INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

The main vectors influencing the levels of fuel consumption in a bus company. The lines, the bus driver and the vehicles

Rodrigo Manuel da Silva Vargas

Dissertation for obtaining the degree of master in

Civil Engineering

Jury

President: Professor José Álvaro Pereira Antunes Ferreira
Supervisor: Professor João António de Abreu e Silva
Vowel: Professor Filipe Manuel Mercier Vilaça Moura

October 2011

Abstract

The definition of parameters that take a part on the fuel consumption of the service provided by transportation companies is a complex but useful tool to optimally manage the companies' resources. The development of a methodology that can provide this is information is an important step for developing more sustainable services of transport.

To develop such a methodology it was used stepwise multiple regression models for each of the main actors in the service of a company of road transportation: bus lines, drivers and vehicles. For this purpose was gathered data confined to the months of October 2009, November 2009, March 2010, May 2010 and September 2010, with the last two being for validation purposes only. The dimension of the data samples collected for developing the three models consisted in 87 routes, 488 drivers and 105 vehicles. The collected data permitted to describe, with precision, each of the main actors, which were then adapted to each model.

As conclusion, the main variables detected as influential to fuel consumption are the type of vehicle, the commercial speed, the slope greater than 5% and the localization of bus stops. Besides these variables, there were some driving events that were also selected for being harmful to the fuel's consumption such as the air loading with excessive rotation, the sudden longitudinal decelerations and the rotations above the stipulated maximum value.

Keywords: consumption, fuel, route, driver, bus, energy efficiency, commercial speed, multiple linear regression.

1. Introduction

Nowadays, mobility is an essential part of an increasingly urbanized and complex world. Thus transportation, both of people and goods, occupies a central part in modern societies and has become more extensive and far-reaching than ever. Therefore it represents an important share in the consumption of fossil fuels, and road transportation is responsible for an expressive majority of that consumption.

Rodoviária de Lisboa S.A. operates in the municipalities of Lisbon, Loures, Odivelas and Vila Franca de Xira, serving 400 000 inhabitants and carrying 200 000 passengers per day. To examine which factors have a higher influence on fuel consumption three models were developed for the referred main vectors. Fuel consumption represents an important parcel of the total operating costs for urban bus operators. Thus in a context of raising fuel price, reducing fuel consumption is an important objective for bus operators. But in order to control fuel consumption it is fundamental to know the main factors that contribute to it.

The main objective of this research is to develop three regression models that can predict the fuel consumption for a bus line, a driver or a vehicle taking into account the variables that are considered as the most influential in this forecast.

2. State of the art

The following topics describe some of the most influential researches that were used as base for the development of the present investigation:

- According to VOLVO, 2009, adding a bus stop at every 10 km suggests an increase of the fuel consumption;

- For positive slopes (ascending) above 5%, the correlation with consumption proved to be very clear (SIMÕES, A. 2005);
- With the formative sessions, Rodoviária de Lisboa reached a 2.6% reduction in specific consumption of the overall company by April 2009 (KOSHAL, R. K. 1970);
- The mass reduction of the average fleet operator Horários do Funchal was analyzed, according to (SIMÕES, A. 2005), reducing the mass of the fleet by 10%. It was predicted that this measure could decrease the fuel consumption by 2%;
- Another factor of considerable influence to the fuel consumption is the idle time with the motor working. A typical gasoline vehicle spends about 1 liter of fuel per hour idling, and diesel one spends about 0.7 liters of fuel per hour at idle (18);
- According to VOLVO, 2009, a speed reduction from 90 km/h to 80 km/h reduces fuel consumption by 6%;
- It was found that the commercial speed of a vehicle has a large effect on consumption, in which the bigger the commercial speed, the lower the fuel consumption (UITP, 2009).

3. Linear regression

The methodology adopted to develop the fuel consumption models was stepwise multiple regression analysis and the dependent variable was the logarithm of the fuel consumption. The data used in these models included the type of bus line, the topography characteristics of each line, the vehicle typology and several variables collected by the onboard information system FM200 which collects driving events and relates them with the driver, the vehicle and the line. Three models were developed concerning bus lines, drivers and vehicles.

