



Spatial analysis and prediction of the flow of patients to public health centres in a middle-sized Spanish city

Isabel Ramos,¹ Juan J. Cubillas,² Francisco R. Feito,^{2,3} Tomas Ureña⁴

¹Department of Cartographic, Geodetic Engineering and Photogrammetry, University of Jaen, Jaen; ²TIC-144 Andalusian Research Plan, Department of Computer Science, University of Jaen, Jaen; ³Department of Computer Science, University of Jaen, Jaen; ⁴Sanitary District, Andalusian Health Service, Jaen, Spain

Abstract

Human and medical resources in the Spanish primary health care centres are usually planned and managed on the basis of the average number of patients in previous years. However, sudden increases in patient demand leading to delays and slip-ups can occur at any time without warning. This paper describes a predictive model capable of calculating patient demand in advance using geospatial data, whose values depend directly on weather variables and location of the health centre people are assigned to. The results obtained here show that outcomes differ from one centre to another depending on variations in the variables measured. For example, patients aged 25-34 and 55-65 years visited health centres less often than all other groups. It was also observed that the higher the economic level, the fewer visits to health centres. From the temporal point of view, Monday was the day of greatest demand, while Friday the least. On a monthly basis, February had the highest influx of patients. Also, air quality and humidity influenced the number of visits; more visits during poor air quality and high rela-

tive humidity. The addition of spatial variables minimised the average error the predictive model from 7.4 to 2.4% and the error without considering spatial variables varied from the high of 11.8% in to the low of 2.5%. The new model reduces the values in the predictive model, which are more homogeneous than previously.

Introduction

Current ways to improve health systems are increasingly using spatial variables, *e.g.* Croner (2003) used data from the Internet and tools based on geographic information systems (GIS) to manage human resources with regard to health, while Gonçalves *et al.* (2014) dealt with the optimal location of a hospital comes. When applying geospatial approaches, the dimension of the resources examined, a key issue in health systems planning, often comes to the fore. Generally, the resources needed in a primary health care centre are quantified from historical data regarding the services required. Thus, the strategy adopted by those responsible for the management of the health resources is commonly based on the average number of patients received, *i.e.* the patient flow. This policy is appropriate if the demand for health care fully depend on such calculations. However, a sudden increase in demand would result in the health centre not being able to attend all patients requesting its service. This leads to delays and slip-ups, *e.g.*, patients cannot get their appointments as scheduled, doctors' agendas collapse and emergencies are more difficult to handle than otherwise. Research, such as that by Lichtner *et al.* (2013), highlights the importance and necessity of efficient planning, competent management and expert use of resources in health services. Indeed, adequate resources management in primary health care leads to higher score for patient satisfaction and reduces health care costs and drug prescriptions (Starfield, 1992). The definition of quality health care has been analysed from different points of view as summarised by Campbell *et al.* (2000). The bottom line is to provide the services needed when required by the patient and using the most appropriate tests and procedures to achieve the results expected without delay. The research described in this paper thus deals with access to health centres indicating the ways and means needed to improve overall outcomes and the delivery of primary health care services by focusing on delivery and management of health services and the medical resources available in the city of Jaen, in Andalusia, south Spain.

Background

In order to know in advance health centres' needs with respect to human and material resources, the University of Jaen, Andalusia (located in southern Spain) and the Jaen Health District have recently

Correspondence: Isabel Ramos, Department of Cartographic, Geodetic Engineering and Photogrammetry, University of Jaen, Campus Las Lagunillas, 23071 Jaen, Spain.
Tel. +34.953.212854 - Fax: +34.953.212855.
E-mail: miramos@ujaen.es

Key words: Spatial analysis; Health care; Resource optimisation; Data mining; Spain.

Acknowledgments: this work has been partially supported by the Andalusian Health Service, Department of Equality, Health and Social Policies of the Junta of Andalusia, Spain.

Received for publication: 19 January 2016.
Revision received: 17 July 2016.
Accepted for publication: 29 August 2016.

