

## Pronóstico de las condiciones del ganado vacuno en cuba a través de modelos ARIMA

### Prognosis of the conditions of cattle livestock in Cuba through ARIMA models



<http://opn.to/a/2qjer>

Ismabel María Domínguez-Hurtado <sup>1\*</sup>, César Soto-Valero <sup>1</sup>

<sup>1</sup>Centro Meteorológico Provincial de Villa Clara, Villa Clara, Cuba

**RESUMEN:** El objetivo de este trabajo se centró en la modelación para predecir las condiciones de los animales mediante el uso de series temporales para Cuba. El análisis se basó en el índice de temperatura y humedad (ITH) calculado a partir de los datos de 64 estaciones meteorológicas del país, agrupadas en tres regiones: oriente, centro y occidente. Se aplicaron los test de *Dickey - Fuller* y *KPSS* para identificar la estacional de las series. Se empleó el método *Holt Winters* de la familia de series temporales de suavizado exponencial. Los mejores modelos resultaron los correspondientes a la zona oriental (1,1, 2) (0, 0,1) <sub>52</sub> y (0,0,1)\_(0,1,1) <sub>12</sub>. Además, se realizó el análisis para la estación meteorológica La Piedra, Manicaragua, Villa Clara.

**Palabras clave:** series temporales, ARIMA, ITH.

**ABSTRACT:** The objective of this work was focused on modeling to predict animal conditions through the use of time series for Cuba. The analysis was based on the temperature and humidity index (ITH) calculated from the data of 64 meteorological stations of the country, grouped into three regions: east, center and west. Dickey - Fuller and KPSS tests were applied to identify the seasonal series. Holt Winters method of the family of time series of exponential smoothing was used. The best models were those corresponding to the eastern zone (1,1, 2) (0,0,1) <sub>52</sub> and (0,0,1) (0,1,1) <sub>12</sub>. In addition, the analysis was carried out for a La Piedra weather station (Manicaragua, Villa Clara, Cuba).

**Key words:** temporal series, ARIMA, THI.

#### INTRODUCCIÓN

El pronóstico de condiciones agrometeorológicas, en sentido general, se refiere a las predicciones esperadas que influyen sobre la producción agropecuaria ([Caraza y Quintero, 1997](#)). La utilidad de los mismos se orienta en proveer una guía para la adopción de decisiones, las que en el área de la producción agropecuaria resultan de interés dada la dependencia de esta actividad con las condiciones de la atmósfera.

En la ganadería, se han utilizado las salidas de los modelos globales empleados en la predicción

del tiempo asociadas con indicadores como el índice de temperatura y humedad (THI) o rangos de temperatura del aire para la evaluación de las condiciones de los animales ([Domínguez et al., 2010](#)). Además de ello, muchas de las técnicas de predicción agrometeorológica se fundan en la relación estadística que existe entre las variables dependientes que han de ser estimadas y las variables agrometeorológicas independientes (precipitación, temperatura), o las variables deducidas (índice de humedad del suelo, influencia de la humedad atmosférica).

\*Autora para la correspondencia: *Ismabel María Domínguez-Hurtado*. E-mail: [ismabel.dominguez@vcl.insmet.cu](mailto:ismabel.dominguez@vcl.insmet.cu)

Recibido: 08/07/2019

Aceptado: 20/01/2020

La estadística inferencial provee varias herramientas para el análisis que resultan de interés para las ciencias meteorológicas. Los modelos autorregresivos integrados de promedios móviles o ARIMA (acrónimo del inglés *autoregressive integrated moving average*, también referido como método de *Box - Jenkins*), donde las estimaciones futuras se sustentan únicamente en los datos previos, han sido aplicados en Cuba para la modelación y pronóstico en diferentes ramas del saber. Por ejemplo, en meteorología se encuentran los trabajos de [Osés et al., \(2017\)](#) y Cabanas (2005), mientras que en la rama agropecuaria se hallan, por ejemplo, los trabajos de [Sánchez \(2013\)](#) y [Guevara et al., \(2004\)](#).

El objetivo general de este trabajo es predecir las condiciones de los animales mediante modelos ARIMA de series temporales para el escenario nacional.

## MATERIALES Y MÉTODOS

Se emplearon los datos de 64 estaciones meteorológicas pertenecientes a la red de estaciones de observación del Instituto de Meteorología. Igualmente se usaron los Boletines de Indicadores Fundamentales de la ganadería. De la primera, se utilizaron los datos de la empresa pecuaria La Vitrina.

A continuación se listan los indicadores seleccionados:

Producción de leche

- Vacas en ordeño (cabezas),
- Producción total de leche (L).

Reproductivos:

- Nacimientos totales (cabezas),

Relativos a la masa total:

- Mortalidad real (cabezas),
- Mortalidad en los terneros (cabezas)

Para la evaluación del estado de los animales se empleó el Índice de Temperatura y Humedad (ITH) mediante la [ecuación 1](#) ([Nienabar y Hahn, 2007](#); [Mujica, 2005 citado por Aguilar, 2006](#)) para un periodo de 21 años (1993-2013).

$$ITH = 0.8t_d + HR(t_d - 14.4) + 46.4 \quad (1)$$

Donde,  $t_d$  es la temperatura del bulbo seco (°C) y  $HR$  representa la humedad relativa expresada en forma decimal. En la [tabla 1](#) aparecen los intervalos adoptados para la valoración.