Multiple regression is a flexible method of data analysis that may be appropriate whenever a quantitative variable (the dependent or criterion variable) is to be examined in relationship to any other factors (expressed as independent or predictor variables). Relationships may be nonlinear, independent variables may be quantitative or qualitative, and one can examine the effects of a single variable or multiple variables with or without the effects of other variables taken into account.

4. Data gathering

The set of data collected can be divided into three main groups: bus lines, drivers and vehicles. The aim of this chapter is to describe and overview the data collected and present the variables in the model as well as the methodology applied for its calculation. The data gathering was confined to the months of October 2009, November 2009 and March 2010. Later, more data was gathered in the months of May 2010 and September 2010 for results validation purposes only. The dimension of the data samples collected for developing the three models consisted in 87 routes, 488 drivers and 105 vehicles.

To describe the routes the following data were collected: length of the route, number of stops, distance between stops, percentage of kilometers corresponding to various intervals of inclination, the rout type and the average commercial speed to each route. To describe the drivers the following data were collected: percentage of each driving event, number of formative classes, drivers age and the drivers experience in years. Finally, to describe the vehicles the following data were collected: average fuel consumption, vehicle type, percentage of miles driven by each type of vehicle, vehicle's mass and vehicle's age. All data was gathered directly from *Rodoviária de Lisboa* except the topography information from the bus lines which was developed adding altitude to the geographical data (longitude and latitude) obtained from the bus transportation company. Then, these data were transformed into variables that could be used in each model using *MySQL*, a relational database management system.

5. Regression analysis

As follows, the results of the three regression models developed will be described.

5.1. Bus lines

Algebraically the results are presented below:

$$\begin{aligned}\log(Consumption) = 1,741 - 0,012\text{NMonit} - 0,005 ComSpeed -\\ -0,004 VeicMini + 1,173\text{x}10^{-5}(\text{ISup5\%})^2 + 0,019\log(MaxDistStops) +\\ +0,48 Ev0140 + 0,002 VeicArtic - 0,001 VeicMini\end{aligned} \quad (1)$$

The results of the model, based on 87 observations, are presented in the table below.

Table 1 – Summary of the bus lines model

R	R square	Adjusted R square	Standard Error of the Estimate
,986	,972	,970	,01554

An R^2 of 0.972 suggests a very high fit: 97.2 % of the variation in fuel consumption of the bus lines can be explained by the developed model.

In the following table are presented the coefficients obtained as well as some statistical information for each variable.

Table 2 - Statistical coefficients of the selected variables of the bus lines model

Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	1,741	,024		71,494	,000		
Average of monitoring sessions of drivers	-,012	,003	-,117	-4,201	,000	,458	2,183
Commercial speed	-,005	,001	-,263	-7,788	,000	,311	3,220
Percentage of Mini-bus use	-,004	,000	-,888	-42,120	,000	,798	1,253
Square Percentage of inclination superior to 5%	1,173E-5	,000	,104	3,713	,000	,450	2,224
Logarithm of the maximum distance between stops	,019	,008	,070	2,269	,026	,368	2,719
Percentage of 0140 event occurrence[1]	,048	,021	,049	2,308	,024	,795	1,257
Percentage of articulated bus use	,002	,000	,238	12,053	,000	,911	1,098
Percentage of midi bus use	-,001	,000	-,167	-8,565	,000	,933	1,071

[1] 0140 Event - Air loading with excessive rotation.

In all cases the results are significant at 1%, except the logarithm of maximum distance between stops and the occurrence percentage of the 0140 event which are significant at 3%. Given that the normal measure used is a significance level of 5% the model provides very strong results.

5.2. Drivers

Algebraically the results are presented below:

$$\log(Consumption) = 1{,}965 - 0{,}145\log(ComSpeed) - 0{,}005 VeicMini - \\ -0{,}002 VeicStand - 0{,}002 VeicMidi + 0{,}002 I\sup 5\% - \\ -0{,}002 AvrLenght + 0{,}003 Ev1007 \tag{2}$$

The results of the model, based on 488 observations, are presented in the table below:

Table 3 – Summary of the drivers model

R	R square	Adjusted R square	Standard Error of the Estimate
,958[a]	,917	,916	,01750

