Time series features and machine learning forecasts

Oscar Claveria[1][*], Enric Monte[2], Salvador Torra[3]

[1] AQR-IREA, University of Barcelona (UB)
[2] Department of Signal Theory and Communications, Polytechnic University of Catalunya (UPC)
[3] Riskcenter-IREA, Department of Econometrics, Statistics and Applied Economics, University of Barcelona (UB)

Abstract. In this study we combine the results of two independent analyses to position Spanish regions according to both the characteristics of the time series of international tourist arrivals and the accuracy of predictions of arrivals at the regional level. We apply a seasonal-trend decomposition procedure based on non-parametric regression to isolate the different components of the series and calculate the main time series features. Predictions are generated with several machine learning models in a recursive multi-step-ahead forecasting experiment. Finally, we summarize all the information from the two previous experiments using categorical principal component analysis. By overlapping the distribution of the regions and the component loadings of each variable along both dimensions, we observe that entropy and dispersion show an inverse relation with forecast accuracy, but the interactions between the rest of the features and accuracy are heavily dependent on the forecast horizon. On this evidence, we conclude that in order to increase forecast accuracy of tourist arrivals at the regional level, model selection should be region-specific and based on the forecast horizon.

Keywords: tourist arrivals; time series features; machine learning; forecast accuracy; categorical principal components analysis

[*] **Corresponding Author:**
Oscar Claveria, University of Barcelona, 08034 Barcelona, Spain. Email: oclaveria@ub.edu

Introduction

The increasing weight of the tourism industry in the gross domestic product (GDP) of most countries explains the growing interest in the sector from economic circles. Bilen, Yilanci, and Eryüzlü (2015) and Sharif, Saha, and Loganathan (2017) find evidence of the relevance of tourism in economic growth. In this context, machine learning (ML) methods are being increasingly applied to anticipate the number of arrivals at the destination level. The most commonly used ML techniques for prediction are neural network (NN) and support vector regression (SVR) models, which can be regarded as an extension of the support vector machine (SVM) mechanism. In recent years, several authors find evidence in favour of SVRs and NNs when compared to more traditional time series models (Akin, 2015; Chen & Wang, 2007; Claveria & Torra, 2014; Hong, Dong, Chen & Wei, 2011).

The search for more accurate forecasts of tourist arrivals has led to an extensive body of research, but in most of these studies time series features are overlooked (Kim & Schwartz, 2013). The main objective of this work is to fill this gap. With this aim, we combine the output of two independent analysis. First, we compute the time series features of international tourist arrivals to the different regions of Spain. In order to do so, we isolate the different components of the series by means of a seasonal-trend decomposition procedure based on non-parametric regression. Second, we compute the out-of-sample forecast accuracy of the predictions of tourist arrivals to every region of Spain in a multi-step-ahead forecasting experiment. We use several ML methods: SVR and NN models, and a Gaussian process regression (GPR) model. GPR is a statistical learning technique originally devised for spatial interpolation, which has recently been used for time series forecasting (Ben Taieb, Sorjamaa, & Bontempi, 2010). By means of dimensionaltiy reduction techniques we synthesize the results of the two previous experiments in two dimensions, along which we project the distribution of Spanish regions.

Spain is one of the world's top tourist destinations after France and the United States. According to the Spanish Statistical Office (https://www.ine.es/), the country received 82 million interantional tourist arrivals in 2017, which represented an 8.9% increase in relation to the previous year. The total expenditure incurred by international tourists increased 12.4% compared to 2016. The average expenditure per tourist stood at 1,061 euros, with an annual increase of more than 3%. Tourism accounts for almost 12% of GDP and provides employment for 13% of the working population. The main source markets are the United Kingdom, France, Germany and Italy.

Spain has seventeen regions. Most tourist arrivals are concentrated in the Mediterranean coast and the islands. Catalonia, Andalusia, Madrid, and the Balearic and the Canary Islands are the regions that receive the higher number of tourists, which account for 85% of total international tourist demand to Spain. The fact that regional markets present marked differences and have evolved in very diverse forms, has led us to conduct the analysis at a regional level.