**Tabla 1.** Escala para la evaluación del ITH.

Valor del ITH		Evaluación
Inferior	Superior	
	74	Normal
75	78	Alerta
79	83	Peligro
84	-	Emergencia

Fuente: [Nienabar y Hahn \(2007\)](#) y [Mujica \(2005\)](#).

Se aplicaron técnicas exploratorias para la identificación de patrones y para ello, los datos fueron divididos y a partir de esto, se crearon series temporales como sigue:

- Valores diarios (cada año con 365 datos por estación).
- Valores semanales (cada año incluyó 52 datos por estación),
- Por valores mensuales (cada año incluyó 12 datos por estación),
- Valores anuales (un total de 21 puntos),
- Por estaciones (valores diarios, semanales, mensuales y anuales para cada una de las 23 estaciones).

Para el análisis y representación se crearon tres grupos: ITH1, que incluyó las estaciones de las provincias más orientales (desde Camagüey hasta Guantánamo), ITH2 (Matanzas hasta Ciego de Ávila) e ITH3 (Pinar del Río, Mayabeque, La Habana e Isla de la Juventud).

Se procedió al examen visual de los datos, mediante la representación gráfica con el propósito de identificar irregularidades en las series. Posteriormente se pasó a la descomposición de los mismos (tendencia, ciclicidad y estacionalidad). Se aplicaron los test de *Dickey - Fuller* y *KPSS* (*Kwiatkowski- Phillips - Schmidt - Shin*) para identificar la estacional de las series.

La prueba *Dickey - Fuller* se basó en asumir que la serie se puede aproximar por un proceso AR(1) con tres variantes: media cero, media diferente de cero y tendencia lineal. En este caso se partió de las hipótesis:

$$H_0: \rho = 0 \quad (2)$$

$$H_1: \rho < 0 \quad (3)$$

La hipótesis nula (2) asume que la serie no es estacionaria. El estadístico de la prueba se denota por  $p$  y su distribución bajo  $H_0$  permite calcular los valores críticos, de tal forma que el criterio de rechazo es  $p < p_{0.05}$ , con  $p$  es valor calculado del estadístico.

Se representó un diagrama estacional como se describe en [Hyndman y Athanasopoulos \(2013\)](#), es decir, semejante a un gráfico de tiempo excepto que los datos se trazaron contra las estaciones en años separados. La elección de un modelo ARIMA óptimo ( $p, d, q$ ) se efectuó mediante una función AUTOARIMA. Se consideró a  $p$  como el parámetro auto regresivo,  $d$  al número de fases de diferenciación no estacionales y  $q$  al parámetro de medias móviles.

Los pronósticos de objetos *STL* (por sus siglas en inglés, *Standard Triangle Language*) se obtuvieron aplicando un método de predicción no estacional a los datos ajustados estacionalmente y re estadificando utilizando el último año del componente estacional. Se empleó el método *Holt Winters* de la familia de series temporales de suavizado exponencial.

Para la evaluación e iteración, se plotearon los ACF y PAFC para los residuales de los modelos resultantes. Todos los análisis se efectuaron mediante la aplicación de la herramienta *R* de [Core Team, \(2018\)](#).

## RESULTADOS Y DISCUSIÓN

### Obtención de las series temporales para las diferentes escalas de trabajo

Se adicionaron los datos diarios, semanales y mensuales para conformar series de tiempo de la suma de valores de ITH. Para el ajuste de la serie de tiempo semanal, primeramente se truncaron los primeros y últimos registro de las series y se asignaron los nuevos valores de principio y fin.

### Chequeo de la estacionalidad de los datos

Para el ajuste a un modelo tipo ARIMA se requiere que las series sean estacionarias, es decir, que su media, varianza y autocovarianza sean invariantes en el tiempo. En la [Tabla 2](#) aparecen los resultados obtenidos para las pruebas seleccionadas.

Como se puede apreciar en cualquiera de los casos, se rechaza la hipótesis nula y todas las series son estacionarias ( $p < 0.05$ ). Lo que se corrobora con los resultados del test KPSS ([Tabla 3](#)).

La autocorrelación en las series es un indicador de cómo cada valor de la serie temporal. En este caso se empleó la función `acf()` para la detección de esta característica.

El panel en la [figura 1](#) muestra los autocorrelogramas, donde se muestran la

**Tabla 2.** Test de *Dickey - Fuller*

Periodicidad	Grupo	<i>Dickey- Fuller</i>	Retardo	p	Hipótesis alternativa
Diaria	I	-27.509	55	0.01	Estacionaria
	II	-12.249	48	0.01	Estacionaria
	III	-24.97	48	0.01	Estacionaria
Semanal	I	-8.6862	10	0.01	Estacionaria
	II	-8.7354	10	0.01	Estacionaria
	III	-8.7835	10	0.01	Estacionaria
Mensual	I	-5.4499	6	0.01	Estacionaria
	II	-8.6185	6	0.01	Estacionaria
	III	-5.1948	6	0.01	Estacionaria

**Tabla 3.** Test de KPSS para el nivel de estacionalidad.

Periodicidad	Grupo	KPSS	Parámetro de truncado del retardo	p
Diaria	I	1.3020	95	0.01
	II	1.4743	78	0.01
	III	1.8322	77	0.01
Semanal	I	2.119900	7	0.01
	II	0.076025	7	0.01
	III	4.392	7	0.01
Mensual	I	2.12620	3	0.01
	II	0.10048	3	0.01
	III	3.6182	3	0.01

correlación entre la serie y sus rezagos. Siempre se tiene que  $\rho(0)=1$ . Las líneas discontinuas en azul representan las bandas de confianza de  $\rho(k)$  de nivel 95% bajo la hipótesis de que la serie es un ruido blanco (incorrelada).