©Copyright I. Ramos *et al.*, 2016
Licensee PAGEPress, Italy
Geospatial Health 2016; 11:452
doi:10.4081/gh.2016.452

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



collaborated on a data mining study to predict patient flows distinguishing between the medical the administrative requirements (Cubillas *et al.*, 2014). This study was undertaken with historical data from the latest five years. The first four years were used to design de prediction model that was finally validated using data from the fifth. Data mining algorithms, such as minimum description length (MDL) (Grünwald, 2003), were used in the study to determine the weight of each attribute used on the target attribute, in particular the total number of patients attended to in the health care centre. Also, multiple regression algorithms such as the generalised linear model (GLM) (Dobson and Barnett, 2002) and the support vector machine (SVM) (Cortes and Vapnik, 1995) with linear kernel and Gaussian kernel were used for predicting the number of patients, who needed attention by the health centres in the city of Jaen. The model estimated analyses how meteorological and air pollution factors affect the patient flow. This predictive model is capable to accurately determine the global number of patients attended to in the health centres, but when used to predict the influx of patients to a specific centre, it was not as accurate as anticipated. In fact, the prediction obtained for each health centre showed different levels of accuracy (Table 1). In this case, the error for each health centre is calculated from the difference in absolute value between the number of patients attending in each health centre and the number of patients predicted from the model without using spatial variables. The differences in the error values between health centres, presumably due to local factors inherent to each health centre, were considered in the previous study (Cubillas *et al.*, 2014) but must be taken into account for the model to be useful (Beltrán-Sánchez *et al.*, 2015). Therefore this is the reason why such predictive model works well when global information about the total health centres of Jaen is needed, since overall mean values of the city of Jaen are introduced. With this model, it is not possible to make a correct prediction influx of patients to a particular health centre. Since, as seen above, this model does not consider local factors that influence each health centre particularly. These types of factors are what we call in this paper as spatial variables.

New contribution

The design of the new model is similar to the model mentioned in the background section. Also the periods of time considered are the same. The main goal of this research was to determine to what extent the use of spatial variables improves the model. To make this comparison, it is necessary to use the previous model with and without the spatial variable during the same period of time. To do so, we increased the level of detail and precision of the previous, predictive model to obtain valid information about each health centre. In order to undertake valid decisions about the management of each individual health centre, a new predictive model was generated. This implies the definition of all the variables that influence patients living within the service area of each health centre to visit it. In other words, we focused on the local factors that determine the physical particulars of the vicinity around the different centres. Variables such as the type of population classified according to age, is an important variable indicating the different health needs of paediatric, young and geriatric clientes (Schäfer *et al.*, 2012). Another factor to be considered is the economic level of the area. In theory, a wealthy area would have a greater number of patients with private health insurance. Usually these people also tend to have a higher cultural level and therefore can more easily access health information, which may sometimes reduce visits to the local doctor. In addition, people with better economic and cultural levels usually have a healthier life style resulting in less disease in the long term. Finally, weather variables, such as temperature and relative humidity, should

also be considered. Specific interpolation algorithms allowed us to obtain the needed, specific values for each location. In short, the design of a new predictive model would depend crucially on the local spatial variables.

Materials and Methods

Study area

Each primary health care centre covers a specific health district in Jaen, which means that each centre serves only the population resident in its district. Each district is comprised of census tracts, which are administrative boundaries of the city conducted by the local government. In the city of Jaen, seven health districts serve the entire population slightly more than 116,000 people covering around 6.58 km². The digitised map of the city of Jaen served as the baseline cartographic information that was integrated into a GIS by use of MapInfo Professional v.11 (MapInfo, 2013). The cartographic basis of the health districts was digitised in addition to the census tracts for each district (Figure 1).

Type of patient data

The predictive model created for this study is based on data from the four years 2007-2010 plus the 2011 data, which were used to validate the calculations and the model output with special reference to the predicted and real values from 2011. We considered the number of patients treated each day at the health care centres during those years, and in order to increase the level of detail of the predictive model, a study of the particular spatial variables involved was also included. The variables for each health centre were recorded with reference to: the type of patient by classification according to age, distinguishing between geriatric, young and paediatric population; the economic level of the service area of the health centre; and the weather variables temperature and relative humidity collected for each health district area for each day during the period considered. These data come from weather stations located in the city of Jaen (REDIAM, 2014).

In order to avoid the potential source of error connected with the difficulty obtaining accurate population data, particularly with reference to the age ranges, from the databases at each health centre, we obtained the census tracts from the National Statistics Institute (INE, 2014). Once all this information had been collected and stored, the number of health centre clients for each health centre was obtained by carrying out specific, spatial queries through the GIS. At this stage each new variable introduced into the database had its own spatial component that were eventually used to generate the prediction model. The percentage of patients by age and economic level were added to each district. The patients assigned to each health centre and classified by age range can be seen

Table 1. Model error when applied to each health centre.

