



Universitat
de les Illes Balears

Análisis y Predicción del Indicador de Presión Humana en les Illes Balears

AUTOR: *Juan Pol Hernández*

Memoria del Trabajo de Fin de Máster

Máster Universitario en Análisis de Datos Masivos en Economía y Empresa

de la

UNIVERSITAT DE LES ILLES BALEARS

Curso Académico *2017-2018*

Fecha Junio 2018

Nombre Tutor del Trabajo Tomás del Barrio Castro

Juan Pol Hernández
Tutor: Tomás del Barrio Castro

Treball de fi de Màster Universitari en Anàlisi de Dades Massives en Economia i Empresa
(MADM)

Universitat de les Illes Balears
07122 Palma de Mallorca
juanpolhernandez@gmail.com

Resumen

En este trabajo se presentan una serie de modelos, que predicen el año 2017 del Indicador de Presión Humana (IPH) de les Illes Balears, Mallorca, Menorca i les Pitiüses¹. Una vez conocido el mejor modelo para cada una de las series, se hace una predicción hasta el año 2030.

Abstract

This paper presents a series of models, which predict the year 2017 of the Human Pressure Indicator (IPH) of the Balearic Islands, Mallorca, Menorca i les Pitiüses. Once the best model for each of the series is known, a prediction is made until the year 2030.

Palabras clave: Series Temporales, IPH, Illes Balears

1. Base de Datos

Las bases de datos utilizadas en este trabajo son las del indicador de presión humana (IPH) de les Illes Balears y de las islas de Mallorca, Menorca y Eivissa-Formentera. Los datos del IPH se pueden conseguir de l'institut d'estadística de les Illes Balears (Ibestat²)[4]. Las series temporales de estos datos están disponibles desde el año 1997 hasta el año 2017, los datos que van desde 1997 hasta 2016, son datos definitivos, y los datos de 2017³ se encuentran en estado provisional. Estos datos se ofrecen en frecuencia diaria. Por lo que cada una de las cuatro series tiene 7305 datos diarios. El programa utilizado ha sido R-Studio, y junto con esta memoria se adjuntan los archivos utilizados.

2. IPH

El IPH estima la carga demográfica real que soporta un cierto territorio diariamente. El IPH, está

formado por la suma de población residente⁴ (R), y de población no residente⁵ (NR). El valor del IPH de un día específico puede ser diferente al de otro día, causado por la variación de la población no residente y Residente. A continuación se puede ver la ecuación⁶ del IPH, que representa la suma de Residentes y no residentes en un momento t:

$$IPH_t = R_t + NR_t \quad (1)$$

En la figura 1 y 2, se visualiza el IPH de les Illes Balears, Mallorca, Menorca y les Pitiüses. Se puede observar que existe estacionalidad en las cuatro series, y que este patrón estacional tiende a repetirse año tras año.

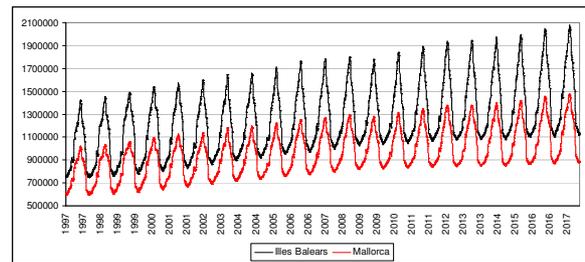


Figura 1: IPH de les Illes Balears y Mallorca

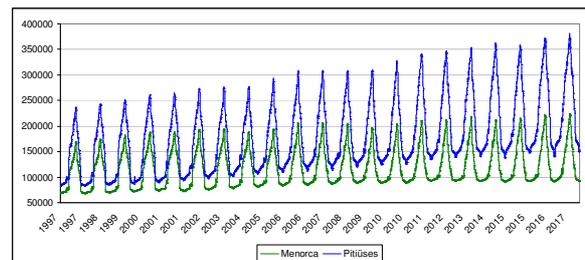


Figura 2: IPH de Menorca y les Pitiüses

Las series temporales vistas en las figuras 1 y 2, son series no estacionarias. Por lo que hay que aplicar logaritmos⁷ para que la varianza sea

¹ Son las Islas de Eivissa, Formentera y otros islotes.

² Las fuentes que utiliza Ibestat para obtener estimar estos datos se pueden ver en su página Web.

³ La última descarga se realizó el día 07/05/2018.

⁴ Datos de las Cifras oficiales de población resultantes de la revisión del Padrón municipal[5]

⁵ Conocida como población estacional, puede permanecer en cada una de las islas durante un cierto tiempo, como diariamente o quincenalmente.

⁶ A la hora de calcular el IPH, se tienen en cuenta otros factores como mortalidad, nacimientos,..etc. Esta ecuación intenta dar una idea general.

⁷ Consiste en aplicar logaritmos neperianos a cada uno de los datos de las series diarias, para conseguir que la varianza sea estable o constante en toda la serie.

estacionaria, y después aplicar el operador de diferencias⁸ para obtener estacionariedad en media tal y como se puede ver en la figura 3. En esta figura se puede observar que la media es constante, y que hay datos que se alejan mucho de la media. Esta variación puede venir explicada por los efectos calendario, algunos de estos días afectan realmente a las series, como por ejemplo la Semana Santa. Y por los valores atípicos, como por ejemplo Outlier Innovacional, Outlier Aditivo y level shift.

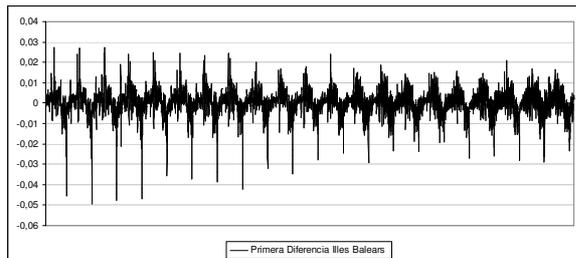


Figura 3: Primera Diferencia IPH de les Illes Balears

En las figura 4 y 5, se pueden comparar los últimos 4 años para cada una de sus islas y en su conjunto. Todas tienen en común que alrededor del día 100, la inclinación de la curva es más elevada, causado por la llegada de turistas. Así como van pasando los años se alcanza un valor del IPH más alto, que suele coincidir con el día 225. Y a partir de aquí empieza a descender. Durante el mes de diciembre y enero la cifra del IPH llega a sus valores mínimos. Y así como nos vamos acercando a los meses de verano esta cifra va aumentando, coincidiendo con los meses de más afluencia turística. Así como van pasando los años, va creciendo la cantidad máxima anual que se alcanza en cada uno de los diferentes años. Destacar que los extremos de Menorca son casi planos, indicando que durante esos meses, las cifras son parecidas a la población residente.

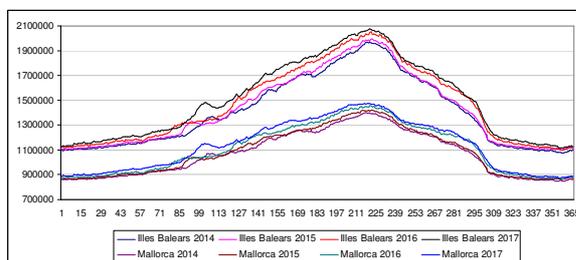


Figura 4: IPH de les Illes Balears y Mallorca por años

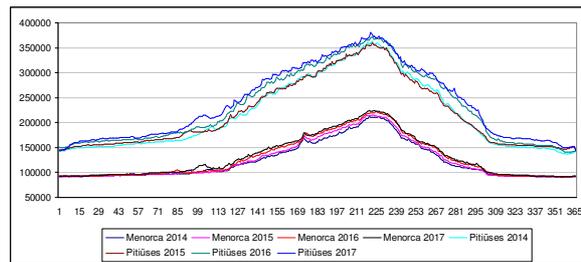


Figura 5: IPH de Menorca y les Pitiüses por años

3. El modelo

El modelo econométrico que puede explicar el crecimiento temporal⁹ del IPH, es el siguiente:

$$\ln(Y_t) = \delta + \mu t + \varphi f_t + \sum_{j=1}^6 \gamma_j d_{j,t} + \sum_{j=1}^k \left[\alpha_j \cos\left(\frac{2\pi j}{365} t\right) + \beta_j \sin\left(\frac{2\pi j}{365} t\right) \right] + \varepsilon_t \quad (2)$$