The 0.917 value of R^2 still suggests a high fit and that 91.7% of the variation in fuel consumption applied to drivers can be explained by the developed model. In the following table are presented the coefficients obtained as well as some statistical information for each variable:

Table 4 - Statistical coefficients of the selected variables of the drivers model

Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	1,965	,024		82,230	,000		
Logarithm of commercial speed	-,145	,020	-,161	-7,360	,000	,360	2,779
Percentage of mini bus use	-,005	,000	-,845	-55,150	,000	,736	1,359
Percentage of standard bus use	-,002	,000	-,561	-33,719	,000	,624	1,604
Percentage of midi bus use	-,002	,000	-,325	-21,222	,000	,735	1,360
Percentage of inclination superior to 5%	,002	,000	,310	14,390	,000	,373	2,685
Average length of course	-,002	,000	-,132	-4,906	,000	,237	4,219
Percentage of 1007 event occurrence[2]	,003	,001	,077	5,689	,000	,943	1,060

In all cases the results are significant at 1%. Since the normal measure used is a significance level of 5% the model provides very strong results.

[2] 1007 event - Sudden longitudinal decelerations.

5.3. Vehicles

Algebraically the results are presented below:

$$\log(Consumption) = -1,737+0,776\log(\text{VeicMass})-0,006\text{ComSpeed}-0,003\text{AvrgLenght}+0,017\text{Ev1067}+0,195\log(\text{DriversAge}) \quad (3)$$

The results of the model, based on 105 observations, are presented in the table below:

Table 5 – Summary of the vehicles model

R	R square	Adjusted R square	Standard Error of the Estimate
,977[a]	,954	,952	,02393

An R^2 of 0.954 suggests a very high fit: 95.4 % of the variation in fuel consumption of the bus lines can be explained by the developed model. In the following table are presented the coefficients obtained as well as some statistical information for each variable:

Table 6 - Statistical coefficients of the selected variables of the vehicles model

Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
Constant	-1,737	,141		-12,311	,000		
Logarithm of vehicle's mass	,776	,020	,912	38,079	,000	,805	1,242
Commercial speed	-,006	,001	-,195	-4,735	,000	,273	3,666
Average length of course	-,003	,001	-,140	-3,325	,001	,261	3,831
Percentage of 1067 event occurrence[3]	,017	,004	,108	4,838	,000	,931	1,075
Logarithm of the average driver's age	,195	,077	,056	2,516	,013	,927	1,079

In all cases the results are significant at 1%, except for the logarithm of the average driver's age, which is significant for 2%. Once the normal measure used is a significance level of 5% the model provides very strong results.

6. Discussion on regression analysis

The models developed in this research aim to serve *Rodoviária de Lisboa* in the forecasting of the fuel consumption of a bus-line, driver or vehicle by quantifying some variables' parameters. This can be used either to detect odd cases that show fuel consumption values that are too different from the predicted values but also to define new strategies that can optimize the use of the company resources. In this chapter the selected variables for each model will be described.

[3] 1067 event - Rotations above the stipulated maximum value (1900rpm for articulated bus and 2000rpm for the rest).

6.1. Bus-lines

6.1.1 Type of vehicle

The selected variables in this model to describe the type of vehicle were the percentage of mini-bus, percentage of midi-bus and percentage of articulated bus. The variable relative to mini-bus has the biggest weight in fuel consumption calculation. For 1% increase on the percentage of mini-bus and midi-bus the fuel consumption can decrease 0.33% and 0.08%, respectively. For 1% increase in the percentage of articulated-bus the fuel consumption can increase about 0.14%.

6.1.2 Commercial Speed

The commercial speed was selected as one of the most important variables in this model. The unstandardized coefficient of -0.005 indicates that the increase of commercial speed causes fuel consumption to decrease. The elasticity can go up to 0.19, confirming that this is not only an important variable to the passengers satisfaction with the service but is also a good parameter to control fuel consumption.

6.1.3 Average number of monitoring sessions

This variable shows a standardized coefficient of -0.117. The selection of the average number of monitoring sessions confirmed the importance that must be assigned to this initiative. The elasticity can vary up to 0.07.

6.1.4 Percentage of course with a slope greater than 5%

This variable has a maximum elasticity of 0.045 and a standardized coefficient of +0.104 which means the increase of the percentage of courses with high slopes also increases the fuel consumption. This result is in line with other studies referred in the state of the art.