Se observa que dentro de las series de tiempo hay una fuerte correlación entre cada valor y el valor anterior. Esto significa que los valores de las series temporales no son independientes entre sí. También hay una correlación estadísticamente significativa para el segundo y tercer rezagos. En este caso, el retardo predeterminado es  $10\log_{10}\left(\frac{N}{m}\right)$  donde  $N$  es el número de observaciones y  $m$  el número de series.

Una forma simple de lidiar con la autocorrelación es la diferenciación. Diferenciar significa restar valores previos de los valores actuales. La diferenciación de orden 1 significa que se sustrae el valor anterior del valor actual. En este caso, todavía hay autocorrelaciones estadísticamente significativas.

Los gráficos mostrados en la [Figura 2](#) muestran las series temporales.

Una vez obtenidas las series temporales se confeccionaron los diagramas estacionales como se describe en [Hyndman y Athanasopoulos \(2018\)](#).

Los resultados de la aplicación del pronóstico automático mediante suavizado exponencial (método *Holt Winters*), tanto para la escala semanal como mensual, aparecen en el panel representado en la [Figura 3](#). Para el primer lapso la predicción se

realizó para 52 semanas (un año); mientras que en el segundo, se determinó usar 36 meses (tres años).

Para el pronóstico de la serie temporal se empleó la función `auto.arima()`, mediante la cual se determinaron los valores óptimos  $(p,d,q)$  ([Hyndman, 2008](#)).

En la [Tabla 4](#) aparecen los estadísticos obtenidos para cada uno de los modelos ARIMA y STL. Los estadísticos Raíz del Error Cuadrático Medio (RMSE), el Error Medio Absoluto (MAE) y el Error Medio Absoluto Porcentual (MAPE), miden la magnitud de los errores, o sea, el mejor será el que muestre el valor más pequeño; y el Error Medio (ME) y el Error Medio Porcentual (MPE) miden el sesgo estadísticamente y se considera que el mejor modelo da un valor próximo a 0 ([Sánchez, 2013](#)). Otro criterio lo dan [Sánchez y Poveda \(2006\)](#), para quienes el desempeño de los modelos se evalúa para todas las series mediante el error cuadrático medio (RMS). Sobre los correspondientes en este análisis, tanto para la escala semanal como mensual, se observa que el de mejor ajuste corresponde al grupo oriental:  $(1,1,2)$   $(0,0,1)_{52}$  y  $(0,0,1)(0,1,1)_{12}$ .

Debe considerarse que los modelos ARIMA, por lo general predicen bien para horizontes de tiempo ( $h$ ) fuera de la muestra cortos y medios. El desempeño de los pronósticos de distintos modelos puede variar según sea la amplitud de dicho horizonte. Cuando se pronostica una serie, el valor de  $h$  hasta que sería apropiado pronosticar, estará determinado por la incertidumbre con que se

pronostica para cada  $h$  (si la amplitud del intervalo de predicción es muy amplia, la información que se brinda será trivial o poco precisa) (Blaconá, 2018).

Para el caso de la aplicación de esta herramienta en el caso particular de una estación meteorológica se escogió la 78308, con una escala temporal diaria para el análisis. En la Figura 4 aparece la serie temporal obtenida al efecto.

Para la aplicación del pronóstico automático mediante suavizado exponencial (*Holt - Winters*), como los subsiguientes, requieren el análisis de estacional. La trama estacional de acuerdo con Hyndman y Athanasopoulos (2018) aparece en la Figura 5 a modo de gráfica de tiempo, excepto que los datos se representan en función de las estaciones en años separados.