Donde $\ln(Y_t)$, es el valor del IPH de un día t en logaritmos neperianos. δ es la constante. μ es el coeficiente asociado a la tendencia temporal t . φ es un vector que recoge los coeficientes asociados a las variables ficticias f_t que representa los efectos calendario¹⁰. γ_j ($j=1, \dots, k$) son los parámetros asociados a los días de la semana, y $d_{j,t}$ ($j=1, \dots, 6$) son seis variables ficticias, que indican el día de la semana. $\cos(2\pi j t / 365)$ y $\sin(2\pi j t / 365)$ para $j=1, \dots, k$, son un conjunto de 365 de variables trigonométricas ficticias que recogen la estacionalidad determinista anual del IPH utilizando la transformada de Fourier¹¹, y sus coeficientes asociados son α_j y β_j . Y ε_t es el término de perturbación.

La ecuación (2) se estima utilizando mínimos cuadrados ordinarios. El término de perturbación ε_t , sigue diferentes procesos (ARIMAxSARIMA, BSM, TBSM o ANN), los cuales se verán en la metodología. Para el modelo PAR, se utiliza el mismo modelo pero excluyendo las variables ficticias días de la semana, ya que una vez estimada la ecuación 2, se crea una matriz $n \times 7$ con el término de perturbación, para separar los datos en días de la semana. Y para los modelos BATS Y TBATS no se aplica el modelo que se ha visto aquí, ya que utilizan una estructura propia que se verá en el apartado de metodología.

⁹ Se puede ver un modelo parecido en Haldrup et al.[12] o Sansó[15].

¹⁰ Los efectos calendario son aquellos días de fiesta que se aprueban por el Govern de les Illes Balears[1].

¹¹ Se puede consultar el capítulo 11 de la referencia[19], para entender la transformada de Fourier.

⁸ En el apartado 4.1 se explicará.

4. Metodología

En este apartado se van a comentar siete modelos econométricos de series temporales que son los que utilizan en este trabajo. No se plantea explicar en profundidad cada uno de ellos, pero si el lector está interesado puede acudir a las referencias.

4.1. ARIMA x SARIMA

El modelo ARIMAxSARIMA¹² es un conjunto de procesos lineales. Estos procesos lineales son: los procesos autoregresivos AR(p), porque x_t se obtiene a partir de la autoregresión de sus valores pasados. En los procesos de media móvil MA(q), x_t se obtiene a partir de las medias móviles de las variables de ruido blanco. Y el operador de diferenciación I(d)¹³, es necesario para que la media sea estacionaria. Ver Sansó[14] para más información.

$$\Phi_p(L^s)\phi_p(L)\Delta_s^d\Delta^d x_t = \delta + \Theta_q(L^s)\theta_q(L)\varepsilon_t \quad (3)$$

Donde:

- δ : es la constante.
- $\phi_p(L)=(1-\phi L-\phi L^2, \dots, -\phi L^p)$: Es un polinomio autoregresivo de grado p.
- $\theta_q(L)=(1-\theta L-\theta L^2, \dots, -\theta L^q)$: Es un polinomio de media móvil de grado q.
- $\Phi_p(L^s)=(1-\Phi L^s-\Phi L^{2s}, \dots, -\Phi L^{ps})$: Es un polinomio autoregresivo estacional de grado P+s.
- $\Theta_q(L^s)=(1-\Theta L^s-\Theta L^{2s}, \dots, -\Theta L^{qs})$: Es un polinomio de media móvil estacional de grado Q+s.
- $\Delta^d=(1-L)^d$: Término que indica el número d de diferencias regulares.
- $\Delta^D=(1-L)^D$: Término que indica el número D de diferencias estacionales.
- $\varepsilon_t \sim (0, \sigma^2)$: Término de perturbación, o Ruido Blanco¹⁴, con media cero y varianza constante.

4.2. Modelos estructurales

Los modelos estructurales que se verán son: los modelos Basic Structural Model (BSM) y Trigonometric Basic Structural Model (TBSM). Además de aplicar el filtro de Kalman. Para más información ver Ngai Hang Chan[2], Harvey[13], y Wei[19]

¹² También conocida como metodología Box-Jenkins

¹³ Operador de diferenciación Δ . $\Delta = I-L$ donde L es el operador del retardo $Lx_t = x_{t-1}$. $\Delta x_t = (1-L)x_t = x_t - x_{t-1}$. La estacionariedad se puede conseguir a partir de ir aplicando sucesivas diferenciaciones hasta conseguir la estacionariedad en media

¹⁴ Es un proceso estocástico con media cero, varianza constante y no existe autocorrelación.

4.2.1. Basical Structural model (BSM)

El modelo que se utiliza es un modelo estructural básico con efectos estacionales, el cual es una representación local level model with seasonal effects, en el que el nivel (level) en cada momento es la suma del valor del período anterior y un elemento aleatorio. Se puede representar de la siguiente manera:

$$y_t = \mu_t + \gamma_t + \varepsilon_t \quad (4)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \xi_t \quad (5)$$

$$\beta_t = \beta_{t-1} + \zeta_t \quad (6)$$

$$\gamma_t = \sum_{j=1}^{s-1} \gamma_{t-j} + w_t \quad (7)$$

Donde:

- $t=s, \dots, n$. s es igual a la periodicidad de los datos.
- μ_t : Tendencia estocástica.
- γ_t : Componente estacional estocástico.
- β_t : Componente de inclinación.
- $\varepsilon_t, \xi_t, \zeta_t$ y w_t : Componentes irregulares con media cero y varianza constante, los cuales son independientes y no correlacionados entre ellos.

4.2.2. BSM trigonométrico (TBSM)

Es un modelo parecido al anterior, pero con la diferencia de que ahora la parte estacional, tiene una representación trigonométrica. La ecuación para la parte estacional es la siguiente:

$$\gamma_t = \sum_{j=1}^{[s/2]} \gamma_{jt} \quad (8)$$

$$\left. \begin{aligned} \gamma_{jt} &= \gamma_{s,j-1} \cos(v_j) + \gamma_{j,t-1}^* \sin(v_j) + w_{jt} \\ \gamma_{jt}^* &= -\gamma_{j,t-1} \sin(v_t) + \gamma_{j,t-1}^* \cos(v_j) + w_{jt}^* \end{aligned} \right\} \quad (9)$$

Donde:

- $v = 2\pi j/s$, $j=1, 2, \dots, [s/2]$
- γ_t : Componente estacional estocástico.
- γ_t^* : Se utiliza para la construcción del modelo, su representación puede ser ignorada.
- w_{jt} y w_{jt}^* : Procesos de ruido blanco que no están correlacionados entre ellos, y tiene una varianza común.

Los modelos anteriores se pueden representar en forma de espacio de los estados (State Space):

$$y_t = Z\alpha_t + \varepsilon_t \quad (10)$$

$$\alpha_t = T\alpha_{t-1} + R\eta_t \quad (11)$$

$$\alpha_0 \sim N(a_0, P_0) \quad (12)$$

Donde:

- Z , T , y R : Son unas matrices de ceros y unos.
- $\alpha_t = (\mu_t, \beta_t, \gamma_t, \dots, \gamma_{t-S+2})$: Son los parámetros.
- ε_t : Término de perturbación con media cero y varianza constante σ_ε^2 .
- η_t : Término de perturbación con media cero y matriz de covarianza $Q = [\sigma_\xi^2, \sigma_\zeta^2, \sigma_w^2 I]$.
- α_0 : La media y la varianza del vector estado inicial.

Se supone que α_0 y P_0 son conocidos o elegidos de antemano. Dados los parámetros de varianza, se puede aplicar el filtro Kalman y/o un suavizador (smoothing) para extraer una estimación de los componentes.