6.1.5 Maximum distance between stops

This variable was transformed into its logarithm so the elasticity from a log-log model is equal to the unstandardized coefficient, +0.019. As expected, this coefficient is positive, once it tends to increase de commercial speed too.

6.1.5 Percentage of occurrence of the event 1040 (air loading with excessive rotation)

This event was selected as influential to the consumption increase with a standardized coefficient of +0.049, a positive coefficient as it was expected too. Its maximum elasticity is 0.04, which is a low value but still has some weight to fuel consumption prediction. This result should be communicated to bus drivers, given its importance to the fuel consumption reduction.

6.2. Drivers

6.2.1 Type of vehicle

The type of vehicle was described in this model with three variables: mini-bus, midi-bus and standard bus. The maximum elasticity obtained was 0.5 for midi-bus, 0.2 for mini-bus and 0.15 for standard bus. The unstandardized coefficients are negative for all the variables.

6.2.2 Percentage of course with a slope greater than 5%

This variable has a maximum elasticity of 0.11 and a standardized coefficient of +0.310 which means the increase of the percentage of courses with high slopes also increases the fuel consumption. The importance of this result relates mainly to the definition of new bus lines.

6.2.3 Commercial Speed

The logarithm of commercial speed was also selected in this model. The standardized coefficient is -0.161 and indicates that the increase of commercial speed causes fuel consumption to decrease. Since the variable was transformed into its logarithm, the elasticity has the same value as the unstandardized coefficient, which is -0.145.

6.2.4 Average length of course

The average length of course has a standardized coefficient of -0.132. The maximum elasticity is about 0.045 which shows that this variable has a low elasticity.

6.2.5 Percentage of occurrence of the event 1007 (sudden longitudinal deceleration)

This event was select as influential to the consumption increase with a standardized coefficient of +0.077. This variable also represents the aggressiveness of the driving style and that is probably the reason for being selected as a factor that increases the fuel consumption. Its maximum elasticity is about 0.035.

6.3. Vehicles

6.3.1 Type of vehicle

The type of vehicle was described in this model with only one variable: the vehicle mass logarithm. The unstandardized coefficient is 0.776 and the standardized coefficient is 0.912. Since the mass was transformed into its logarithm, the elasticity is the same as the unstandardized coefficient. The coefficients are positive like it was expected and according to the other two models.

6.3.2 Commercial Speed

The commercial speed was selected in the three models and always as one of the most important variables. The standardized coefficient is -0.915 and the elasticity can go up to 0.21, representing a strong variable in the model since the only variable with higher values is the vehicle mass.

6.3.3 Average length of course

The average length of course is present in two of the three models. In the present model, this variable has a standardized coefficient of -0.140 with a maximum elasticity of about 0.07, so it has almost twice of the elasticity present in the previous model. This variable has a big correlation with commercial speed, so it's expected that the coefficients are both negative.

6.3.4 Percentage of occurrence of the event 1067 (rotations above the stipulated maximum value)

This variable has 0.108 of standard coefficient and a maximum 0.07 elasticity. Like it was expected, the rotations above the stipulated limit increase the fuel consumption, since they represent more effort to the vehicle's motor. From now on, the company should pay particular attention to this event in the monitoring sessions.