Health care centre	Error (%)
Belén	6.60
El Valle	7.16
Federico del Castillo	2.54
Fuentezuelas	11.75
La Magdalena	6.67
San Felipe	5.32
Virgen de la Capilla	11.77



in Table 2. The table reveals substantial differences with respect to the age range of the patients between the various health centre service areas under study. For example, Federico del Castillo has 18.4% of the paediatric population, while Belen has only 12.6%. Another important difference is the low percentage of geriatric people using the Fuentezuelas health centre (1.87%) compared with the Virgen de la Capilla centre (5.04%). These differences show that the range age factor should be postulated as an influential factor in the predictive model.

Economic levels

Because of the absence of any kind of indices measuring the economic level of each zone of the city of Jaen, we considered the value of a standard property in each area as a valid rating. A three-room apartment, measuring approximately between 80-90 m² floor surface that has access to a garage and a storage room outside, the most common type of property in this city, was therefore used as the standard. We researched this information manually throughout each street and inserted the data

found into the GIS database (Figure 2). By assigning the average value for each street to the corresponding census tract an adequate approximation of the average economic status of patients belonging to each health centre could be made (Table 3).

Weather variables

Values of the weather variables, such as temperature or relative humidity, were measured at five weather stations located in different parts of the city (Figure 3). In order to assign these meteorological values to the various health districts, the GIS tools were used. Most of the meteorological variables were mapped or analysed by spatial interpolation techniques, *e.g.*, by application of Kriging spatial interpolation methods (Xuan *et al.*, 2015).

Although Jaen is not an overly large city, the map of average temperatures and relative humidity shows significant differences between the health districts. The main reason is the orography of the area where the city is located. These differences in environmental values coincide with



Figure 1. Health districts and census tracts in Jaen, Spain.

Table 2. Patients assigned to each health care centre classified by age.

Age range (years)	Belen	El Valle	Federico del Castillo	Fuente-zuelas	La Magdalena	San Felipe	Virgen de la Capilla
0-14	12.70	15.66	18.40	18.04	16.74	14.85	12.06
14-24	10.90	11.84	9.77	13.17	12.66	12.60	8.55
25-34	17.69	17.04	15.64	16.17	17.53	16.04	15.80
35-44	15.72	17.38	18.66	18.84	15.39	14.99	17.47
45-54	14.23	14.20	13.79	15.84	14.21	15.55	14.10
55-64	11.41	9.42	8.21	8.82	8.92	9.52	10.92
65-74	7.88	6.43	6.31	4.34	6.12	6.51	8.80
75-84	5.92	4.67	5.59	2.91	5.40	6.48	7.27
>85	3.54	3.37	3.64	1.87	3.03	3.45	5.04

Source: National Statistics Institute (INE). Values are expressed as percentage.



the different types of population belonging to each health district, which results in health demands varying from one health centre to another. February is the colder month, which means that the demand for health care is high in all health centres. However, as seen in Table 4, the percentage of patients actually attended to compared with the total number of registered persons in each health district varies from one to another confirming that the different particulars of each health district leads to an uneven demand for health care.

Spatial variables

Knowledge of the role played by the spatial variables in our predictive model is essential to the efforts to improve its effectiveness. Depending on location, these variables have an unequal influence on the outcomes and it is consequently crucial to analyse this issue. Therefore, the MDL algorithm was applied to determine the weight of each variable (Table 5). The influence of the new variables on the prediction model is confirmed by the high weights obtained.

Results

With respect to the type of population seeking the attention of the health services, it can be observed that patients from 0-14 years and 75-84 years (the paediatric and geriatric populations) constitute the two groups that most positively influenced the number of medical visits. Out of all groups, patients aged 25-34 and 55-65 were those who influenced the number of visits most negatively, *i.e.* the groups with the least frequent medical visits (Table 6).