4.2.3. Filtro de Kalman

El filtro de kalman es un conjunto de ecuaciones matemáticas que sirven para identificar el estado oculto de un sistema dinámico lineal, y debe existir ruido blanco. La ganancia de retroalimentación del error del filtro de kalman, es capaz de escogerla de forma óptima cuando se conocen las varianzas de los ruidos que afectan al sistema. Permiten calcular un estimador lineal, insesgado y óptimo del estado de un proceso en cada momento del tiempo, teniendo como base la información disponible en el momento t-1, y junto con la información disponible en el momento t, permite actualizar dichas estimaciones.

El filtro de kalman asume que la media y la varianza del vector estado inicial son conocidos $\alpha_0 \sim N(\alpha_0, P_0)$. Estos datos se pueden obtener, a partir de una experiencia previa, o darles un valor cero¹⁵.

El filtro de Kalman está formado por el siguiente sistema de ecuaciones:

Predicción del vector estado

$$E[\alpha_{t+1} | I_t] = a_{t+1|t} = T a_t \quad (13)$$

Dispersión de la predicción

$$VAR[a_{t+1|t}] = P_{t+1|t} = T P_t T' + R Q R' \quad (14)$$

Error de predicción

$$e_{t+1|t} = y_{t+1|t} - Z a_{t+1|t} \quad (15)$$

Dispersión del error de predicción

$$VAR[e_{t+1|t}] = f_t = Z P_{t+1|t} Z' + h \quad (16)$$

Ganancia de Kalman

$$G_{t+1} = P_{t+1|t} Z' f_t^{-1} \quad (17)$$

¹⁵ Este procedimiento funciona para bases de datos grandes, donde al final los valores tendera a su valor real

Estimación del vector estado

$$E[\alpha_{t+1} | I_{t+1}] = a_{t+1} = a_{t+1|t} + G_{t+1} e_{t+1|t} \quad (18)$$

Dispersión de la estimación del vector estado

$$P_{t+1} = (I - G_{t+1} Z) P_{t+1|t} \quad (19)$$

4.3. BATS y TBATS

El acrónimo del modelo BATS es Box and Cox transformation, ARMA errors, trend and seasonal components. El modelo Bats es un modelo que combina State Space Model, transformación Box-Cox¹⁶ y el modelo ARMA para los residuos. Hay mucho autores que hablan sobre este tema como De Livera et al. [3].

4.3.1. BATS

Este modelo funciona bien siempre y cuando haya homocedasticidad¹⁷, en el caso de que nos encontremos con el problema de la no linealidad, se aplica la transformación Box-Cox. El modelo BATS no funciona bien cuando la estacionalidad es compleja o de alta frecuencia. Para dar solución a este problema aparece el modelo TBATS, que introduce una representación trigonométrica de los componentes estacionales basados en las series Harvey [13] y West y Harrison Fourier [20]

La representación de un Modelo BATS se formularia de la siguiente manera:

$$y_t^{(w)} = \begin{cases} \frac{y_t^{(w)} - 1}{w}; & w \neq 0 \\ \log y_t & w = 0 \end{cases} \quad (20)$$

$$y_t^{(w)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t \quad (21)$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \quad (22)$$

$$b_t = (1 - \phi) b + \phi b_{t-1} + \beta d_t \quad (23)$$

$$s_t^{(i)} = s_{t-m_i}^i + \gamma_i d_t \quad (24)$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (25)$$

¹⁶ Es la ecuación 20, se puede calcular λ mediante algún procedimiento de máxima verosimilitud o aplicando diversos λ para ver cual hace que la varianza sea constante al largo del tiempo.

¹⁷ Homocedasticidad: La varianza de los errores es constante a lo largo del tiempo. El término contrario es la heterocedasticidad.

Donde:

- i = la i -ésima estacionalidad.
- $y_t^{(w)}$: Representa la transformación Box-Cox con el parámetro w , donde la y_t es la observación en el momento t .
- m_1, m_2, \dots, m_T : Son los periodos estacionales.
- l_t : Es el local level¹⁸ en el periodo t .
- b : Es la tendencia a largo plazo, y b_t es la tendencia a corto plazo en el periodo t .
- $s_t^{(i)}$: representa el i -ésimo componente estacional en el momento t .
- $\{d_t\}$: En un proceso ARMA (p,q) y $\{\varepsilon_t\}$ es un proceso de ruido blanco Gaussiano con una media zero y una varianza σ^2 constante
- α, β , y γ_t : Son los parámetros de suavizado (The smoothing parameters) para $i=1, \dots, T$.
- ϕ : Es el coeficiente asociado a la tendencia temporal t .

4.3.2. TBATS

Las ecuaciones son las mismas que el modelo BATS, menos la parte estacional.

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)} \quad (26)$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \text{sen} \lambda_j^{(i)} + \gamma_1^{(i)} d_t \quad (27)$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \text{sen} \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t \quad (28)$$

Donde:

- i = la i -ésima estacionalidad.
- $\gamma_1^{(i)}$ y $\gamma_2^{(i)}$: Son los parámetros de suavizado (smoothing parameters).
- $\lambda_j^{(i)} = 2\pi j/m_i$. Conjunto de variables trigonométricas ficticias que recogen la estacionalidad determinista anual.
- $s_{j,t}^{(i)}$: Es el nivel estocástico del componente estacional i -ésimo como.
- $s_{j,t}^{*(i)}$: Es el crecimiento estocástico del i -ésimo componente estacional, necesario para describir los cambios estacionales a lo largo del tiempo.
- k_i : La cantidad de armónicos requeridos para el componente estacional i -ésimo.

4.4. Red Neuronal Artificial (ANN)

4.4.1. ANN

Una red neuronal artificial (Artificial neural network), se puede considerar como una red de neuronas que se organizan en capas: los datos iniciales (inputs layer), una o varias capas ocultas

(hidden layer) - en cada una de estas capas ocultas pueden existir diferentes nodos- y una capa final (Output layer). Las redes neuronales están formadas por nodos (que es donde se calcula el modelo) y de aristas (que son las que conectan los nodos). Estas aristas pueden tener diferentes pesos; es decir, hay aristas que tendrán más importancia que otras. En el caso de utilizar una red neuronal para series temporales, se utiliza un modelo autoregresivo AR(p), y el tipo de red neuronal que se utilizará es del tipo *multilayer feed-forward network*, con una capa oculta de n nodos. Para datos estacionales, se agregan los últimos valores observados de la misma estación que los inputs. Los datos iniciales saldrán de la capa input, y se dirigirán a los nodos que se encuentran en la capa oculta, por lo que cada arista tendrá un peso diferente aleatorio¹⁹ inicial, y estos se irán actualizando usando los datos observados. Dentro de cada nodo de la capa oculta, se utiliza un modelo y su resultado es luego modificado por una función de activación no lineal. De los nodos de la capa oculta saldrán los resultados modificados que se dirigirán hacia la capa final a través de las aristas con unos pesos, donde se volverán a calcular, (exactamente igual que los nodos de la capa oculta) para obtener el resultado final. Este proceso se puede ir repitiendo las veces que sea necesarios hasta que se encuentren los pesos óptimos, y poder dar un resultado final. Y finalmente se tendrá una red neuronal entrenada.

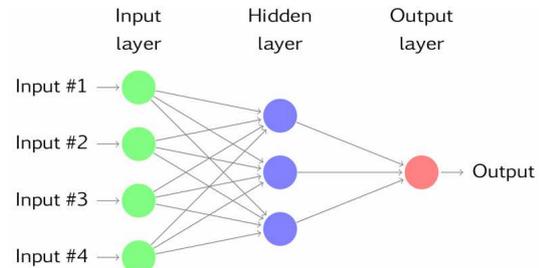


Figura 5: Red neuronal

Tal y como se puede ver en Franses y Dijk [8], o en Teräsvirta et al.[17]. Una red neuronal puede ser representada matemáticamente de la siguiente manera:

$$y_t = \beta_0^j w_t + \sum_{j=1}^q \beta_j G(\gamma_j^j w_t) + \varepsilon_t \quad (29)$$

Donde:

- $w_t = (1, y_{t-1}, \dots, y_{t-p})$: Consiste en una constante, y de p retardos para y_t

¹⁸Es la suma del valor del período anterior y un elemento aleatorio.

