]; Gardner, 1985 [?]; McKenzie, 1986 [?]; Sweet, 1985 [?]). En tal caso, sería mejor usar un modelo de regresión en lugar de los métodos de alisado exponencial. Newbold y Bos (1989) [?] criticaron fuertemente cualquier enfoque basado en esta hipótesis. Otros autores buscaron intervalos de predicción mediante la equivalencia entre los métodos de alisado exponencial y los modelos estadísticos. Johnston y Harrison (1986) [?] establecieron variantes de predicción para los métodos SES, y los métodos de Holt de alisado exponencial para modelos de espacio de estado con múltiples fuentes de error. Yar y Chatfield (1990) [?] obtuvieron intervalos de predicción para el método aditivo de Holt-Winters derivando el equivalente método ARIMA subyacente. Chatfield y Yar (1991) [?] discutieron intervalos aproximados de predicción para el método multiplicativo de Holt-Winters, asumiendo que los errores en las predicciones a un paso fueran independientes. Koehler et al. (2001) [?] también derivó una fórmula aproximada para la previsión de varianza para el método multiplicativo de Holt-Winters, distinguiéndose de Chatfield y Yar (1991) [?] solamente en cómo la

desviación estándar del error en la predicción a un paso era estimada. Ord et al (1997) [?] y Hyndman et al. (2002) [?] usaron el modelo de espacio de innovación subyacente para simular rutas de muestras futuras, y de este modo obtuvieron intervalos de predicción para todos los métodos de alisado exponencial. Hyndman, Koehler, Ord y Snyder (2005) [?] usaron modelos de espacio de estado para derivar intervalos de predicción analíticos para 15 de los 30 métodos, incluyendo todos los métodos usados habitualmente. Proporcionaron el enfoque algebraico más comprensible hasta la fecha para manejar el problema de la distribución de predicción para la mayoría de los métodos de alisado exponencial.

## 2.7. Espacio de parámetros y propiedades del modelo

Una práctica habitual es restringir los parámetros de alisado al rango de 0 a 1. Sin embargo, ahora que los modelos estadísticos subyacentes están disponibles, el espacio natural (invertible) de parámetros se puede usar en su lugar. Archibald (1990) [?] mostró que es posible para parámetros de alisado dentro de los intervalos usuales producir modelos no invertibles. Consecuentemente, al predecir, el impacto del cambio en los valores pasados de la serie no es despreciable. Intuitivamente, tales parámetros producen predicciones pobres y deteriora el rendimiento de la predicción. Lawton (1998) [?] también discutió este problema.

## 3. Modelos ARIMA

### 3.1. Preámbulo

Los primeros intentos de estudiar series temporales, particularmente en el siglo 19, estaban generalmente caracterizados por la idea de un mundo determinista. Fue la mayor contribución de Yule (1927) la que impulsó la noción de aleatoriedad en series temporales postulando que cada serie temporal puede ser considerada como la realización de un proceso estocástico. Basado en esa sencilla idea, se desarrolló desde entonces un gran número de métodos de series temporales. Slutsky, Walker, Yaglom y Yule formularon el concepto de los modelos autorregresivos (AR) y media móvil (MA). El teorema de la descomposición de Wold condujo a la formulación y solución del problema de predicción lineal de Kolmogorov (1941). Desde entonces, una gran cantidad de literatura ha aparecido en el área de series temporales, tratando estimación de parámetros, identificación, comprobación de modelos y predicción, véase Newbold (1983) para un estudio anterior.

La publicación *Time Series Analysis: Forecasting and Control* de Box y Jenkins (1970) integró el conocimiento existente. Además, estos autores desarrollaron una ciclo coherente, versátil e iterativo en tres etapas para la identificación, estimación y verificación de series temporales (conocido como el enfoque Box-Jenkins). El libro ha tenido un enorme impacto en la teoría y práctica moderna de análisis y predicción de series temporales. Con la llegada de los ordenadores, se popularizó el uso de los modelos ARIMA y se extendió a muchas áreas de la

ciencia. De hecho, la predicción de series temporales discretas mediante modelos ARIMA univariantes, modelos de función de transferencia (regresión dinámica), y modelos ARIMA multivariados han generado bastantes artículos en IJF. A menudo estudios de naturaleza empírica, que usan uno o varios métodos/modelos como punto de referencia para comparar.

(incluir tabla)

### 3.2. Univariado

El éxito de la metodología de Box-Jenkins se basa en el hecho de que los modelos pueden imitar el comportamiento de los diversos tipos de series y sin requerir muchos parámetros a estimar en la elección final del modelo. Sin embargo, a mediados de los años sesenta, la selección de un modelo era en gran medida una cuestión subjetiva del investigador, no había ningún algoritmo para especificar un modelo único. Desde entonces, muchas técnicas y métodos han surgido para añadir rigor matemático en el proceso de búsqueda de un modelo ARMA, incluyendo el criterio de información Akaike (AIC), el error de predicción final de Akaike (FPE) y el criterio de información bayesiano (BIC). A menudo, estos criterios se reducen a la minimización de errores de predicción a un paso (en la muestra), con un término de penalización para el sobreajuste. FPE también ha sido generalizado para predicción a varios pasos (véase, por ejemplo, Bhansali, 1996 [?] y Bhansali, 1999 [?]), pero esta generalización no ha sido utilizada en el área aplicada. Este también parece ser el caso de criterios basados en principios de validación cruzada y validación por división muestral (véase, por ejemplo, West, 1996), haciendo uso de errores de predicción fuera-de-muestra; véase Peña y Sánchez (2005) para un enfoque relacionado a considerar.