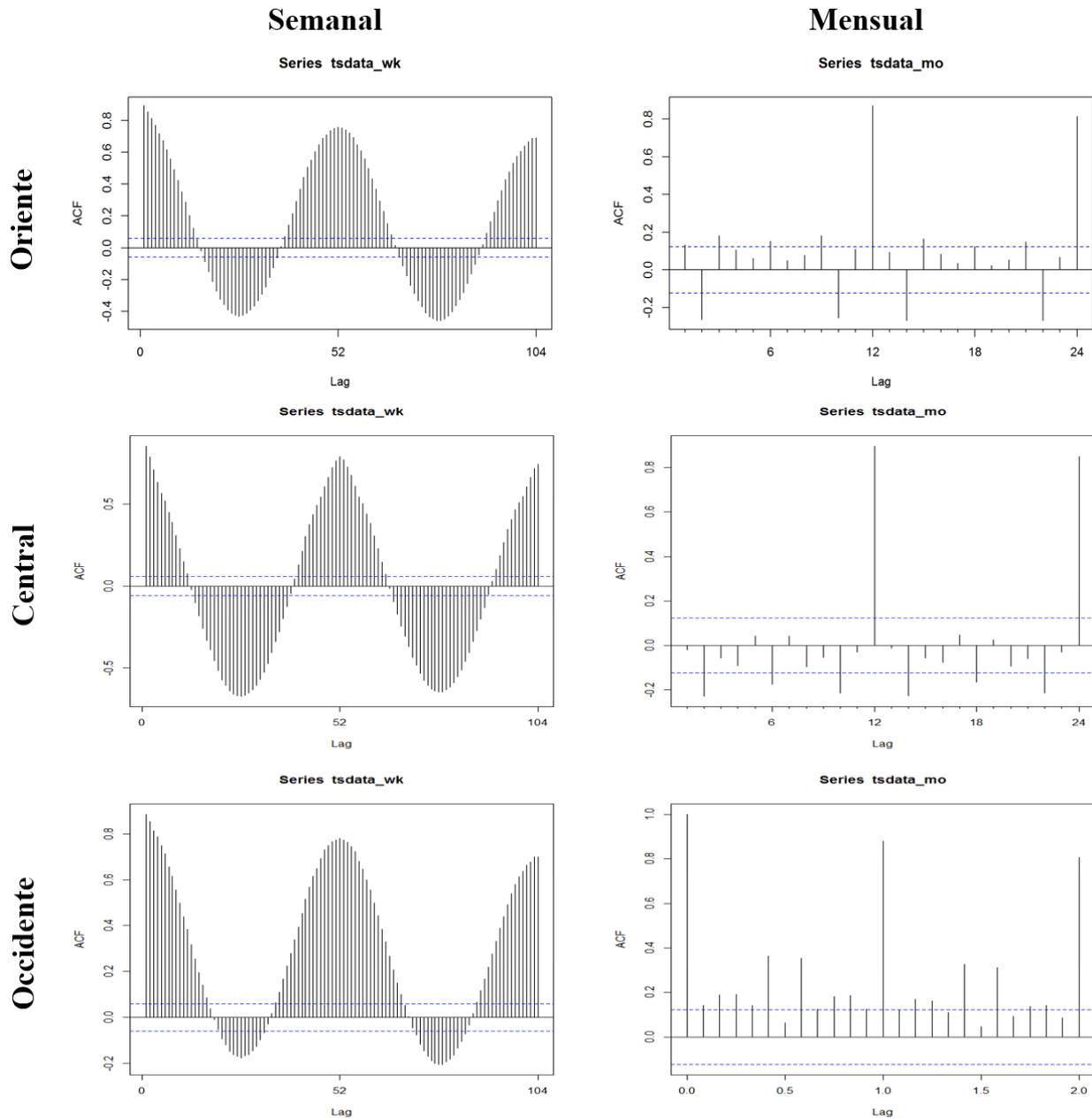


Figura 1. Autocorrelogramas para las series.