¹⁹ En el programa R-Studio he aplicado una semilla (set seed), para obtener siempre los mismos resultados. Además al tener las aristas un peso diferente, un dato puede ir a un nodo con un peso, y a otro nodo con otro peso.

- $\beta_j, j=1, \dots, q$: Son los parámetros y se suelen llamar fortalezas de conexión (connection strengths), o pesos.

- $\gamma_j, j=1, \dots, q$: Son vectores de parámetros.

- $\beta_0^j w^t$: Es la unidad lineal, algunos usuarios asumen que $\beta_0=0$.

- $\varepsilon_t \sim (0, \sigma^2)$: Término de perturbación.

La función logística sigmoidal, es aplicada como función de activación no lineal.

$$g(x) = \frac{1}{1 + \exp^{-x}} \quad (30)$$

4.5. Periodic Autoregression (PAR)

4.5.1. PAR

Para resumir lo que es un modelo PAR (p), me he basado Franses and Paap[9]. Tal y como se puede ver en Ghysels and Osborn[10], Los procesos periódicos son procesos en que los coeficientes cambian con las estaciones del año. Además los modelos PAR asumen que las observaciones correspondientes a cada estación pueden ser descritas utilizando diferentes modelos Franses[7]. Por lo que los modelos PAR son útiles para captar procesos que combinan diferentes estacionalidades.

Un modelo periódico autoregresivo de orden PAR(P), se puede ver en la ecuación 31 con doble subíndice: τ periodos ($\tau=1,2,\dots,N$), y s representa la estación ($s=1,2,\dots,S$). En este trabajo, las estaciones son los días de la semana con lo que $s=1, 2, 3, 4, 5, 6$ y 7 . Y τ se refiere a la semana, donde N es la cantidad de semanas que hay.

$$y_{s\tau} = \mu_s + \phi_{1s} y_{s-1,\tau} + \dots + \phi_{ps} y_{s-p,\tau} + \varepsilon_{s\tau} \quad (31)$$

Donde:

- μ_s : Es la constante para cada estación.

- $\phi_{1s}, \dots, \phi_{ps}$: Son los parámetros autoregresivos de orden p_s , los cuales pueden variar con la estación s ($s=1,2,\dots,S$). Si la base de datos tiene frecuencia diaria, los datos se ordenan por semanas, $s=1,2,\dots,7$.

- $\varepsilon_{s\tau}$: Es el término de perturbación. Se asume que es un proceso de ruido blanco con una varianza constante σ^2 y una media de cero.

Se pueden calcular los modelos PAR, utilizando la representación multivariante²⁰. Así que la ecuación 31 puede expresarse en términos

²⁰ Ver referencias [11] y [18], y además en Haldrup et al.[12] la denominaron representación de vector días, al trabajar con datos diarios con vectores de dimensiones 7×1

multivariantes como un VAR(P) siendo $P = \text{int}[(p+S-1)/S]$:

$$\Phi_0 Y_\tau = \mu + \Phi_1 Y_{\tau-1} + \Phi_2 Y_{\tau-2} + \dots + \Phi_p Y_{\tau-p} + E_\tau \quad (32)$$

Donde:

- La matriz $Y_i = [y_{i1}, \dots, y_{iS}]$: Compuesta por el vector $S \times 1$ son las observaciones de cada una de las estaciones, si trabajamos con las observaciones de una semana será de 7×1 , que corresponderá con una semana completa.

- $\mu_i = [\mu_1, \dots, \mu_s]$: Son las constantes del modelo que forman un vector $S \times 1$.

- $E_i = [\varepsilon_{i1}, \dots, \varepsilon_{is}]$: Son las innovaciones con valor esperado de cero y la matriz de varianzas y covarianzas es igual a $\Omega = E(E_\tau E_\tau^T) = \text{diag}(\sigma_1^2, \dots, \sigma_s^2)$.

Las matrices Φ_j de orden $S \times S$ con $j=1, \dots, P$

$$\Phi_0 = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -\phi_{1,2} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ -\phi_{p-1,s} & -\phi_{p-2,s} & \dots & -\phi_{1,s} & 1 \end{pmatrix}$$

$i=1, \dots, P$

$$\Phi_i = \begin{pmatrix} \phi_{ip,1} & \phi_{ip-1,1} & \dots & \phi_{(i-1)p+2,1} & \phi_{(i-1)p+1,1} \\ \phi_{ip+1,2} & \phi_{ip,2} & \dots & \phi_{(i-1)p+3,2} & \phi_{(i-1)p+2,2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \phi_{(i+1)p-1,s} & \phi_{(i+1)p-2,s} & \dots & \phi_{ip+1,s} & \phi_{ip,s} \end{pmatrix}$$

Finalmente se puede reescribir el modelo²¹, como:

$$y_\tau = \Phi_0^{-1} \Phi_1 y_{\tau-1} + \Phi_0^{-1} \varepsilon_\tau \quad (33)$$

5. Predecir

Una vez conocidos los coeficientes de la ecuación 2 (sin el término de perturbación ε_0 , para hacer una predicción hay que ir añadiendo más días $t+h$ a la ecuación 34, donde h son los días que se quieren estimar. Después hay que predecir el término de perturbación ε_t , utilizando los modelos vistos en la metodología. Con las excepciones del modelo PAR (que utiliza el mismo modelo pero sin incluir las variables ficticias días de las semana) y de los modelos BATS y TBATS que utilizan sus propios algoritmos. Al final para obtener los nuevos datos diarios, hay que sumar ambas partes, tal y como se muestra en la ecuación 35, para cada uno de los modelos. $\text{Ln}(Y_{t+h}^*)$ es el resultado de la

²¹ Ver Franses y Paap[9]

predicción para el día $t+h$ de la parte determinista del modelo, y ε_{t+h} , es el resultado predicho que se ha obtenido de los modelos de la metodología. Habrá una predicción para ε_{t+h} , diferente para cada uno de los modelos de la metodología, cosa que permitirá comparar que modelo predice mejor. Al final al sumar ambas parte, se obtiene D_{t+h} , el valor predicho para un momento o día $t+h$.

$$\ln(Y_{t+h}^*) = \delta + \mu(t+h) + \varphi_{t+h} + \sum_{j=1}^6 \gamma_j d_{j,t+h} + \sum_{j=1}^k \left[\alpha_j \cos\left(\frac{2\pi j}{365}(t+h)\right) + \beta_j \sin\left(\frac{2\pi j}{365}(t+h)\right) \right] \quad (34)$$

$$D_{t+h} = \ln(Y_{t+h}^*) + \varepsilon_{t+h} \quad (35)$$

6. Fórmulas

Las fórmulas para calcular los errores de predicción serán las siguientes:

6.1. Error cuadrático medio (MSE)

El error cuadrático mide la media de los errores al cuadrado. Donde \hat{y}_i son los i valores estimados, y_i son los i valores reales, y n es el número total de valores.

$$MSE = \frac{1}{n} \sum_{i=0}^n \left(\hat{y}_i - y_i \right)^2 \quad (36)$$

6.2. Error Medio Absoluto (MAE)

El error medio absoluto mide la media de los errores en valor absoluto. Donde \hat{y}_i son los i valores estimados, y_i son los i valores reales, y n es el número total de valores.

$$MAE = \frac{\sum_{i=0}^n \left| \hat{y}_i - y_i \right|}{n} \quad (37)$$

6.3. Error porcentual absoluto medio (MAPE)

El Error porcentual absoluto medio mide la media entre la variación porcentual en valor absoluto, y se expresa en porcentaje. Donde \hat{y}_i son los i valores estimados, y_i son los i valores reales, y n es el número total de valores.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|}{n} * 100 \quad (38)$$

7. Resultados prediciendo el año 2017

Los datos diarios que van del año 1997 hasta 2016, son los datos de entrenamiento (7305 datos diarios), y dejando el año 2017 (365 datos diarios) para probar o comparar con la predicción de los diferentes modelos. Para les Illes Balears, se podrá ver si es mejor predecir con sus datos, o sumando las predicciones que se obtienen de Mallorca, Menorca y les Pitiüses. Con las predicciones de los modelos, se han creado tres nuevas series diarias²² combinando los modelos para ver cual de ellos predice mejor dentro de unos intervalos de tiempo, y comparando sus resultados con los datos reales del año 2017. A estos intervalos de tiempo, les he llamado combinación trimestral, combinación mensual y combinación diaria. El modelo que predice mejor por meses, trimestres o días se ha obtenido de las formulas de los errores utilizadas en el apartado 6, pero poniendo intervalos y escogiendo el que tiene un error menor por intervalo. Para la combinación mensual y trimestral se dirán que modelos predicen mejor, pero para la combinación diaria se dará su frecuencia absoluta y relativa²³.