Methodology – ML models

In this study we use different ML techniques: GPR, linear and polynomial SVR, and MLP and RBF NN architectures. We denote a time series of length N as $\{x^0, x^1, x^2, \cdots, x^N\}$. We refer to the training set as $D^{TRN} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$, which consists of a set of tuples (x_t, y_t), where y_t is the target value and x_t the observation vector at time t, constructed from a segment of the time series consisting of the past p samples $x_t = [x_t \quad x_{t-1} \quad \cdots \quad x_{t-p-1} \quad x_{t-p}]^T$. Similarly, we refer to the test database as $D^{TST} = \{(x_1^{tst}, y_1^{tst}), (x_2^{tst}, y_2^{tst}), \ldots, (x_N^{tst}, y_N^{tst})\}$. The set of tuples is assumed to be drawn from the following process:

$$y_t = f(x_t) + \varepsilon \qquad \text{with } \varepsilon \sim N(0, \sigma^2) \qquad (1)$$

GPR is as a method of interpolation based on the assumption that the inputs have a joint multivariate Gaussian distribution characterized by an analytical model of the structure of the covariance matrix (Rasmussen & Williams, 2006). The model assumes a kernel $k(x_i, x_j)$ that gives a similarity measure between two given vectors x_i, x_j. The covariance matrix is denoted as $K(X, X)$. We also define a vector of similarity between the test vector x_t^{tst} and the rest of elements of D^{TRN} as $K(x_t^{tst}, X) = [k(x_t^{tst}, x_1) \quad \cdots \quad k(x_t^{tst}, x_N)]$.

The joint distribution of the variables is the conditional Gaussian distribution, parametrized by $K(x_t^{tst}, X)$ and $K(X, X)$. Thus, as the forecast value we use the mean of the ex-post distribution:

$$y_t^{tst} = f(x_t^{tst}) + \varepsilon = \mu = K(x_t^{tst}, X)[K(X, X) + \sigma^2 I]^{-1} y \qquad (2)$$

If we define a vector $\{\beta_0, \beta_1, \cdots, \beta_N\}$ as $\beta = [K(X, X) + \sigma^2 I]^{-1} y$, expression (2) can be rewritten as:

$$y_t^{tst} = f(x_t^{tst}) + \varepsilon = \beta_0 + \sum_{j=1}^{N} \beta_j k(x_t^{tst}, x_j) \tag{3}$$

The functional form of the covariance matrix that we use in this experiment is a radial basis kernel with a linear trend to account for the trend component present in most of the time series over the training period. Alternative sets of kernels are discussed in MacKay (2003). See Claveria, Monte, and Torra (2016) and Wu, Law, and Xu (2012) for a detailed description of the model used in this study.

The SVR mechanism can be regarded as an extension of SVMs to construct data-driven and nonlinear regressions (Drucker, Burges, Kaufman, Smola, & Vapnik, 1997). With a SVR we can define an approximation of the regression function within a tube generated by means of a set of support vectors v_i that belong to D^{TRN}. This is done by introducing restrictions on the structure or curvature of the set of functions over which the estimation is done (Vapnik, 1998; Schölkopf & Smola, 2002). The model consists of a linear combination of the outputs of a kernel $K(v_i, x_t)$:

$$y_t = f(x_t) + \varepsilon = \beta_0 + \sum_{j=1}^{q} \beta_j K(v_i, x_t) \tag{4}$$

The weights $\{\beta_j, \beta_0\}$ combine the outputs of the kernel function and are estimated in D^{TRN}. The kernel gives a measure of similarity between points in the feature space. In this experiment we use both a linear and a polynomial kernel for the estimation of the regression. See Chen and Wang (2007) and Hong, Dong, Chen, and Wei (2011) for a detailed description of the model.

NNs are used to identify related temporal patterns. In this study we use two NN models: MLP and RBF. The MLP architecture is a feed-forward NN based on supervised learning. The topology consists of a hidden layer with q units, followed by a linear combination of the outputs of the hidden layer plus a bias term:

$$y_t = f(x_t) + \varepsilon = \beta_0 + \sum_{j=1}^{q} \beta_j g\left(\sum_{i=1}^{p} w_t^T x_t + w_{0j} \right)$$
$$g(.) = \frac{1}{1 + e^{-(.)}} \tag{5}$$

Where g is the non-linear function of the units in the hidden layer. The number of neurons is estimated by cross-validation. The weights β_j correspond to the connections of the hidden units j to the output units. The parameters are denoted as $w_t^T = [w_{1j} \quad w_{2j} \quad \cdots \quad w_{pj}]^T$, and are estimated in D^{TRN} using the Levenberg-Marquardt

algorithm. The training is done by iteratively estimating the value of the parameters by local improvements of the cost function.