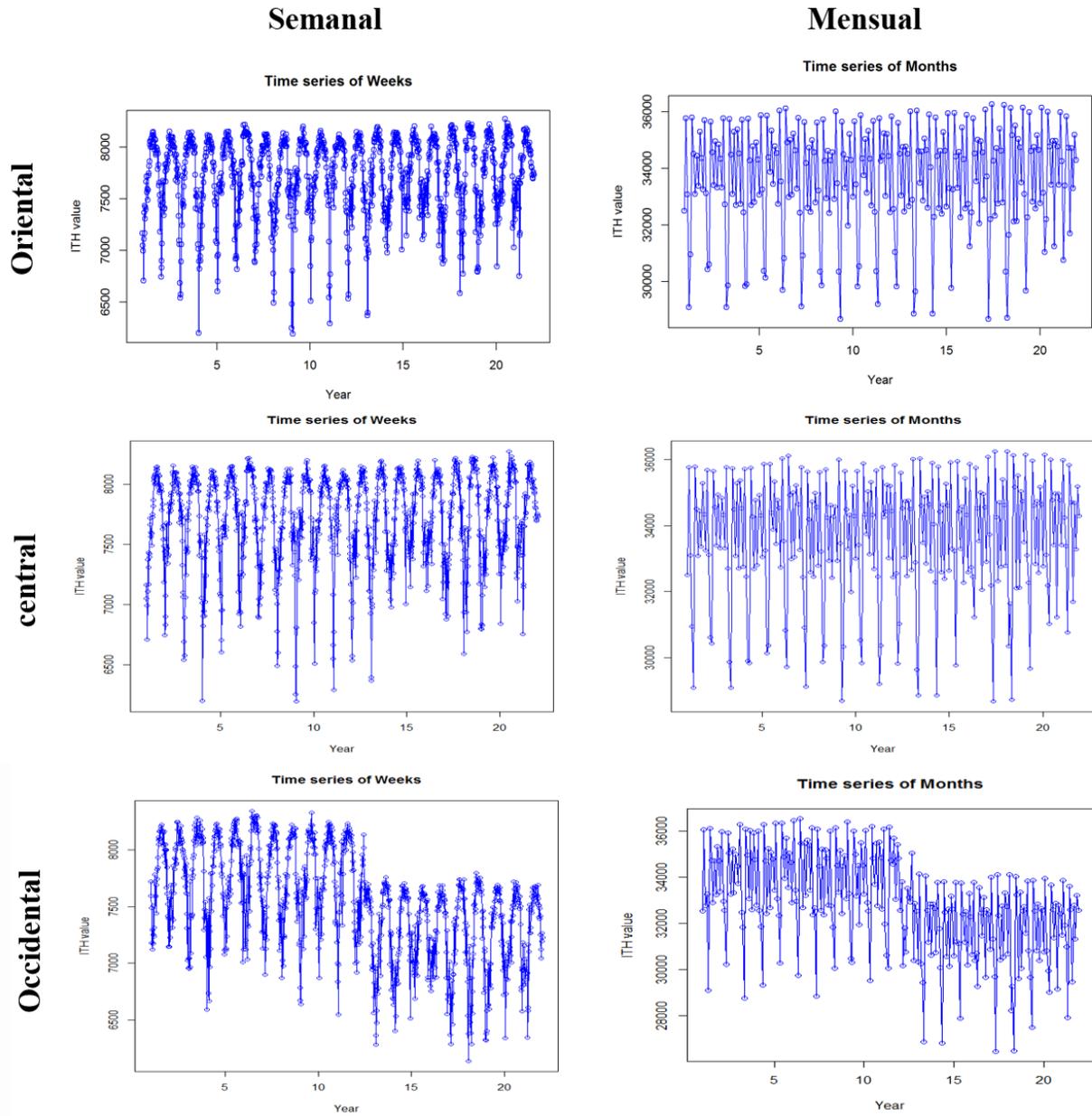
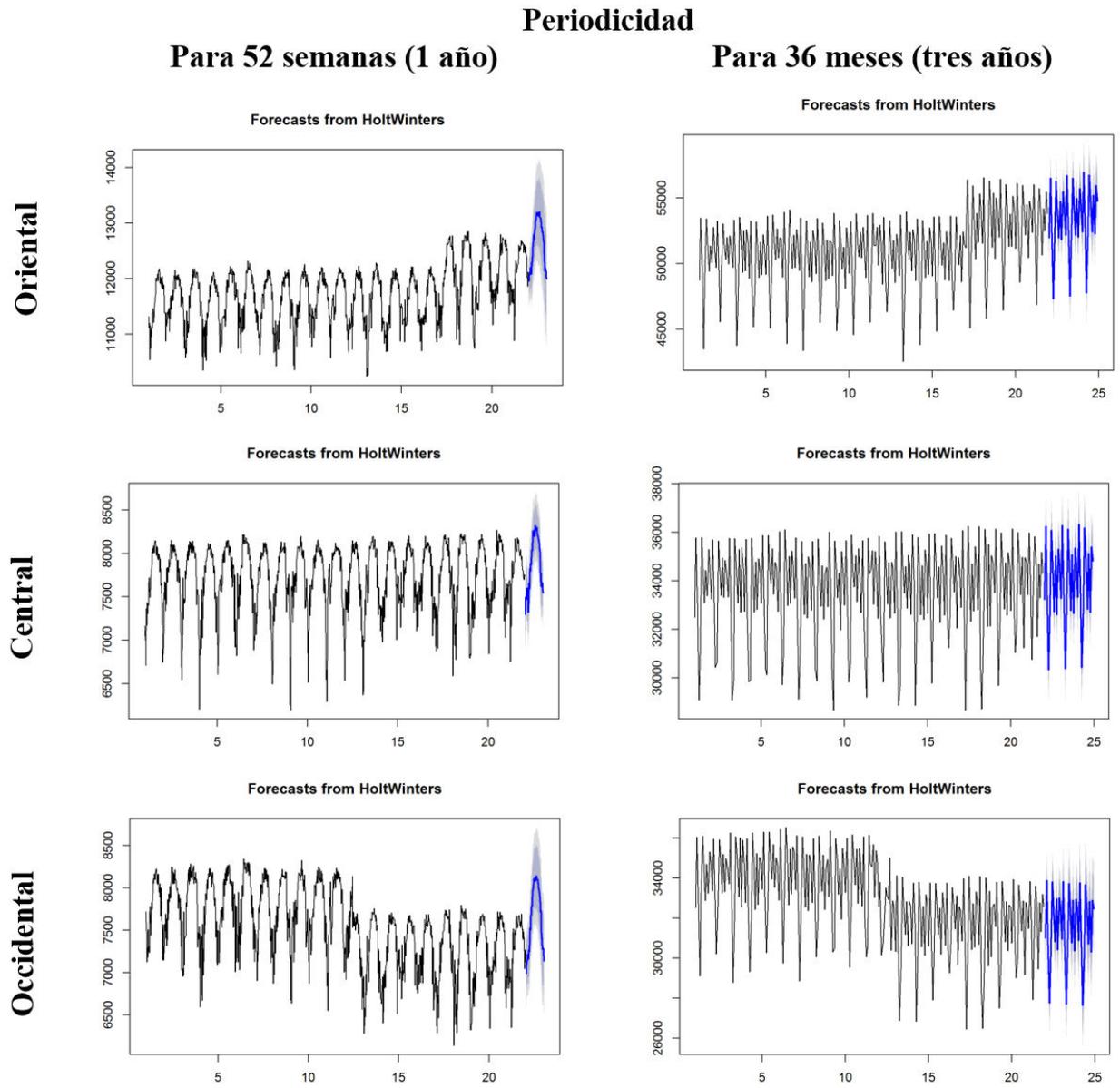