Se indicará que modelo ARIMAxSARIMA ha sido elegido. He mirado que modelo tenia un AIC menor, y también he aplicado el principio de parsimonia²⁴. El paquete de R-Studio utilizado para calcular la red neuronal, utiliza la siguiente expresión: $NNAR(p,P,k)_m$ el cual es equivalente a $ARIMA(p,0,0)(P,0,0)_m$, donde: p es el grado autoregresivo de la parte Arima, P es el grado autoregresivo de la parte Sarima, k es la cantidad de nodos utilizados dentro de la capa oculta, y m indica la frecuencia de los datos, en este caso es de 365. Se mostrará para el modelo PAR, que modelo VAR ha sido elegido utilizando el criterio de AIC.

7.1. Illes Balears

El modelo ARIMAxSARIMA(0,1,0)(1,1,1) es el elegido para les Illes Balears. La red neuronal

²² Estas series diarias, representan la predicción del año 2017, por lo cada serie esta formada por 365 datos diarios

²³ Sobre los 365 días predichos.

²⁴ El principio de Parsimonia: En igualdad de condiciones, siempre se escogerá el modelo más sencillo.

utilizada es NNAR(1,1,2)₃₆₅. Y para el modelo PAR, se ha utilizado un VAR(7). La red neuronal es el modelo que tiene unos errores menores respecto a los otros modelos, se puede observar en la tabla 1. Si predecimos les Illes Balears con las predicciones de Mallorca, Menorca y les Pitiüses, los errores pueden ser menores o mayores dependiendo del modelo y el tipo de error. Pero si se selecciona la combinación diaria, este es el que tiene un error más bajo.

La combinación trimestral ha elegido como mejor modelo para el primer trimestre el modelo ARIMAxSARIMA, y para los otros tres trimestres el modelo ANN. Pero si predecimos les Illes Balears con los datos predichos de las islas su combinación trimestral es igual, menos en el tercer trimestre que el modelo elegido es el PAR. La combinación mensual selecciona para los tres primeros meses el modelo ARIMAxSARIMA, el cuarto mes es el modelo ANN, el quinto mes es el modelo PAR, los siguientes cinco meses el modelo elegido es ANN, y los dos últimos meses PAR y ARIMAxSARIMA. Y utilizando los datos de las 3 islas para predecir en su conjunto los resultados son parecidos pero los meses que cambian son el séptimo y noveno mes, eligiendo el modelo PAR, y los meses ocho y doce escogen el modelo TBATS. En la combinación diaria el modelo ANN es el que tiene una mayor frecuencia absoluta y relativa de 134, y 36,6% respectivamente, seguido por el modelo PAR con una frecuencia absoluta y relativa de 83 y 22.73%. Y utilizando las predicciones aunadas de las 3 islas, el modelo ANN tiene una frecuencia absoluta y relativa de 125 y 34%, seguido del modelo PAR con una frecuencia absoluta y relativa de 78 y 21.31%.

En la figura 6 y 7 se muestran los resultados en forma gráfica, donde la línea negra son los valores reales, la roja la predicción del modelo, la verde la predicción agregada de las tres islas, y de color azul los intervalos de confianza.

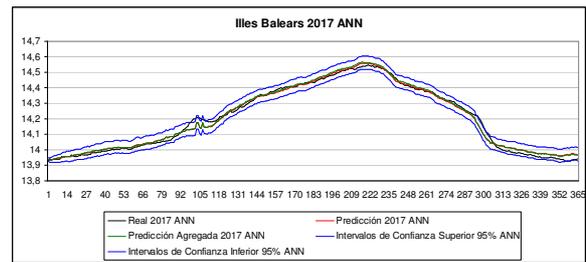


Figura 6: Predicción ANN para les Illes Balears

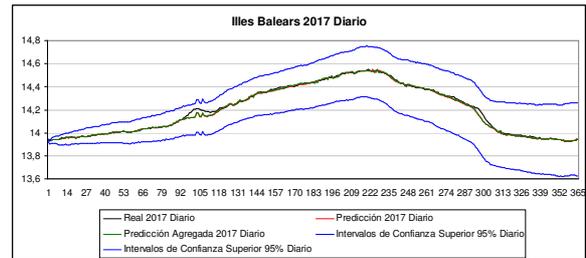


Figura 7: Predicción combinación Diaria para les Illes Balears

7.2. Mallorca

El modelo ARIMAxSARIMA(0,1,0)(2,1,1) es el elegido para Mallorca. La red neuronal utilizada es NNAR(2,1,4)₃₆₅. Y para el modelo PAR, se ha utilizado un VAR(7).

Para Mallorca, y según la tabla2, el modelo que tiene unos errores menores es ANN. Y para la combinación diaria los resultados son mejores, cosa que se puede observar gráficamente en la figura 8 y 9. La combinación trimestral está formada por los modelos BATS, ANN, PAR Y TBATS. La combinación mensual elige el modelo BATS para los tres primeros meses, el modelo ANN para los meses 4,6 y 10, el modelo PAR para los meses 5,7,8, y 9, y para el modelo TBATS los meses 11 y 12. Para la combinación diaria, el modelo ANN tiene una frecuencia absoluta de 110, y relativa de 30.13%. Seguido por el Modelo BATS (frecuencia absoluta de 83 y frecuencia relativa de 22.73%) el modelo PAR (frecuencia absoluta de 79 y frecuencia relativa de 21.64%).

Illes Balears Errores						
	MAPE	MAPE Agregado	MAE	MAE Agregado	MSE	MSE Agregado
ARIMAxSARIMA	0,160612449	0,142768332	0,022912179	0,020356025	0,000945424	0,000809409
BSM	0,150815323	0,140199825	0,021510114	0,019983448	0,000854713	0,000786384
TBSM	0,162225323	0,140243502	0,023133501	0,019990319	0,000931238	0,000786819
BATS	0,17309903	0,205237268	0,024622173	0,029239434	0,001032648	0,001292905
TBATS	0,163986536	0,147665758	0,023380785	0,02104266	0,000892633	0,000760999
ANN	0,103143512	0,10371736	0,014600305	0,014706328	0,000360985	0,00035073
PAR	0,107501208	0,109305462	0,015218027	0,015484381	0,000433264	0,00043086
Trimestral	0,093821994	0,09008478	0,013294245	0,012776435	0,000340181	0,00031278
Mensual	0,080983687	0,076095954	0,011503128	0,010812564	0,000270719	0,000244036
Diario	0,059497429	0,054421845	0,008466389	0,007744038	0,000210181	0,000187125

Tabla1: Errores Illes Balears

Mallorca Errores			
	MAPE	MAE	MSE
ARIMAxSARIMA	0,13269575	0,018514275	0,000543174
BSM	0,135588267	0,018940822	0,000588668
TBSM	0,135097927	0,018869696	0,00058226
BATS	0,163636138	0,022920125	0,000907944
TBATS	0,14858347	0,020788233	0,000677125
ANN	0,110379525	0,015313228	0,000368184
PAR	0,117877277	0,01632633	0,000461156
Trimestral	0,078096452	0,010883085	0,000289996
Mensual	0,063320349	0,008815641	0,000185078
Diario	0,04791723	0,006680906	0,000150375