Existen diferentes métodos para estimar los parámetros de un modelo ARMA (véase Box et al., 1994 [?]). Aunque estos métodos son asintóticamente equivalentes, en el sentido de que las estimaciones tienden a la misma distribución normal, hay grandes diferencias en las propiedades con muestra finitas. En un estudio comparativo de los paquetes de software, Newbold, Agiakloglou y Miller (1994) [cita] mostraron que esta diferencia puede ser muy sustancial y, como consecuencia, puede influir en las predicciones. En consecuencia, recomendaron el uso de máxima verosimilitud completa. El efecto de que los errores de estimación para los parámetros limitan las predicciones también fue algo observado por Zellner (1971). Él utilizó un análisis bayesiano y derivó la distribución predicha de observaciones futuras tratando los parámetros del modelo ARIMA como variables aleatorias. Más recientemente, Kim (2003) [cita] consideró una predicción y estimación de los parámetros de modelos AR con pequeñas muestras. Encontró que las estimaciones de parámetros con sesgo corregido producían predicciones más precisas que las de mínimos cuadrados. Landsman y Damodaran (1989) [cita] presentaron la evidencia de que la estimación de parámetros de ARIMA James-Stein mejoraba la precisión de predicción con respecto a otros métodos, bajo el criterio de MSE.

Si una serie temporal se sabe que sigue un modelo ARIMA univariado, las predicciones que usan observaciones desagregadas son, en términos de MSE, al

menos tan buenas como aquellas que usan observaciones agregadas. Sin embargo, en aplicaciones prácticas, existen otros factores a considerar, como valores perdidos en series desagregadas. Ledolter (1989) [cita] y Hotta (1993) analizaron el efecto producido de un outlier aditivo en el intervalo de predicción cuando se estiman los parámetros del modelo ARIMA. Cuando el modelo es estacionario, Hotta y Cardoso Neto (1993) mostraron que la pérdida de eficiencia usando datos desagregados no era alta, incluso si el modelo no es conocido. Por tanto, la predicción podría hacerse tanto con modelos desagregados como agregados.

El problema de añadir información externa (prior) en las predicciones de ARIMA univariado fue considerado por Cholette (1982) [?], Guerrero (1991) [?] y de Alba (1993) [?].

Como alternativa a la metodología de ARIMA univariado, Parzen (1982) [cita] propuso la metodología ARARMA. La idea clave es que una serie temporal es transformada de un filtro AR de memoria larga, a un filtro de memoria corta, evitando así un operador de diferenciación más duro. Además, se usa una fase de identificación diferente al convencional de Box-Jenkins. En la competición M (Makridakis et al., 1982 [cita]), los modelos ARARMA consiguieron el MAPE más bajo para horizontes de predicción mayores. Por lo tanto, es sorprendente encontrar que, a parte de artículo de Meade y Smith (1985) [cita], la metodología ARARMA todavía no haya despegado en el área aplicada. Su valor definitivo, quizá pueda ser juzgado mediante el estudio de Meade (2000), quien comparó la capacidad de predicción de un método ARARMA automático y no automático.

El modelado automático de ARIMA univariado, ha mostrado una capacidad de predicción a un paso tan precisa como la de otros métodos competentes (Hill y Fildes, 1984, Liber, 1984, Poulos et al., 1987 y Texter y Ord, 1989). Varios proveedores de software han implementado métodos de predicción automáticos de series temporales (incluso métodos multivariados); véase Geriner y Ord (1991) [?], Tashman y Leach (1991) [cita], y Tashman (2000) [cita]. A menudo estos métodos actúan como caja negra. La tecnología de sistemas expertos (Mélard & Pasteels, 2000) se puede usar para evitar este problema. Algunas directrices en la elección de un método de predicción automático se da en Chatfield (1988) [?].

Mejor que adoptar un modelo AR para todos los horizontes de predicción, Kang (2003) investigó empíricamente el caso de seleccionar un modelo AR de forma diferenciada para cada horizonte en predicciones multi-paso. El rendimiento de predicción del procedimiento multi-paso parece depender entre otras cosas del criterio de selección de orden óptimo, los periodos de predicción, los horizontes de predicción y la serie a predecir.

### 3.3. Función de transferencia

La identificación de modelos de función de transferencia puede ser difícil cuando hay más de una variable de entrada. Edlund (1984) [?] presentó un método en dos pasos para identificación de una función respuesta impulso cuando varias variables de entrada están correladas. Koreisha (1983) [cita] estableció varias relaciones entre funciones de transferencia, implicaciones causales y especificación de modelo econométrico. Gupta (1987)