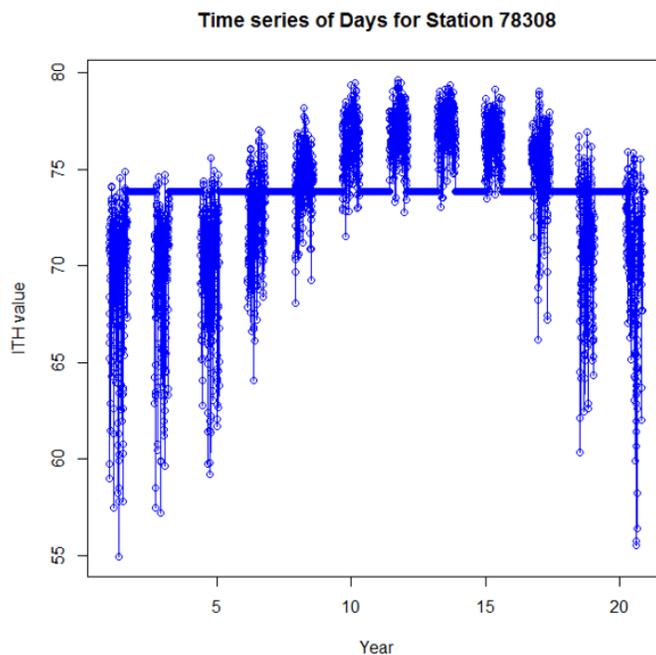
Figura 2. Series temporales para los valores medios del ITH.



**Figura 3.** Pronóstico automático mediante suavizado exponencial (método *Holt Winters*)

**Tabla 4.** Comparación de modelos probados para la serie índice de temperatura y humedad (ITH).

Escala	Modelo	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
<b>Grupo 1 (Oriental)</b>								
Semana	(1,1,2) (0,0,1) [52]	0.672346	219.5853	157.2042	-0.01666238	1.365004	0.6784801	0.01182631
	STL + ETS (M, N, N)	2.685848	165.9827	121.0931	0.002996591	1.052445	0.5226275	0.1330594
Mes	(0,0,1) (0,1,1) [12]	-14.16601	855.5038	632.3386	-0.05865995	1.257348	0.8519141	0.07934173
	STL + ETS (M, N, N)	49.05009	622.7427	450.1271	0.08150345	0.8971779	0.6064308	0.04596662
<b>Grupo 2 (Central)</b>								
Semana	(1,0,1) (1,0,0) [52]	0.9132933	184.7285	127.7953	-0.05741515	1.70961	0.795167	-0.007524508
	STL + ETS (M, N, N)	0.9999345	132.3185	93.09202	-0.01872396	1.242576	0.5792366	0.170529
Mes	(1,0,2) (0,0,2) [12]	-2.697734	947.1285	723.3742	-0.14271560	2.203124	1.583557	-0.0005898661
	STL + ETS (M, N, N)	26.45073	414.3554	285.7820	0.06279701	0.8780415	0.6256125	0.08074876
<b>Grupo 3 (Occidental)</b>								
Semana	(1,1,2) (0,0,2) [52]	0.04125073	196.7555	141.0974	-0.04851988	1.921384	0.7013239	0.02477567
	STL + ETS (M, N, N)	-1.566856	148.7209	108.583	-0.0606489	1.481001	0.5397114	0.1550966
Mes	(1,0,0) (1,1,0) [12]	2.935849	782.174	566.6811	-0.02771096	1.785012	0.9263086	-0.02684771
	STL + ETS (M, N, N)	-35.08419	523.9949	374.2767	-0.1320103	1.175063	0.6118005	0.03096434



**Figura 4.** Serie temporal (escala diaria). Estación 78308.

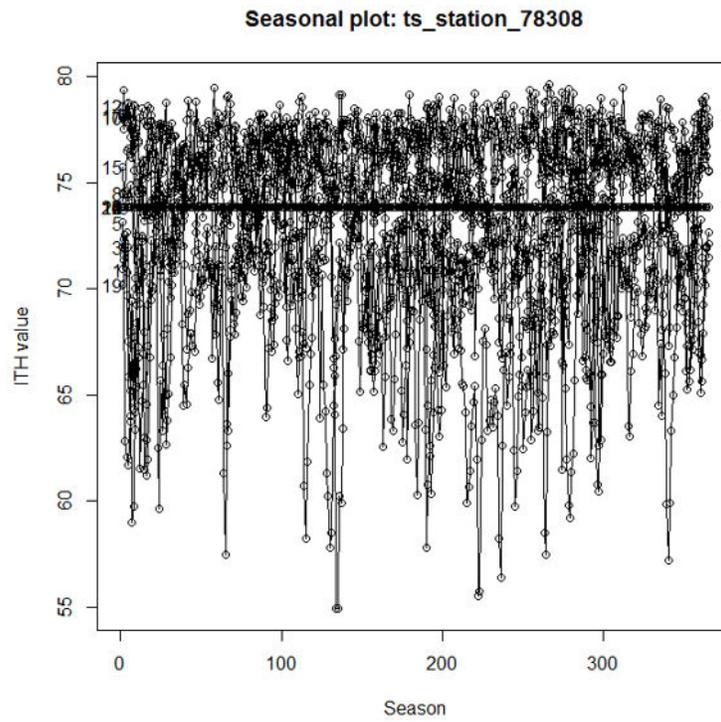


Figura 5. Seasonal plot. Estación 78308.