Tabla2: Errores Mallorca

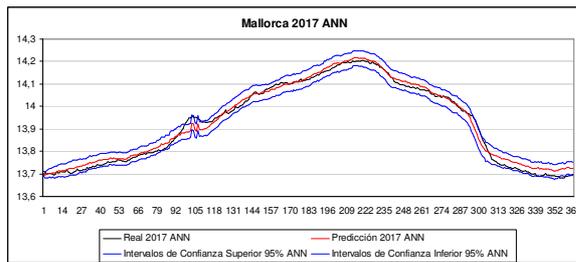


Figura 8: Predicción ANN para les Illes Balears

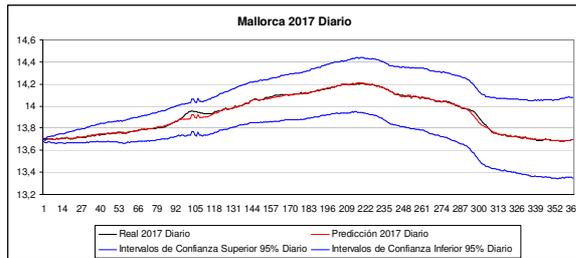


Figura 9: Predicción combinación Diaria para les Illes Balears

7.3 Menorca

El modelo ARIMAxSARIMA(0,1,0)(2,1,1) es el elegido para Menorca. La red neuronal utilizada es NNAR(2,1,2)₃₆₅. Y para el modelo PAR, se ha utilizado un VAR(8).

Si se observan los errores de la tabla 3, se puede ver que para los errores MAPE y MAE el mejor modelo es TBATS, y para el error MSE el mejor modelo es ANN. Si escogemos la combinación diaria el error disminuye. La figura 10, muestra la representación gráfica del mejor modelo que predice Menorca, y la figura 11 es el gráfico de la combinación diaria. La combinación trimestral, selecciona por orden trimestral los siguientes modelos: PAR, ANN, ANN, y TBATS. La combinación mensual está formada por el modelo BATS para el primer mes, el modelo PAR para los meses 2 y 7, el modelo ANN para los meses 3, 4 y 9, y para el modelo TBATS el resto de meses. Y para la combinación diaria, el modelo ANN tiene una frecuencia absoluta y relativa de 115 y 31.5% respectivamente, seguido de por el método TBATS con una frecuencia absoluta de 114, y relativa de 31.23%.

Menorca			
	MAPE	MAE	MSE
ARIMAxSARIMA	0,164721264	0,019421756	0,000715712
BSM	0,166205075	0,01959874	0,000724604
TBSM	0,166143165	0,019577836	0,000715683
BATS	0,164684267	0,019450835	0,000770694
TBATS	0,135258053	0,015961831	0,000516686
ANN	0,136357285	0,016004637	0,000457093
PAR	0,167285666	0,019698625	0,000730192
Trimestral	0,112458326	0,013267709	0,000388125
Mensual	0,095227188	0,011222253	0,000300402
Diario	0,066643261	0,007855638	0,00019918

Tabla3: Errores Menorca

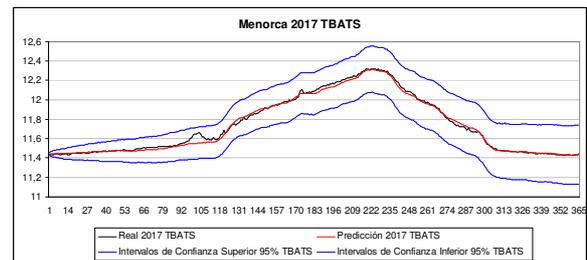


Figura 10: Predicción TBATS para Menorca

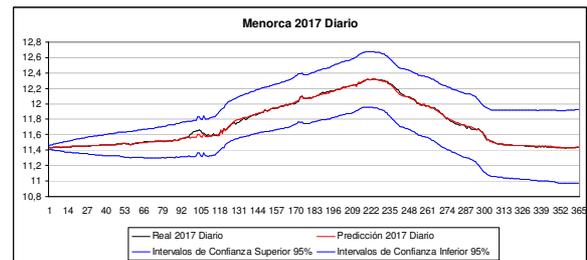


Figura 11: Predicción combinación Diaria para Menorca

7.4. Eivissa y Formentera

El modelo ARIMAxSARIMA(0,1,0)(2,1,1) es el elegido para les Pitiüses. La red neuronal utilizada es NNAR(2,1,2)₃₆₅. Y para el modelo PAR, se ha utilizado un VAR(4).

El modelo que mejor predice a les Pitiüses es el modelo ANN. Pero si coge la combinación diaria, el resultado obtenido es el que tiene un error menor. Las figuras 12 y 13, muestran la representación de los modelos ANN y diario. Para el tercer trimestre de la combinación trimestral el mejor modelo es el TBATS, y para el resto de trimestres el mejor modelo es ANN. La combinación mensual escoge el modelo PAR para los meses 5 y 6, el modelo TBATS para el mes 7, el modelo TBSM para el mes 8, y para el resto de meses el modelo ANN. La frecuencia absoluta es 231, y la relativa de 63% a favor de la red neuronal en la combinación diaria.

Pitiüses			
	MAPE	EA	MSE
ARIMAxSARIMA	0,682248986	0,083642264	0,00916874
BSM	0,573372824	0,070263985	0,006785766
TBSM	0,594120881	0,072805516	0,007205979
BATS	0,72048234	0,088320002	0,009943695
TBATS	0,474077294	0,05810337	0,004630793
ANN	0,226372669	0,028089078	0,001282755
PAR	0,339902118	0,041977758	0,002528582
Trimestral	0,197315733	0,02437197	0,001034763
Mensual	0,181763202	0,022406909	0,00093265
Diario	0,155910167	0,019153154	0,000807933

Tabla4: Errores de les Pitiüses

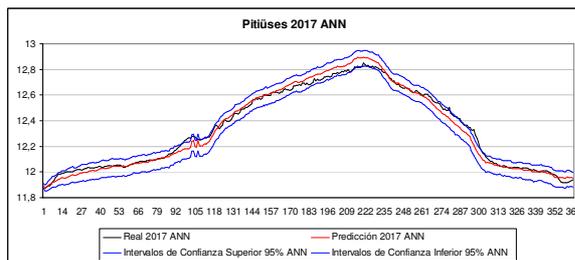


Figura 12: Predicción ANN para les Pitiüses

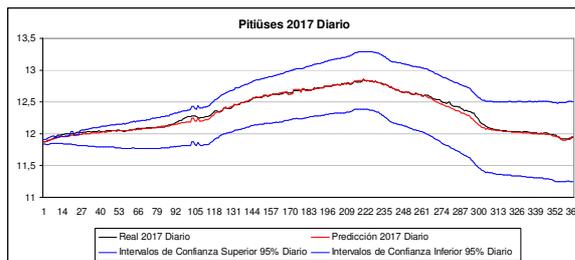


Figura 13: Predicción combinación Diaria para les Pitiüses

8. Predicción hasta 2030

La idea de predecir a largo plazo es un ejercicio. Hay que decir que estos datos predichos son el resultado de haber entrenado un modelo con los datos que van del año 1997 al 2016 (7305 datos diarios), y prediciendo hasta el año 2030 (5113 datos diarios). Y los resultados que se obtendrán posiblemente cambien en un futuro, ya que no se puede saber que ocurrirá.