RBF NNs are hybrid networks that combine both supervised and non-supervised learning. The hidden layer is formed by a set of radial functions centred each at a centroid. The model consists on a linear combination of the outputs of q Gaussians plus an offset:

$$y_t = f(x_t) + \varepsilon = \beta_0 + \sum_{j=1}^{q} \beta_j g_j(x_t) + \varepsilon$$

$$g_j(x_t) = \exp\left(-\sum_{j=1}^{q} \left\|(x_t - \mu_j)\right\|^2 \Big/ 2\sigma_j^2\right)$$

(6)

Where μ_j is the centroid vector for unit j. For each centroid, the variance σ_j^2 indicates the width over the input space of the Gaussian function. The number of centroids and the spread of each centroid are determined by cross-validation in D^{TRN}. See Claveria, Monte, and Torra (2015; 2017) for a detailed description of the NN models applied in this study.

Results

In this section we first compute the time series features of international tourist arrivals to each region of Spain. Data is obtained from the Spanish Statistical Office, and include the monthly number of arrivals at a regional level over the time period 1999:01 to 2014:01. We apply the "Seasonal and Trend Decomposition using Loess" (STL) procedure proposed by Cleveland, Cleveland, McRae, and Terpenning (1990) to isolate the seasonal and trend components of the time series. STL presents several advantages over the classical decomposition method and X-12-ARIMA (Bergmeir, Hyndman, & Benítez, 2016).

Given that $Y_t = T_t + S_t + R_t$, where T_t denotes the trend component, S_t the seasonal component and R_t the residual, we can assess the strength of each component by computing two measures introduced by Wang, Smith-Miles, and Hyndman (2009):

$$\text{strength of trend} = 1 - \frac{V(R_t)}{V(Y_t - S_t)}$$

(7)

$$\text{strength of seasonality} = 1 - \frac{V(R_t)}{V(Y_t - T_t)}$$

(8)

Additionally, for each time series we compute the skewness, the kurtosis, the coefficient of variation, the first autocorrelation, and two additional statistics: the Box-Cox lambda (λ_1) and spectral entropy (H).

The lambda value indicates the power to which all data should be raised in order to be normally distributed (Box & Cox, 1964). The Box-Cox power transformation searches within the interval [-5;5] until the optimal value is found in:

$$Y_t^{(\lambda)} = \begin{cases} \dfrac{(Y_t + \lambda_2)^{\lambda_1} - 1}{\lambda_1} & \text{if } \lambda_1 \neq 0 \\ \log(Y_t + \lambda_2) & \text{if } \lambda_1 = 0 \end{cases} \tag{9}$$

Entropy describes the complexity of a system and can be used as a proxy for the predictability of a given time series (Goerg, 2013). H can be obtained as:

$$H = -\int_{-\pi}^{\pi} f_y(\omega) \log f_y(\omega) d\omega \tag{10}$$

Where $f_y(\omega)$ is the spectral density of Y_t. Low values of H are indicative of more signal, suggesting that a time series is easier to forecast (Kang, Hyndman, & Smith-Miles, 2017). Figure 1 contains the reuslts of the descriptive analysis of the data in a bar chart for each feature. We can observe that the time series features vary across regions. The Canary Islands present the highest level of strength of trend, but the lowest in terms of the strength of seasonality, Box-Cox lambda, and spectral entropy. On the other hand, the Balearic Islands present the highest values of entropy and dispersion.
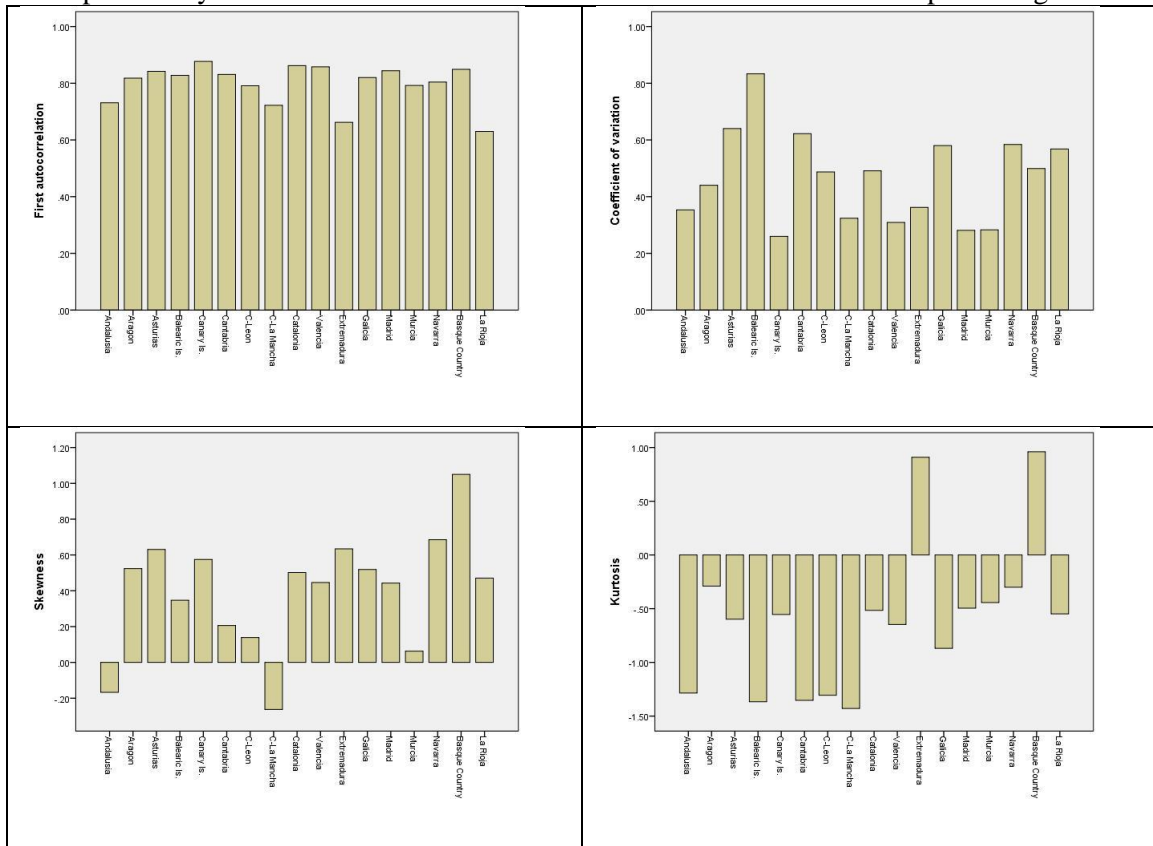
Figure 1. Time series features of international tourist arrivals to Spanish regions
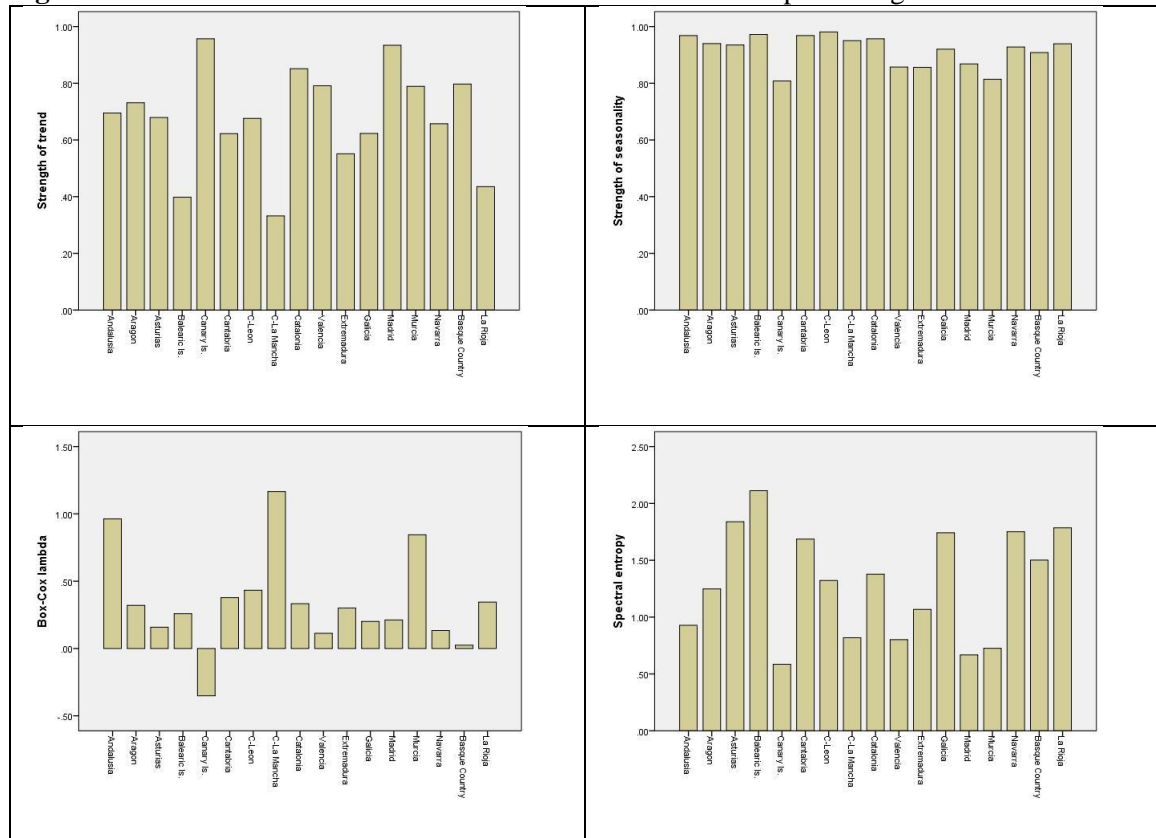
Note. C-Leon stands for Castilla-Leon and C-La Mancha stands for Castilla-La Mancha.