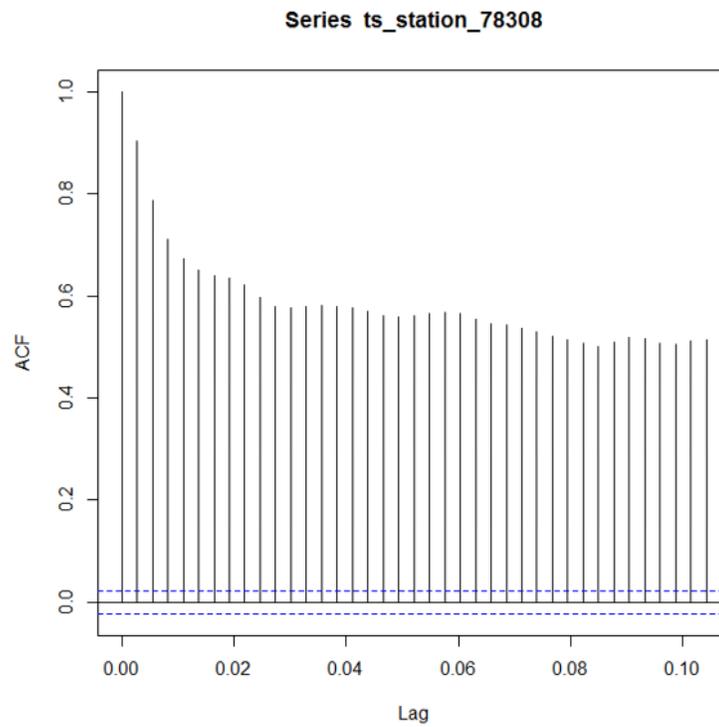
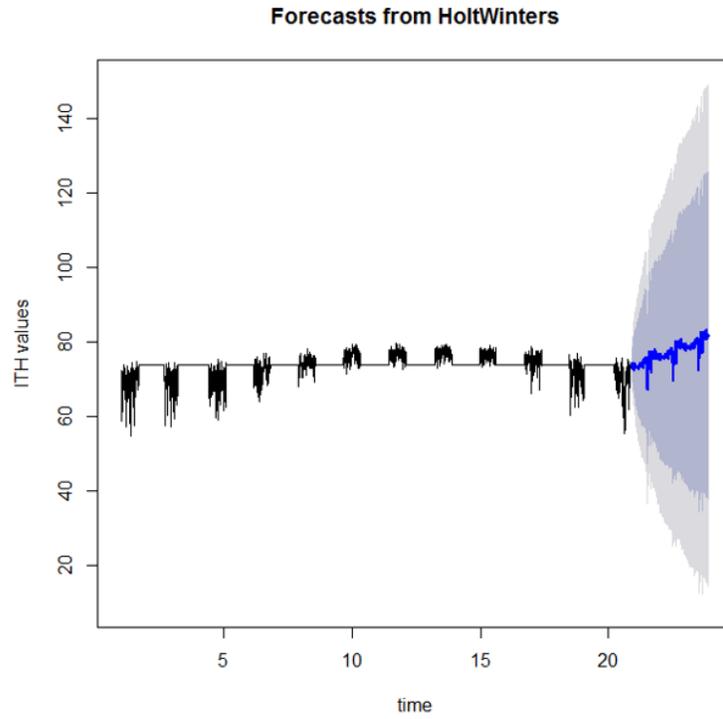
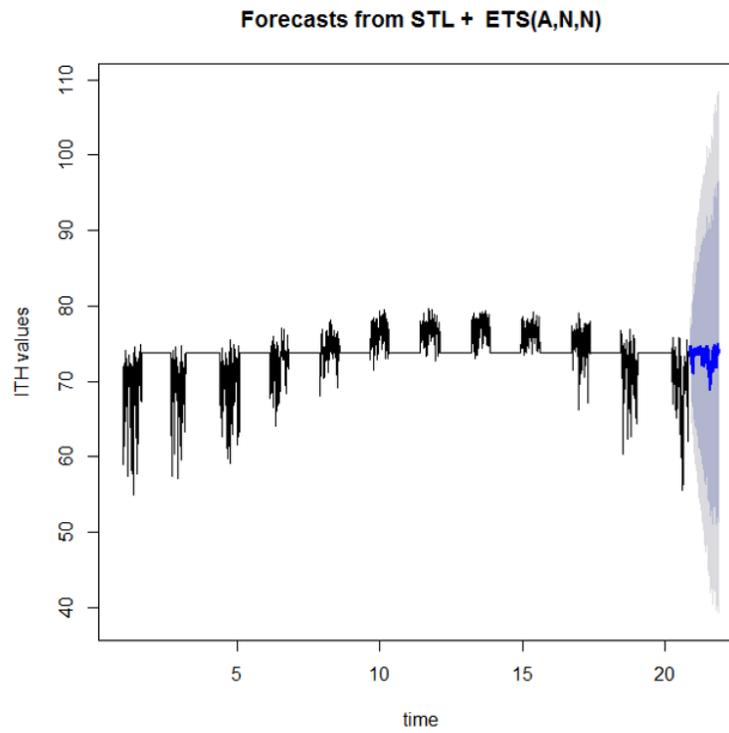


Figura 6. Autocorrelación para la serie temporal.