From the daily point of view, Monday was the day of greatest patient demand and Friday the least. When the months were considered,

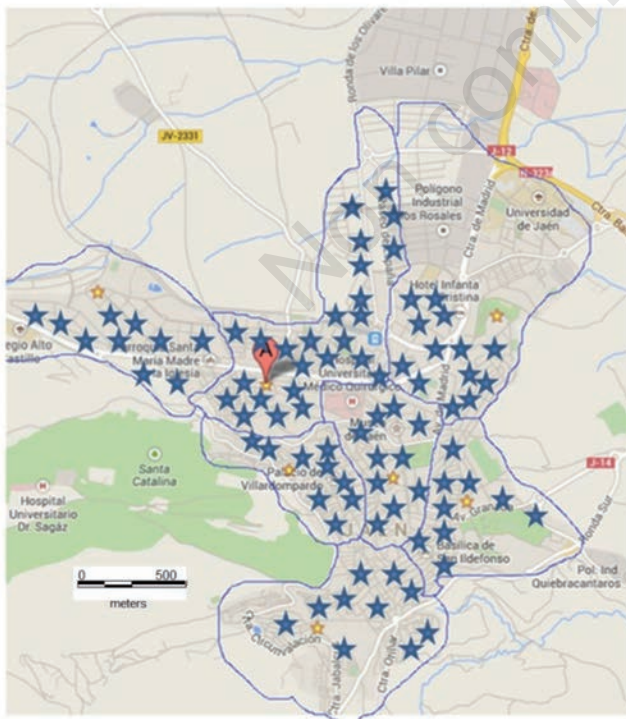


Figure 2. Properties in Jaen, Spain used to calculate the economic level of each district.

February had a higher influx of patients than the rest of the year. The average temperature was found to be more influential than the maximum and minimum values. It was also observed that the higher the economic level, the fewer visits to health centres (regression coefficient - 0.034). Also the poor air quality factor influenced increased visits to the centres (coefficient 0.008). Finally, relative humidity is also influential, the higher the relative humidity in the zone, the higher the number of patients visiting the doctor.

As a result, the addition of the spatial variables analysed into the pre-

Table 3. Average price in the various areas.

Health care centre	Average price (€)
Belén	139,125
El Valle	109,250
Federico del Castillo	161,000
Fuentezuelas	137,000
La Magdalena	94,500
San Felipe	101,500
Virgen de la Capilla	244,750

Prices of apartments in Jaen city are taken from Websites Real States.

Table 4. Percentage of patients demanding health care on a specific date.

Health care centre	Patients assigned (n)	Patients who attended (%)
Belén	8987	2.59
El Valle	16,148	3.16
Federico Castillo	24,959	2.98
Fuentezuelas	8030	2.84
La Magdalena	12,221	2.83
San Felipe	19,239	2.82
Virgen de la Capilla	20,405	2.21N

Data refer to the arbitrary date of 14 February 2011.

Table 5. Effect of the spatial variables on the various target attributes.

Attribute	Rank	Weight	
Age (years)	0-14	1	0.731
	15-24	1	0.731
	25-34	1	0.731
	35-44	1	0.731
	45-54	1	0.731
	55-64	1	0.731
	65-74	1	0.731
	75-84	1	0.731
>85	1	0.731	
Economic level	1	0.731	
Visits	1	0.731	
Month	3	0.064	
Minimum temperature	4	0.052	
Maximum temperature	5	0.046	
Mean temperature	6	0.04	
Air quality	7	0.007	
Relative humidity	8	-0.008	
Day of the week	9	-0.019	



diction model minimised its average error from 7.40 to 2.41% (Table 7). The error of the predictive model without considering spatial variables varied from 11.77% in the health centre of Virgen de la Capilla to 2.54% in Federico del Castillo. Although these values would be accepted as good results the new model improves them in two aspects: it reduces the errors of the predictive model for each health care centre in Jaen, and in addition these values are more homogeneous than those resulted from the previous model.

Discussion

The study confirms that the most influential factors needed to predict the demand on health centres are the type of population and the economic level in the service area (which are specific for each health centre), followed by more general factors such as month of the year, temperature, environmental quality, relative humidity and the day of the week. Other parameters obtained from the model are regression coefficients, which allow an understanding of the relationship between the current performance at each health centre and the spatial variables as they show how each variable influences patient demand. The regression coefficients in Table 6 have been rescaled in order to clarify the analysis of the influence of each variable. Its four columns contain: i) the attributes of the predictive model; ii) the values of these attributes; iii) the standardised coefficients; and iv) the regression coefficients. The coefficients of this table refer to how many standard deviations (SD) a dependent variable would change per SD increase in the predictor variable. It was felt useful to answer the question of which of the independent variables have the greater effects on the dependent variable in multiple regression analysis when the variables are measured in different units of measurement. The analysis of the standardised coefficients provides an understanding about in which way spatial variables add or subtract patient visits to the health centres.

Table 6. Regression and standardised coefficients.