Figure 1. (cont.)
Descriptive analysis – Time series features of international tourist arrivals to Spanish regions

Note. C-Leon stands for Castilla-Leon and C-La Mancha stands for Castilla-La Mancha.

Next, we compute the out-of-sample forecast accuracy of the predictions generated with the different ML models presented in Section 2. The first 52% of the observations were used as the initial training set, the next 33% as the validation set, and and the last 15% as the test set. This partition is done sequentially. The estimation of the models is done recursively for different forecast horizons (1, 2, 3, 6 and 12 months) during the out-of-sample period (2013:01-2014:01). All models are implemented using Python. Forecast accuracy is measured by means of the Mean Absolute Percentage Error (MAPE). Results are presented in Table 1.

Table 2 contains the number of regions for which we obtain the lowest MAPE with each model. We find that MLP NNs outperform the rest models in most regions, and that regardless of the ML method, forecast accuracy improves for longer horizons.

Table 1
Out-of-sample forecast accuracy – MAPE (2013:01-2014:01)

Forecast horizon	Support Vector Regression models		Gaussian Proces Regression	Neural Network models	
	SVR linear	SVR poly	GPR	RBF	MLP
Andalusia					
h=1	0.34	0.45	0.36	0.39	0.36
h=2	0.47	0.40	0.45	0.44	0.44
h=3	0.45	0.46	0.42	0.45	0.42
h=6	0.33	0.40	0.38	0.40	0.35
h=12	0.13*	0.26	0.14	0.28	0.17
Aragon					
h=1	0.39	0.33	0.36	0.30	0.30
h=2	0.44	0.43	0.41	0.37	0.36
h=3	0.37	0.35	0.37	0.37	0.39
h=6	0.29	0.32	0.27	0.30	0.22*
h=12	0.30	0.30	0.29	0.27	0.24
Asturias					
h=1	0.78	0.88	0.75	0.67	0.61
h=2	0.99	0.94	0.91	0.75	0.80
h=3	0.65	0.68	0.75	0.73	0.78
h=6	0.40	0.59	0.41	0.46	0.34
h=12	0.28*	0.37	0.29	0.42	0.32
Balearic Islands					
h=1	5.46	8.16	5.18	4.35	3.98
h=2	7.19	9.46	6.72	5.72	5.68
h=3	5.43	4.63	5.99	6.05	5.66
h=6	0.91*	2.59	0.96	3.14	1.31
h=12	1.60	2.88	1.66	2.62	1.69
Canary Islands					
h=1	0.46	0.54	0.46	0.42	0.42
h=2	0.43	0.45	0.44	0.41	0.42
h=3	0.44	0.47	0.44	0.41	0.42
h=6	0.41	0.42	0.42	0.41	0.41
h=12	0.47	0.52	0.45	0.40*	0.43
Cantabria					
h=1	1.13	1.39	1.07	0.95	1.03
h=2	1.38	1.70	1.19	1.00	1.03
h=3	1.13	0.96	1.03	1.05	1.05
h=6	0.33	0.60	0.37	0.51	0.38
h=12	0.29*	0.50	0.33	0.57	0.32
Castilla-Leon					
h=1	0.60	0.74	0.61	0.59	0.52
h=2	0.76	0.82	0.70	0.65	0.66
h=3	0.70	0.49	0.66	0.66	0.66
h=6	0.31	0.47	0.31	0.39	0.26
h=12	0.17*	0.35	0.18	0.36	0.23
Castilla-La Mancha					
h=1	0.32	0.39	0.33	0.31	0.28
h=2	0.41	0.43	0.37	0.36	0.36
h=3	0.41	0.51	0.37	0.38	0.38
h=6	0.26	0.32	0.27	0.28	0.23
h=12	0.11	0.22	0.10*	0.21	0.14
Catalonia					
h=1	0.42	0.46	0.39	0.37	0.38
h=2	0.51	0.51	0.47	0.42	0.44
h=3	0.45	0.38	0.42	0.42	0.42
h=6	0.40	0.37	0.32	0.31	0.30
h=12	0.33	0.41	0.32	0.30	0.26*