**Figura 7.** Pronóstico Holt Winters (Estación 78308).



**Figura 8.** Pronóstico STL + ETS (A,N,N).

Seguidamente se muestran los resultados obtenidos para la predicción automatizada utilizando un suavizado exponencial (método *Holt Winters*).

De acuerdo con la literatura consultada, la modelación ARIMA se ha aplicado fundamentalmente en el pronóstico de la producción lechera ([Deluyker et al., 1990](#); [Sánchez, 2013](#)). Otros trabajos dentro de esta línea, se aproximan al empleo de una estimación del inventario ganadero bovino, tal y como muestran [Cuenca et al. \(2008\)](#) y [Pérez \(2014\)](#).

### CONCLUSIONES

Se logró predecir las condiciones de los animales mediante modelos ARIMA de series temporales para el escenario nacional. Estos métodos de análisis dan una alternativa útil a los modelos lineales clásicos para el análisis de series como las del índice de temperatura y humedad.

Tanto para la escala semanal como mensual, los mejores ajustes se obtuvieron en el grupo correspondiente a la zona oriental de Cuba. Se observa que el mejor corresponde al grupo oriental de Cuba: (1,1, 2)(0,0,1)<sub>52</sub> y (0,0,1)(0,1,1)<sub>12</sub>; con errores medios porcentuales (MPE) igual a -0.017 y -0.06 respectivamente.

### REFERENCIAS

Aguilar Ramírez, M., 2006. *Antecedentes: El equilibrio térmico y los parámetros ambientales. En: Control ambiental en alojamientos ganaderos de Navarra II*. [En línea] Available at: <http://www.itgganadero.com/itg/portal/documentos.asp?Id=181> [Último acceso: 2018 ].

Blaconá, M. T., 2018. *Breve reseña de métodos de análisis y pronósticos probabilísticos en series de tiempo*. [En línea] [Último acceso: 22 Agosto 2018 ].

Caraza Hernández, R. & Quintero Bermúdez, E., 1997. Elementos de climatología. En: *Agrometeorología*. Santa Clara: Universidad Central Marta Abreu de Las Villas. Publicaciones CDICT, pp. 295 - 351.

Cuenca Jiménez, N. J., Chavarro Miranda, F. & Díaz Gantiva, O. H., 2008. El sector de la

ganadería bovina en Colombia. Aplicación de series de tiempo al inventario ganadero. *Revista de la Facultad de Ciencias Económicas*, XVI(1), pp. 165-177.

Deluyker, H. A. y otros, 1990. Modeling Daily Milk Yield in Holstein Cows Using Time Series Analysis. *Journal of Dairy Science*, pp. 539-548.

Domínguez-Hurtado, I. M., Moya-Álvarez, A. S. & Estrada-Moreno, A., 2010. Aplicación de la salida del modelo GFS/NCEP en la predicción agrometeorológica. *Ingeniería Agrícola y Biosistemas*, 2(1), pp. 13-19.

Guevara Viera, G. y otros, 2004. Efecto del cambio de precio de la leche sobre algunos indicadores de producción lechera y pronóstico a corto plazo. *Revista de Producción Animal*, 16(2).

Hyndman, R. J., 2008. Automatic Time Series Forecasting: the forecast Package for R. *Journal of Statistical Software*.

Hyndman, R. J. & Athanasopoulos, G., 2013. The forecaster's toolbox. En: *Forecasting: principles and practices*. s.l.:s.n., pp. 14-33.

Hyndman, R. J. & Athanasopoulos, G., 2018. Time series graphics. En: *Forecasting: Principles and Practice*. Segunda ed. s.l.:Monash University.

Moraes, I. M., s.f. *Crecimiento, tecnología y competitividad en la ganadería uruguaya entre 1870-1930*. [En línea] Available at: <http://cdn.fee.tche.br/jornadas/1/s7a3.pdf> [Último acceso: 2 Julio 2019 ].

Nienaber, J. A. & Hahn, G. L., 2007. Livestock production system management responses to thermal challenges. *International Journal of Biometeorology*, Volumen 52, pp. 149-157.

Osés Rodríguez, R. y otros, 2017. Modelación y predicción de la fasciolosis en Villa Clara, Cuba. *Biotempo*, 14(1).

Pérez V, G. J., 2014. Los ciclos ganaderos en Colombia, 1950 - 2001. 46, p. 46.

R Core Team, 2018. *R: A language and environment for statistical computing*, Vienna: s.n.

- Sánchez Molina, J. & Poveda Jaramillo, G., 2006. Aplicación de los métodos MARS, Holt - Winters y ARIMA generalizado en el pronóstico de los caudales medios mensuales en ríos de Antioquía. *Meteorología Colombiana*, Marzo, Issue 10, pp. 36-46.
- Sánchez Suárez, L., 2013. *Pronóstico de la producción de leche, mediante modelos ARIMA. Caso UBPC "Maniabo"*, La Habana: s.n.

Los autores de este trabajo declaran no presentar conflicto de intereses.

Los autores de este trabajo declaran presentar una participación igualitaria en la concepción, ejecución y escritura de la investigación.

Este artículo se encuentra bajo licencia [Creative Commons Reconocimiento-NoComercial 4.0 Internacional \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)