El modelo que ha hecho una mejor predicción para el año 2017, cosa que se ha visto en el apartado anterior, es el que se ha elegido para hacer esta predicción a largo plazo. Para hacer esta predicción se ha utilizado para les Illes Balears, Mallorca, Menorca y les Pitiüses el modelo ANN, y la red neuronal para cada base de datos ha sido la misma que se ha utilizado para predecir el año 2017. Para Menorca se descarta el modelo TBATS, porque a largo plazo el modelo BATS y TBATS mostraban que los valores del IPH se mantenían

constantes durante los años. La tendencia era plana y los valores máximos obtenidos eran exactamente iguales a los del año anterior con una diferencia que entre -0.15% y 0.15% . Para el resto de modelos la diferencia de crecimiento de un año respecto al anterior se encuentra entre el 1% y el 3% . No se ha elegido ninguna de las combinaciones, ya que a largo plazo no funcionan bien.

En las figuras 14, 15, 16 y 17 son las representaciones gráficas de las predicciones hechas para cada una de las bases de datos hasta el año 2030. En todas las series se puede ver de color rojo la parte predicha que van de los años 1997-2017, y de color rojo la parte de los valores que van del año 1997 al 2017. Aparece la serie diaria del año 2017 real y predicha para comparar como se ajustan. En común, las cuatro series tienen una tendencia creciente, y así como van pasando los años el dato diario superior de un año supera el dato superior del año anterior.

8.1. Illes Balears 2030

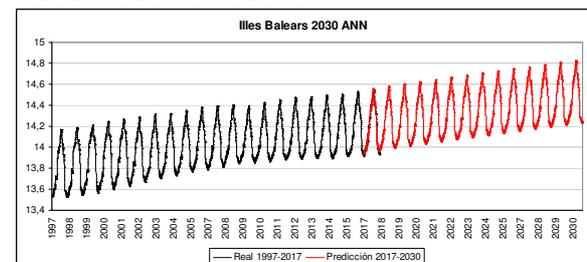


Figura 14: Predicción ANN Illes Balears 2030

8.2 Mallorca

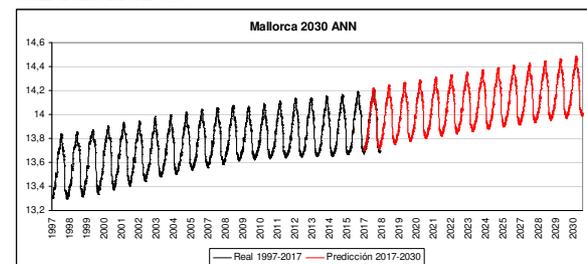


Figura 15: Predicción ANN Mallorca 2030

8.3 Menorca

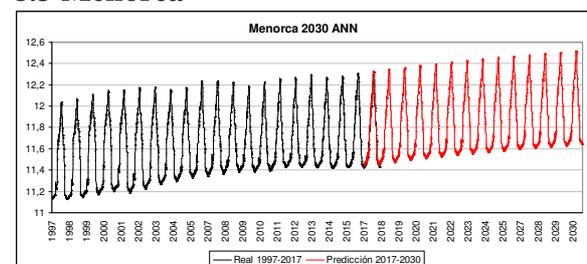


Figura 16: Predicción ANN Menorca 2030

8.4. Eivissa- Formentera

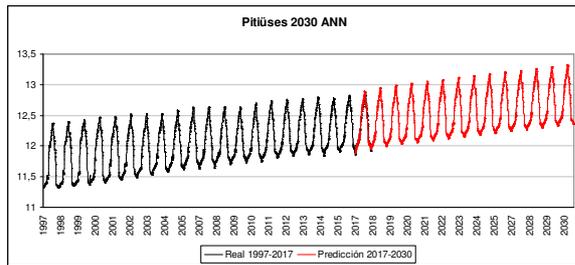


Figura 17: Predicción ANN Pitiüses 2030

9. Efecto de las predicciones sobre el territorio.

Las predicciones que se han obtenido, muestran que así como vayan pasando los años aumenta el valor del IPH. Pero la superficie del terreno es fija, a no ser que ocurran algún suceso futuro que modifique el territorio. En las siguientes tablas se han obtenido los valores máximos y mínimos que alcanzara el IPH, y las densidades máximas y mínimas por superficie²⁵, por línea de costa y por línea de playa²⁶.

En las siguientes tablas se puede ver los resultados para cuatro años. Para el año 2030 en las Islas Baleares se estima que el IPH alcanzará un máximo cercano a 2.7 millones de personas, Mallorca alcanzará 1.9 millones, Menorca alcanzará 273 mil personas y les Pitiüses 611 mil personas. El porcentaje de no residentes se mantiene y se mantendrá constante durante los años que van del 1997 al 2030. Por lo que su porcentaje es parecido: para les Illes Balears es de un 44,94%, para Mallorca un 38,83%, para Menorca un 58,47% y para les Pitiüses es de 62,07%. Estas dos últimas islas seguirán teniendo días durante el año, donde su población residente se verá superada en número por la población no residente. La densidad máxima está correlacionada positivamente con el valor del IPH, por lo que habría más personas por kilómetro cuadrado. Pero si tenemos en cuenta que la población no residente suele hospedarse cerca de la costa²⁷, la densidad máxima que se alcanzaría en la costa sería mucho mayor. Por ejemplo en Mallorca durante el año 2030 tendría una densidad de superficie 535, y la densidad en línea de costa sería de 2666 personas por kilómetro de costa, esto supone una diferencia de un 398%. La mayoría de la población no residente suele acudir a les Illes Balears para estar en las playas²⁸, y también

²⁵ Superficie de les Illes Balears, obtenidos de Ibestat[6]

²⁶ Algunos datos útiles sobre líneas de costas en las Illes Balears, obtenidos de Socib[16]

²⁷ No se incluyen los islotes, por estar deshabitados.

²⁸ Incluye las playas de arena, y las playas de gravas y guijarros

muchos residentes acuden a las diferentes playas. Así que en un foco importante de población durante la mayor parte del día en verano. En Mallorca para el año 2030, se podría llegar a tener unas 28 personas por metro de playa, en Menorca la cifra es inferior ya que habría unas 6 personas, y en les Pitiüses alcanzaría las 13 personas por metro de playa.

Illes Balears ANN Máximos y Mínimos					
Año	IPH Máximo	Residentes	Porcentaje de no Residentes	Densidad Máxima (Personas/Km2)	Densidad Mínima (Personas/Km2)
2017	2102563	1123564	46,56217364	421,7964565	225,398858
2020	2247193	1233191	45,12307642	450,8108712	247,3911373
2025	2489789	1365929	45,13874403	499,4780588	274,0199363
2030	2744834	1511213	44,94336447	550,6428437	303,1654235

Tabla5: Illes Balears Máximos Y Mínimos

Illes Balears ANN Máximos y Mínimos de las Líneas de Costa y Playa				
Año	Densidad Máxima por Línea de Costa (Personas/Km Costa)	Densidad Mínima por Línea de Costa (Personas/Km Costa)	Densidad Máxima por Línea de Playa (Personas/Km Playa)	Densidad Mínima por Línea de Playa (Personas/Km Playa)
2017	1454,55745	777,2838842	13132,80945	7017,887911
2020	1554,613134	853,1238615	14036,18542	7702,626744
2025	1722,44105	944,9527933	15551,4587	8531,725564
2030	1898,881885	1045,460479	17144,49572	9439,182522

Tabla6: Illes Balears Máximos y Mínimos de las líneas de Costa y Playa

Mallorca ANN Máximos y Mínimos					
Año	IPH Máximo	Residentes	Porcentaje de no Residentes	Densidad Máxima (Personas/Km2)	Densidad Mínima (Personas/Km2)
2017	1495976	886954	40,71070939	411,4681542	243,9565497
2020	1604604	978442	39,02285747	441,346215	269,1204351
2025	1771589	1083270	38,85320381	487,2753876	297,9533725
2030	1945113	1189744	38,83419533	535,0032567	327,2393172