Note. * Model with the lowest MAPE

Table 1 (cont.)
Out-of-sample forecast accuracy – MAPE (2013:01-2014:01)

| Forecast horizon	Support Vector Regression models		Gaussian Proces Regression	Neural Network models	
	SVR linear	SVR poly	GPR	RBF	MLP
Valencia					
h=1	0.30	0.49	0.30	0.31	0.27
h=2	0.36	0.33	0.36	0.33	0.33
h=3	0.35	0.34	0.37	0.34	0.36
h=6	0.34	0.28	0.34	0.34	0.35
h=12	0.29	0.52	0.27	0.30	0.25*
Extremadura					
h=1	0.32	0.38	0.33	0.37	0.33
h=2	0.41	0.40	0.39	0.40	0.39
h=3	0.38	0.42	0.36	0.39	0.37
h=6	0.35	0.40	0.39	0.37	0.33
h=12	0.19*	0.36	0.20	0.30	0.21
Galicia					
h=1	0.86	1.01	0.85	0.76	0.72
h=2	1.07	1.07	0.98	0.87	0.89
h=3	0.82	1.05	0.88	0.90	0.92
h=6	0.53	0.75	0.55	0.63	0.46
h=12	0.34*	0.90	0.35	0.52	0.39
Madrid (Community)					
h=1	0.27	0.35	0.26	0.26	0.24
h=2	0.26	0.29	0.25	0.25	0.24
h=3	0.26	0.19	0.25	0.25	0.24
h=6	0.26	0.26	0.25	0.25	0.23*
h=12	0.24	0.42	0.24	0.25	0.25
Murcia (Region)					
h=1	0.24	0.23	0.26	0.22	0.23
h=2	0.29	0.28	0.32	0.25	0.27
h=3	0.26	0.23	0.34	0.28	0.30
h=6	0.25	0.21	0.29	0.27	0.31
h=12	0.20	0.26	0.19	0.20	0.17*
Navarra					
h=1	0.59	0.67	0.59	0.57	0.53
h=2	0.72	0.80	0.71	0.62	0.65
h=3	0.65	0.61	0.63	0.62	0.63
h=6	0.38	0.55	0.33	0.47	0.33
h=12	0.31	0.45	0.31	0.40	0.30*
Basque Country					
h=1	0.41	0.39	0.41	0.37	0.44
h=2	0.46	0.43	0.44	0.39	0.41
h=3	0.39	0.38	0.39	0.39	0.40
h=6	0.42	0.36	0.35	0.33*	0.34
h=12	0.39	0.43	0.38	0.36	0.35
La Rioja					
h=1	0.56	0.73	0.59	0.57	0.53
h=2	0.77	0.62	0.71	0.67	0.68
h=3	0.77	0.65	0.70	0.70	0.69
h=6	0.46	0.64	0.48	0.45	0.30
h=12	0.19	0.39	0.19*	0.39	0.23

Note. * Model with the lowest MAPE

Table 2
Number of regions with lowest MAPE by model and forecast horizon

Forecast horizon	Support Vector Regression models		Gaussian Proces Regression	Neural Network models	
	SVR linear	SVR polynomial	GPR	RBF	MLP
h=1	2	0	0	5	10
h=2	0	3	0	9	5
h=3	2	11	3	1	0
h=6	3	1	0	2	10
h=12	8	0	2	1	6

Due to the observed cross-regional differences, we finally project the Spanish regions according to the characteristics of tourist arrivals time series and to the forecast accuracy obtained in each region for each forecast horizon. To that end, we combine the results of both the descriptive analysis and the forecasting experiment, and summarize all the information in two components using the multivariate positioning approach proposed by Claveria (2016; 2017). First, we rank each region according to its time series features and the obtained MAPE values in increasing order. These rankings are then used as the input for a categorical principal components analysis (CATPCA).

Along both dimensions, the biplot in Figure 2 overlaps the object scores (regions) and the rotated component loadings (Table 3). The coordinates of the end point of each vector are given by the loadings of each variable on the two components. Long vectors are indicative of a good fit. The variables that are close together in the plot, are positively related; the variables with vectors that make approximately a 180º angle with each other, are closely and negatively related; finally, non-related variables correspond with vectors making a 90º angle.