Attribute	Value	Standardised coefficient	Regression coefficient
Day of the week	Monday	0.102	34.749
	Tuesday	0.075	25.544
	Wednesday	0.014	4.765
	Thursday	0.008	1.234
	Friday	-0.162	-55.203
Age (years)	0-14	0.078	4.781
	15-24	-0.022	-1.973
	25-34	-0.187	-31.210
	35-44	0.011	1.121
	45-54	0.043	9.735
	55-64	-0.123	-15.28
	65-74	-0.030	-3.857
	75-84	0.102	14.448
>85	0.059	11.503	
Relative humidity		0.057	0.449
Month	January	0.030	19.29
	February	0.039	19.22
	March	0.022	10.96
	April	-0.001	-0.414
	May	0.003	1.663
	June	-0.106	-52.551
	July	-0.319	-157.451
	August	-0.392	-193.725
	September	-0.206	-101.654
	October	0.031	15.246
	November	0.012	5.839
	December	-0.087	-42.953
Economic level		-0.034	-14.541
Good air quality		0.008	3.611
T _{max}		0.099	1.762
T _{mea}		0.148	2.949
T _{min}		-0.051	-1.188

T_{max}, maximum temperature; T_{mea}, mean temperature; T_{min}, minimum temperature.

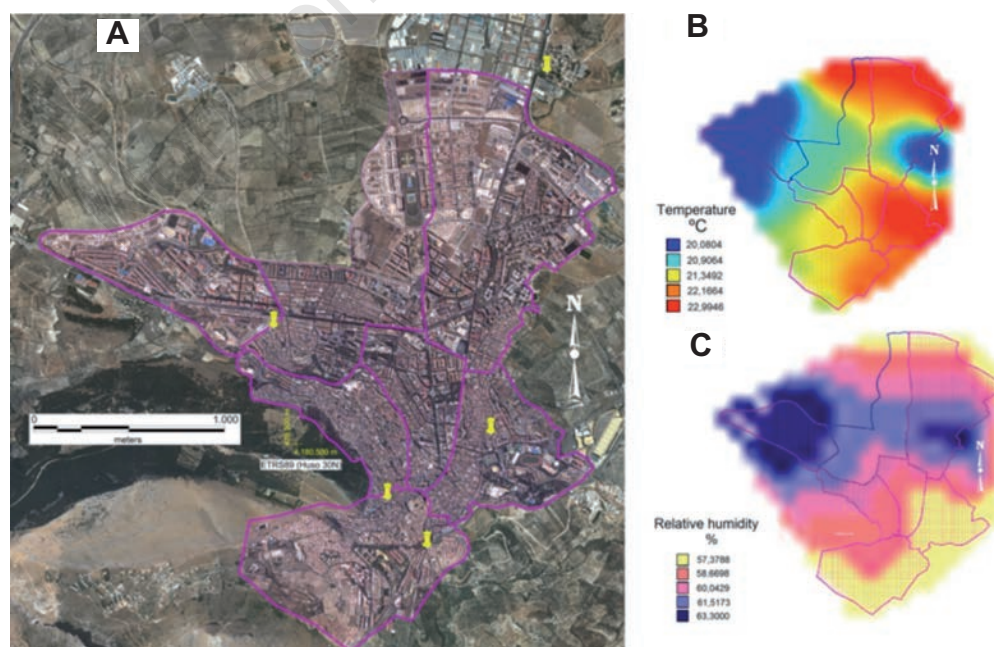


Figure 3. Mean temperature and relative humidity in Jaen, Spain. Yellow pins indicate the location of the weather stations (A); overview of the mean temperature (B); overview of the relative humidity (C).



From a statistical point of view, a relationship between certain environmental, social factors and the number of patients who visit their health care centres has been established. Particularly, the spatial variables are those that best model the peculiarity of each health care centre location. Therefore, in order to make correct decisions in advance about management issues in the health centre, these new variables (age range and economic data) have been taken into account. In this sense, for the environmental variables their spatial component also have been considered as their different values from one place to other have been introduced in the model, instead of considering average values. The economic level is one of the most influential variables. It was determined that the higher economic level the people in the health district have, the fewer the number of visits to the health centre are registered. The main reason may be that most people with high economic level have private health insurance; therefore they do not use the public health services. In addition, people with better economic and cultural levels usually have a healthier living styles and, therefore, have less disease in the long term. Regarding the type of population who attended the health care centre, it was confirmed that the greater the number of people in the 0-14 years old bracket assigned to a health centre, the greater the number of visits to it, while the 25 to 34 years old bracket uses there are fewer medical services demanded. In this sense, health centres with many paediatric people registered will have much more visits than others. Especially, during the critical months in the school period where pandemics like the flu is widespread among children. Nowadays cities have expansion areas where young families with pre-children live, whereas older people, as those in geriatric age, should live in the city centre. These issues reinforce the importance of using local factors better than equal average values for all health care centres in the city. Furthermore, the results obtained show that local variables influence the influx of people to the health centre in a different way from one place to others. Different values for spatial variables provoke differences in patient demands data.