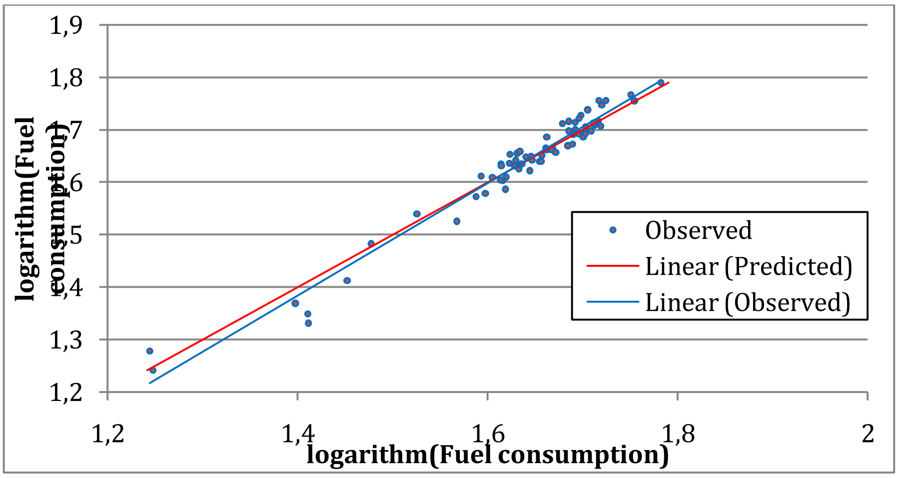
6.3.5 Average driver's age

The average driver's age was also transformed into its logarithm. It has a standardized coefficient of 0.056 with a maximum elasticity of about 0.195, the same as the unstandardized coefficient. The selection of this variable confirms what was already expected by the company itself.

7. Models validations

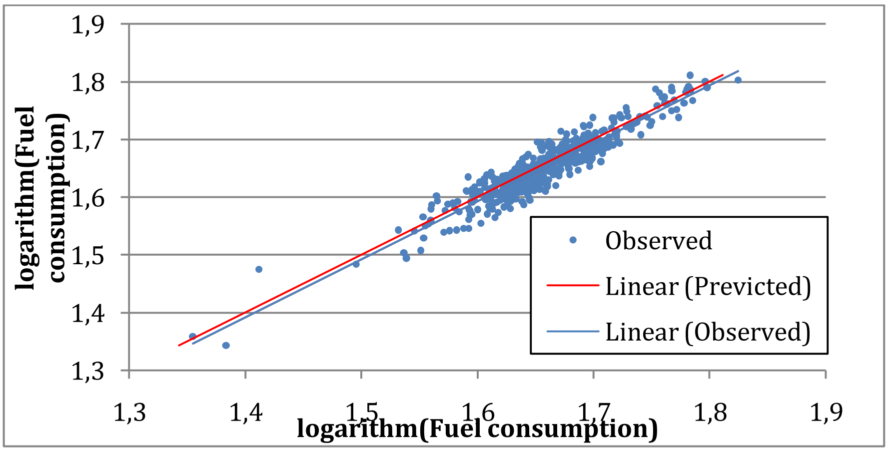
The gathered data, in the months of May 2010 and September 2010, was introduced, for validation purposes only, in the regression models developed. In the next graphics, it's shown the comparison between the predicted and observed values for logarithmic fuel consumption.

The determination coefficient for the bus lines validation sample was R^2 = 0.953. This is a very high relationship with the developed model.



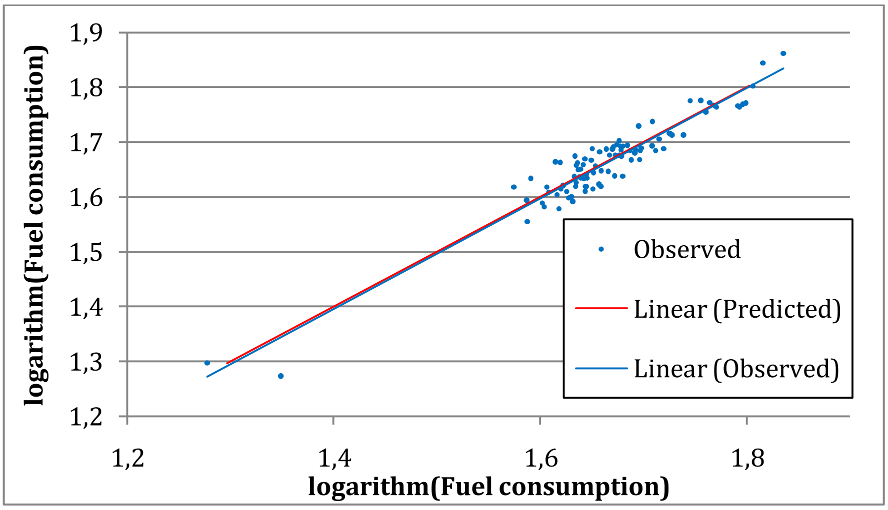
Picture 1 - Comparison between predicted and observed values of fuel consumption logarithm for the bus lines model.