Tabla7: Mallorca Máximos Y Mínimos

Mallorca ANN Máximos y Mínimos de las Líneas de Costa y Playa				
Año	Densidad Máxima por Línea de Costa (Personas/Km Costa)	Densidad Mínima por Línea de Costa (Personas/Km Costa)	Densidad Máxima por Línea de Playa (Personas/Km Playa)	Densidad Mínima por Línea de Playa (Personas/Km Playa)
2017	2050,405762	1215,671031	21649,43623	12835,79716
2020	2199,292492	1341,066338	23221,47326	14159,7974
2025	2428,16425	1484,745066	25638,04106	15676,84515
2030	2665,999176	1630,681195	28149,24745	17217,72793

Tabla8: Mallorca Máximos y Mínimos de las líneas de Costa y Playa

Menorca ANN Máximos y Mínimos					
Año	IPH Máximo	Residentes	Porcentaje de no Residentes	Densidad Máxima (Personas/Km2)	Densidad Mínima (Personas/Km2)
2017	225489	92837	58,82875684	324,5599347	133,6253599
2020	237887	98762	58,48379285	342,4062369	142,1539749
2025	257025	106505	58,56244904	369,9512503	153,2989317
2030	273000	113375	58,47068825	392,9450106	163,1873281

Tabla9: Menorca Máximos Y Mínimos

Menorca ANN Máximos y Mínimos de las Líneas de Costa y Playa				
Año	Densidad Máxima por Línea de Costa (Personas/Km Costa)	Densidad Mínima por Línea de Costa (Personas/Km Costa)	Densidad Máxima por Línea de Playa (Personas/Km Playa)	Densidad Mínima por Línea de Playa (Personas/Km Playa)
2017	621,1824118	255,7485212	5067,173382	2086,218274
2020	655,338843	272,0716253	5345,797753	2219,370787
2025	708,0578512	293,4022039	5775,842697	2393,370787
2030	752,0661157	312,3278237	6134,831461	2547,752809

Tabla10: Menorca Máximos y Mínimos de las líneas de Costa y Playa

Pitiüses ANN Máximos y Mínimos					
Año	IPH Máximo	Residentes	Porcentaje de no Residentes	Densidad Máxima (Personas/Km2)	Densidad Mínima (Personas/Km2)
2017	398040	143948	63,83586597	608,3229566	219,9947294
2020	454506	170221	62,54811656	694,6193918	260,1479573
2025	529714	198357	62,55396949	809,5594261	303,1480744
2030	611643	232020	62,06604008	934,7726615	354,596608

Tabla11: Pitiüses Máximos Y Mínimos

Pitiüses ANN Máximos y Mínimos de las Líneas de Costa y Playa				
Año	Densidad Máxima por Línea de Costa (Personas/Km Costa)	Densidad Mínima por Línea de Costa (Personas/Km Costa)	Densidad Máxima por Línea de Playa (Personas/Km Playa)	Densidad Mínima por Línea de Playa (Personas/Km Playa)
2017	1127,912124	407,8996523	8560,004056	3095,65134
2020	1287,917257	482,3491074	9774,322581	3660,666667
2025	1501,031454	562,0770757	11391,69892	4265,741935
2030	1733,193539	657,4695381	13153,63441	4989,698925

Tabla12: Pitiüses Máximos y Mínimos de las líneas de Costa y Playa

10. Conclusión

En este trabajo se ha creado un modelo que permite hacer predicciones del IPH de les Illes Balears, Mallorca, Menorca, y les Pitiüses. Este modelo está formado por dos partes: una parte que recoge la parte determinista²⁹ del IPH, y la segunda parte consiste en aplicarle al término de perturbación diferentes modelos econométricos, para poder ver que modelo predice mejor. Para el modelo PAR se utiliza el mismo modelo pero sin las variables ficticias días de la semana. Otro modelo que se ha utilizado, a parte del anterior, son los modelos BATS y TBATS, los cuales tienen un sistema de funciones propio.

Los resultados obtenidos han mostrado que los errores entre los diferentes modelos son parecidos. Pero el mejor modelo, ha sido la red neuronal artificial para les Illes Balears, Mallorca y les Pitiüses, y el modelo TBATS para Menorca. Las combinaciones mensuales, trimestrales y diarias de los modelos dan unos mejores resultados, por lo que se podría optar por esta combinación a corto plazo.

A la hora de querer hacer una predicción hasta el año 2030, muestra una tendencia creciente, causada por los datos utilizados para entrenar el modelo. Cualquier hecho futuro podría modificar los resultados a largo plazo.

Los efectos de las predicciones sobre el territorio asumen que el valor del IPH máximo y mínimo que se alcanzará cada año es superior al de años anteriores. El porcentaje de no residentes muestra que se mantendrá constante. La densidad de población por superficie y por línea de costa y playa, también aumenta a causa de aumentar el IPH (siendo el territorio un valor fijo).

²⁹ Ecuación 2, sin el término de perturbación.

11. Referencias

- [1] CAIB. Govern de les Illes Balears (2018): http://www.caib.es/sites/calendarilaboral/ca/an_y_2018/
- [2] Chan, N.A (2004): "Time Series Applications to Finance". John Wiley & Sons, Inc., Publication.
- [3] De Livera, A.M., Hyndman, R.J. and Snyder, R.D. (2011): "Forecasting Time Series with Complex Seasonal Patterns Using Exponential Smoothing". Journal of the American Statistical Association, 106, 1513-1527.
- [4] IBESTAT. (2018): "Indicador de Pressió Humana³⁰".
- [5] IBESTAT. (2018): "Resum Demogràfic³¹".
- [6] IBESTAT. (2018): "Superfície de les Illes Balears³²".
- [7] Franses, P.H. (1996): "Periodicity and Stochastic trends in economic time series". Oxford University Press, Oxford.
- [8] Franses, P.H y Van Dijk, D. (2003). "Non-linear Time Series Models in Empirical Finance". Cambridge University Press
- [9] Franses, P.H. y Paap, R. (2004): "Periodic Time Series Models". Oxford University Press, Oxford.
- [10] Ghysels, E. y Osbron, D.R. (2001): "The Econometric Analysis of Seasonal Time Series". Cambridge University Press, Cambridge.
- [11] Gladyshev, E.G. (1961): "Periodically Correlated Random Sequences". Sov.Math2, 385-388.
- [12] Haldrup, N., Hylleberg, S., Pons, G. y Sansó, A. (2007): "Common Periodic Correlation Features and the Interaction of Stocks and Flows in Daily Airport Data". Economics Working Papers 2005-03, Department of Economics and Business Economics, Aarhus University.
- [13] Harvey, A.C. (1989): "Forecasting, Structural Time Series and the Kalman Filter", Cambridge University Press.
- [14] Sansó, A. (2004): "Anàlisi univariant de Sèries temporals: Procesos estocàstics". ResearchGate
- [15] Sansó, A. "Projeccions de la càrrega demogràfica diària de les Illes Balears per a

³⁰ <https://ibestat.caib.es/ibestat/estadistiques/poblacio/estudis-demografics/indicador-pressio-humana-iph/e91ffb58-6bdd-457c-bd25-ed2a201f57ae>

³¹ <https://ibestat.caib.es/ibestat/estadistiques/poblacio/resum-demografic/definitius-resum-demografic/c29176ea-c2c4-4d65-98c1-23d9c47df46a>

³² <https://ibestat.caib.es/ibestat/estadistiques/entorn-fisic/territori/3f61e929-71c5-41fa-93ae-329e42a63ef0>

diferents horitzons temporals”. Fundació Bit, 2016.

- [16] SOCIB (2016): “Algunos datos útiles sobre líneas de costa en las Illes Balears”. Nota Interna.
- [17] Teräsvirta, T. Van Dijk, D. y Medeiros, M.C. (2005) “Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination”. *International Journal of Forecasting* 21 (2005) 755– 774
- [18] Tiao, G.C. y Grupe, M.R. (1980): “Hidden Periodic Autoregressive-Moving Average in Time Series Data”. *Biometrika* 67, 365-373.
- [19] Wei, William, W.S. (2006): “Time series analysis univariate and multivariate methods”. Second Edition. Pearson
- [20] West, M. & Harrison, J. (1997): “Bayesian forecasting and dynamic models”, Second Edition. Springer-Verlag, New York.