The model developed could be a useful tool for the responsible staff to manage the resources in the health centre, especially the possibility of knowing in advanced some external factors, *e.g.*, environmental variables such as temperature and relative humidity, which can be had at a reasonable level of accuracy one, even two weeks in advance, a situation which contributes to anticipation and planning for resources that will be needed.

This research has been specifically developed using data from the health districts of Jaen, and spatial variables values for these various locations. Nevertheless, the model generated could be adapted to any other place following the methodology described in previous sections. In this sense, the key point is to determine the most influential factors in the model, and the availability of data from at least four or five years

Table 7. Improvement of the prediction model by using spatial variables.

Health care centre	Absolute errors without spatial variables (%)	Absolute errors with spatial variables (%)
Belén	6.60	1.99
El Valle	7.16	2.36
Federico del Castillo	2.54	1.87
Fuentezuelas	11.75	3.10
La Magdalena	6.67	2.55
San Felipe	5.32	2.70
Virgen de la Capilla	11.77	2.34

back in order to generate an accurate predictive model. In the case of Jaen, we have been able to predict the number of patients who will need medical care with an absolute error of 2.41%.

Conclusions

Provision of accurate data about the number of patients who would visit the health care centre on a specific date in advance is possible with the model generated. The kind of tools presented here facilitates forward planning, the one important aspect of health care management that must be improved. Information gained is indeed critical for resource managers of health centres, as they would need to provide adequate resources for their services in the case of sudden demand.

References

- Beltrán-Sánchez H, Drummond-Andrade FC, Riosmena F, Pinto G, Palloni A, Novak B, Graham JE, 2015. Contribution of socioeconomic factors and health care access to the awareness and treatment of diabetes and hypertension among older Mexican adults. *Salud Publica Mexico* 57(Suppl.1):6-14.
- Campbell SM, Roland MO, Buetow SA, 2000. Defining quality of care. *Soc Sci Med* 51:1611-25.
- Cortes C, Vapnik V, 1995. Support-vector networks. *J Mach Learn Res* 20:3:273-97.
- Croner CM, 2003. Public health, GIS, and the Internet. *Annu Rev Publ Health* 24:57-82.
- Cubillas JJ, Ramos MI, Feito FR, Ureña T, 2014. An improvement in the appointment scheduling in primary health care centers using data mining. *J Med Syst* 38:89.
- Dobson AJ, Barnett AG, 2002. An introduction to generalized linear models. Chapman & Hall, London, UK.
- Gonçalves J, Ferreira JA, Condessa B, 2014. Making regional facility location decisions: the example of Hospital do Oeste Norte, Portugal. *Geospat Health* 9:1-6.
- Grünwald P, 2003. Advances in minimum description length: theory and applications. Massachusetts Institute of Technology Press, Cambridge, MA, USA.
- INE, 2014. Available from: <http://www.ine.es/>
- Lichtner V, Venters W, Hibberd R, Cornford T, Barber N, 2013. The fungibility of time in claims of efficiency: the case of making transmission of prescriptions electronic in English general practice. *Int J Med Inform* 82:1152-70.
- MapInfo, 2013. User guide MapInfo v.11.0. Pitney Bowes software Inc. One Global View, Troy, NY, USA.
- REDIAM, 2014. Available from: <http://www.juntadeandalucia.es/medioambiente/site/rediam>
- Schäfer I, Hansen H, Schön G, Höfels S, Altiner A, Dahlhaus A, König HH, 2012. The influence of age, gender and socio-economic status on multimorbidity patterns in primary care. First results from the multicare cohort study. *BMC Health Serv Res* 12:89.
- Starfield B, 1992. Primary care and health. A cross-national comparison. *J Am Med Assoc* 268:2032.
- Xuan TN, Ba TN, Khac PD, Quang HB, Thi Nhat TN, Van QV, Thanh HL, 2015. Spatial interpolation of meteorological variables in Vietnam using the kriging method. *J Inf Process Syst* 11:134-47.