The determination coefficient for the drivers validation sample was R^2 = 0.846. Despite being the lower determination coefficient, it still represents a very good explanation of the reality.



Picture 2 - Comparison between predicted and observed values of fuel consumption logarithm for the drivers model.

The determination coefficient for the vehicles validation sample was R^2 = 0.902, which translates into a very important relationship with the developed model.



Picture 3 - Comparison between predicted and observed values of fuel consumption logarithm for the vehicles model.

8. Conclusions

This research presents macroscopic fuel consumption models that require varied types of variables as input, in order to most accurately characterize the dependent variable. The development of these models attempts to be a tool for transportation planning and management. All the 3 models presented good adjustment indicators, being the lowest R^2 equal to 0,91. The set of variables that proved to be more influent defining the fuel consumption were always those referring to the type of vehicle performing the services. Besides this, the commercial speed was also present in the three models, as one of the most influential variables in explaining fuel consumption (with a negative sign). Other important variables included in the models comprised the number of formative monitoring sessions with bus drivers (revealing a tendency to reduce consumption), the average age of the drivers (higher the driver age tends to mean higher consumption). Also some driving events were detected as prejudicial to fuel consumption, namely the high rotations and the abrupt braking. Lines with important grades tended to have higher consumption levels.

These results provide important policy results that could help guiding *Rodoviária de Lisboa,* to implement and justify the efficiency of measures aimed at reducing the consumption levels of their bus lines.

9. Limitations

There were some limitations detected in the collecting of the necessary data process to develop the regression models. These limitations are described in the following topics:

- The average fuel consumption daily measures have a low accuracy due to human errors registering the data but also to imprecise measurements of the tank filling, so this methodology should be revised;
- The variables developed for slope don't take into account the direction of movement or whether the inclination is negative or positive. This isn't very precise but it was the chosen method due to difficulties inherent to the separation of movement and inclination directions in several variables;
- The instantaneous number of transported passengers should be known in order to precisely define the real mass being transported;

- There are no Midi vehicles in the data collected for the Vehicles regression model which may disqualify the model for predicting the fuel consumption for this type of vehicle;
- The results should not be directly applied to other companies without a previous validation of their applicability despite the strong correlation with the reality in the 3 developed models.

10. Further work

In order to continue improving the energy efficiency of the service provided by *Rodoviária de Lisboa*, it's important to make some proposals considered relevant in this process:

- Improve the accuracy of the data collected;
- With the developed models detect the odd cases that differ greatly from the results expected;
- Evaluate the cost/benefit of the energy conservation measures;
- Produce similar models for each transport activity center unit;
- Define objectives for drivers' monitoring sessions, in order to stimulate their development and the reduction of fuel consumption;
- Make periodic validation of the experimental results obtained for each model.

11. References

WASHINGTON, Simon P. e outros (2003), *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, Albuquerque.

UITP (2009), *Bus Systems: An efficient mode of transport*, UITP's World Congress and Exhibition, Viena.

TABACHNICK, Barbara G. e Fidell, Linda S. (2007), *Using Multivariate Statistics*, 5th Edition, Pearson Education Inc., Boston.

SIMÕES, A. (2005), Metodologia para Auditorias Energéticas a Frotas de Autocarros Urbanos, Dissertation to obtain Master Degree in Transports, Instituto Superior Técnico, Lisbon.

RL (2008), *Plano de Racionalização Energética da Rodoviária de Lisboa, S.A. Relatório de Progresso anual*, Lisbon, September 8.

MAROCO, João (2007), *Análise estatística com utilização do SPSS*, 3ª Edição, Edições Sílabo, Lisboa.

KOSHAL, R. K. (1970), *Economics of Scale in Bus Transport: Some United States Experience*, Journal of Transport Economics and Policy, 4, pp. 29-36.

HAIR, Joseph F. and others (1995), *Multivariate Data Analysis: With Readings*, 4ª Edição, Prentice Hall College Div.

COHEN, Jacob and others (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd Edition, LEA, New Jersey.

VOLVO, Consumo de combustível, Acessed: 20.July.2010, in: http://www.volvotrucks.com/trucks/portugal-market/pt-pt/trucks/environment/pages/fuel_consumption.aspx.