Figure 2. Biplot
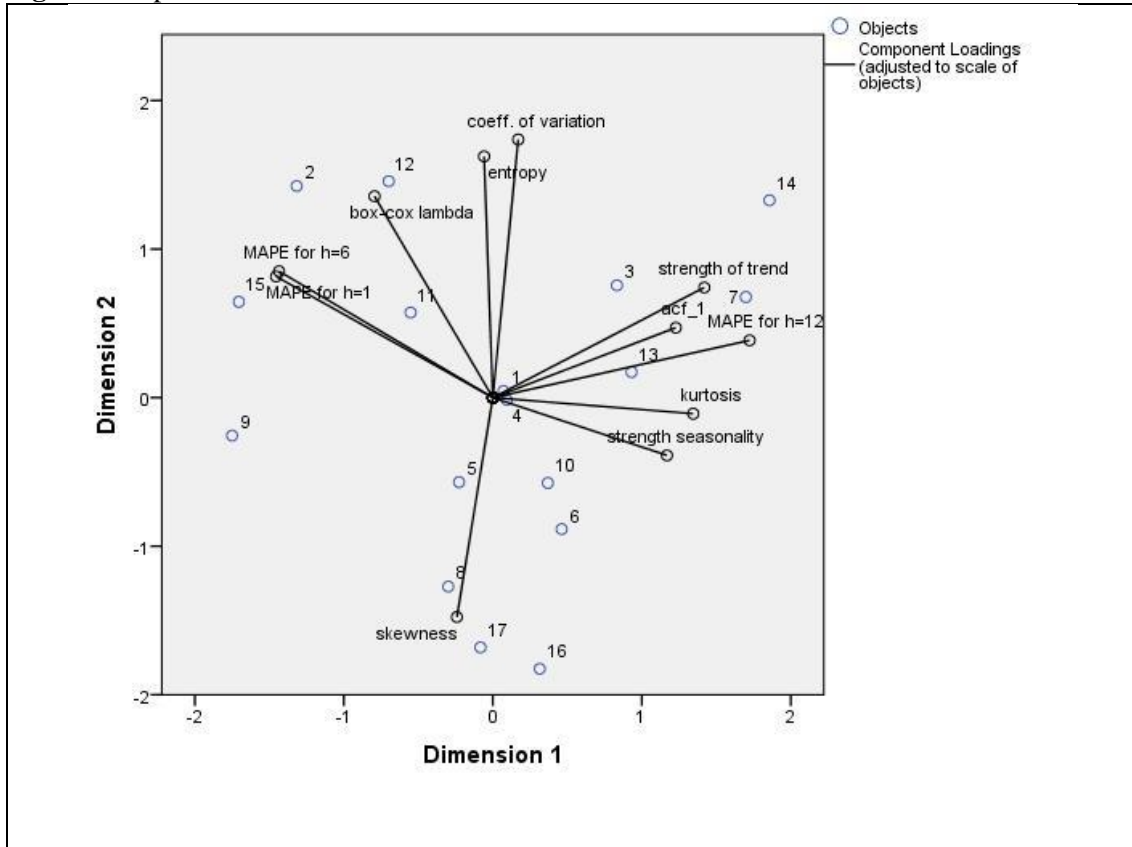
Notes. Rotation Method: Varimax with Kaiser Normalization. Acf_1 denotes the first autocorrelation. For visual clarity, we have coded each region (object) with a number: Andalusia (1), Aragon (2), Asturias (3), Balearic Islands (4), Canary Islands (5), Cantabria (6), Castilla-Leon (7), Castilla-La Mancha (8), Catalonia (9), Valencia (10), Extremadura (11), Galicia (12), Madrid (13), Murcia (14), Navarra (15), Basque Country (16), La Rioja (17).

Table 3. Rotated component loadings

	Dimension	
	1	2
Strength of trend	0.76	0.40
Strength of seasonality	0.63	-0.21
Optimal Box-Cox transformation	-0.43	0.73
Spectral entropy	-0.03	0.87
Autocorrelation function (r1)	0.66	0.25
Coefficient of variation	0.09	0.93
Skewness	-0.13	-0.79
Kurtosis	0.72	-0.06
MAPE for h=1	-0.78	0.44
MAPE for h=6	-0.77	0.46
MAPE for h=12	0.92	0.21

Note. Component loadings indicate Pearson correlations between the quantified variables and the principal components (ranging between -1 and 1).

The proximity observed between entropy and dispersion is indicative of a close and positive relation between them, and of a negative relation with accuracy. The strength of the trend, the first autocorrelation, and the MAPE for h=12 are close to the strength of seasonality and kurtosis, all of which seem negatively related to the rankings based on the MAPE for one- and six-month ahead forecasts. Based on this evidence, we conclude that the interactions of the strength of the seasonal and the trend components with forecast accuracy vary depending on the forecast horizon. This finding is in line with the results obtained by Hassani, Silva, Antonakakis, Filis, and Gupta (2017).

Conclusions

The increasing importance of the tourism industry worldwide is fostering a growing interest in new approaches to tourism modelling and forecasting. In this paper we have combined the results of two independent analyses so as to position Spanish regions according to the features of international tourist arrivals time series and to the accuracy of forecasts obtained for each region.

First, we have applied a seasonal-trend decomposition procedure based on non-parametric regression to isolate the different components of the series and compute their features. Despite certain similarities between regions regarding the strength of seasonality, the rest of the characteristics of the time series markedly differ across regions. Then, we have calculated the out-of-sample accuracy of several machine learning methods in a multiple-step-ahead forecasting experiment, also finding substantially different results between regions, techniques and forecast horizons. Finally, we have ranked the regions according to the obtained results in the two previous analyses and synthesized all the information in two components.

By means of categorical principal components analysis we have projected the distribution of the Spanish regions in two dimensions according to their time series features and the obtained forecasting results. We observed that altough entropy and dispersion appear to be negatively related with forecast accuracy for all horizons, the interactions of the rest of data characteristics with forecast accuracy depend on the length of the prediction horizon. As a result, in order to improve forecast accuracy of tourist arrivals at the regional level, we suggest applying region-specific model selection that takes into account the forecast horizon.

We want to note that we have carried out an exploratory study, and consequently the results cannot be generalized. A question left for further research is the incorporation of additional time series features and the implementation of other forecasting methods. Finally, extending the analysis to regions of other destinations would allow to test if there are significant differences across countries.

References

Akin, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management*, 48, 64–72.

Ben Taieb, S., Sorjamaa, A., & Bontempi, G. (2010). Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10–12), 1950–1957.

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, 32(2), 303–312.

Bilen, M., Yilanci, V., & Eryüzlü, M. (2017). Tourism development and economic growth: a panel Granger causality analysis in the frequency domain. *Current Issues in Tourism,* 20(1), 27–32.

Chen, K. Y., & Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management*, 28(1), 215–226.

Claveria, O. (2016). Positioning emerging tourism markets using tourism and economic indicators. *Journal of Hospitality and Tourism Management*, 29, 143–153.

Claveria, O. (2017). Two-dimensional mapping of Asia Pacific destinations combining tourism and economic indicators. *Asia Pacific Journal of Tourism Research*, 22(7), 720–734.

Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220–228.

Claveria, O., Monte, E., & Torra, S. (2015). Common trends in international tourism demand: Are they useful to improve tourism predictions? *Tourism Management Perspectives*, 16, 116–122.

Claveria, O., Monte, E., & Torra, S. (2016). Modelling cross-dependencies between Spain's regional tourism markets with an extension of the Gaussian process regression model. *SERIEs*, 7(3), 341–357.

Claveria, O., Monte, E., & Torra, S. (2017). Data pre-processing for neural networks-based forecasting: Does it really matter? *Technological and Economic Development of Economy*, 23(5), 709–725.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1), 3–73.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. In M. Mozer, M. Jordan, and T. Petsche (Eds.), *Advances in Neural Information Processing Systems*, 9 (pp. 155–161). Cambridge, MA: MIT Press.

Goerg, G. (2013). Forecastable component analysis. *Proceedings of the 30th International Conference on Machine Learning*, (Icml-13), 64–72.

Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research*, 63, 112–127.

Hong, W., Dong, Y., Chen, L., & Wei, S. (2011). SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. *Applied Soft Computing*, 11(1), 1881–1890.

Kang, Y., Hyndman, R. J., & Smith-Miles, K. (2017). Visualising forecasting algorithm performance using time series instance spaces. *International Journal of Forecasting*, 33(2), 345–358.

Kim, D., & Shwartz, Y. (2013). The accuracy of tourism forecasting and data characteristics: a meta-analytical approach. *Journal of Hospitality Marketing and Management*, 22(4), 349–374.

MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press.

Rasmussen. C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press.

Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond* (Adaptive computation and machine learning series). Cambridge, MA: MIT Press.

Sharif, A., Saha, S., & Loganathan, N. (2017). Does Tourism Sustain Economic Growth? Wavelet-Based Evidence From the United States. *Tourism Analysis: An Interdisciplinary Tourism & Hospitality Journal*, 22(4), 467–482.

Vapnik, V. N. (1998). *Statistical learning theory*. New York: Wiley.

Wang, X., Smith-Miles, K., & Hyndman, R. J. (2009). Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. *Neurocomputing*, 72(10), 2581–2594.

Wu, Q., Law, R., & Xu, X. (2012). A spare Gaussian process regression model for tourism demand forecasting in Hong Kong. *Expert Systems with Applications*, 39(4), 4769–4774.