Predicting Environmental Contribution to Ship’s Average Fuel Consumption Based on Non-Uniform Time Set

Igor Vujović 1,*, Ivica Kuzmanić 1, Miro Petković 1 and Joško Šoda 1

1 University of Split; Faculty of Maritime Studies, Ruđera Boškovića 37, 21000 Split, Croatia, ivujovic@pfst.hr
* Correspondence: ivujovic@pfst.hr

Abstract: Real data obtained from ship in operation are processed and analyzed in this paper. The intention was to provide software which would predict ship’s fuel consumption in some future time instant. It is showed that it is possible to develop such software based on numeric fitting of known data. In order to check how well the prediction of future fuel consumption is, we used only the first half of data for obtaining prediction curve. The second part of data was used to compare different prediction curves goodness. Hence, the presented research used actually a “real future data” and forecasted future data, which are used to numerically evaluate goodness of prediction. The research is of interest for companies logistics, to provide adequate fuel for fleet when and where actually needed. It is concluded that there are several prediction functions which satisfy used statistical quality measures.

Keywords: forecasting of fuel demand; ship’s fuel consumption; data fitting; statistical quality measures; signal processing and analysis

1. Introduction

Contribution of the ship’s environment to the fuel consumption is investigated in the paper. During exploitation, ship is under influence of sea vegetation and microorganisms. These natural phenomena result in growing layers at ship’s hull. Layers impact the speed and fuel consumption of the ship. This is not the only contributing factor. Additional factors could be meteorological, cargo, status of ship’s engines, etc. Therefore, only average trends could be used for prediction of fuel consumption due to growth layers, which are the consequence of environmental contribution. It is interesting topic due to increase efforts in energy efficiency, which includes hybrid technologies [1, 2]. In order to measure reduction of fuel consumption, a models for prediction of classic fuel consumption should be incorporated in the researches [2].

Although there are a lot of references regarding forecasting of fuel demand [3 - 6], there were mostly motivated by energy efficiency and ecology [7 - 9] or forecasting global demand and/or port demand, and not specific ship. Fuel consumption was considered for vessel design [10, 11]. Vessel speed was used for assessment of ship’s fuel consumption in [12]. New trends in this field are in several branches. The effect of biofouling on ship resistance using CFD (Computational Fluid Dynamics) was a topic of [13]. Estimation of the vehicle fuel consumption was a scope of [14]. Authors in [15] used a fuel consumption model based on the Vehicle Specific Power distribution. The traffic condition prediction was linked with the fuel consumption model to predict fuel consumption. “The fuel consumption provided by the On-Board Diagnostic data is used to verify the proposed application, and the forecasting error is less than 20%” [15]. A statistical approach to the ship fuel consumption is presented in [16]. This research could be extended by a speculative approach: is it
possible to subtract model calculated consumption from the total consumption and obtain an environmental impact?

In this paper, average daily fuel consumption is used to forecast ship’s fuel consumption in future. The research was initialized when shipping company approached with idea to forecast ship’s fuel consumption by non-linear method. The paper is organized as follows. The second section defines measures of quality and fitting curves used in the research. The third section presents results. Results are obtained by known data which are not for every consequential day, but for irregular day sequence. Finally, conclusions are given.

2. Curve Fitting and Prediction

In order to compare how good some function fits the real data, several quality measures are used: SSE (The sum of squares due to error, see (2)), R-square (see (6)), RMSE (root mean square error, see (7)) and MSE (Mean squared normalized error, see (1)). The sum of squares of the regression (SSR, see (3)) and the total sum of squares (SST, see (4)) are used to calculate R-square. Measures are defined with help of equations (1) to (7):

\[
MSE = \frac{\| X - Y \|^2}{N},
\]

\[
SSE = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2,
\]

\[
SSR = \sum_{i=1}^{n} w_i (\hat{y}_i - \bar{y})^2,
\]

\[
SST = \sum_{i=1}^{n} w_i (y_i - \bar{y})^2,
\]

\[
R - square = \frac{SSR}{SST} = 1 - \frac{SSE}{SST},
\]

\[
RMSE = \sqrt{MSE},
\]

where \( x, y, X, Y \) denotes signals that are compared, \( w \) weights, and “\( \hat{.} \)” and “\( \bar{.} \)” means estimate and average respectively. In order to fit data, several functions were used. There are expressed in equations (8) to (24), where \( x \) is independent variable, \( \theta \) cycle frequency, \( a, b, c, p, q \) coefficients:

\[
f_1(x) = a \cdot e^{bx},
\]

\[
f_2(x) = a \cdot e^{bx} + c \cdot e^{dx},
\]

\[
f(x) = a_0 + a_1 \cdot \cos(x \cdot \omega) + b_1 \cdot \sin(x \cdot \omega),
\]

\[
f(x) = a_0 + a_1 \cdot \cos(x \cdot \omega) + b_1 \cdot \sin(x \cdot \omega) + a_2 \cdot \cos(2 \cdot x \cdot \omega) + b_2 \cdot \sin(2 \cdot x \cdot \omega),
\]

\[
f(x) = a_1 \cdot e^{[x - \theta_1] / \alpha_1^2},
\]
\[ f(x) = a_1 \cdot e^{-(x-b_1)c_1^2} + a_2 \cdot e^{-(x-b_2)c_2^2}, \]  
\[ f(x) = a_1 \cdot e^{-(x-b_1)c_1^2} + a_2 \cdot e^{-(x-b_2)c_2^2} + a_3 \cdot e^{-(x-b_3)c_3^2}, \]
\[ f(x) = p_1 \cdot x + p_2, \]
\[ f(x) = p_1 \cdot x^2 + p_2 \cdot x + p_3, \]
\[ f(x) = p_1 \cdot x^3 + p_2 \cdot x^2 + p_3 \cdot x + p_4, \]
\[ f(x) = a \cdot x^b, \]
\[ f(x) = a \cdot x^b + c, \]
\[ f(x) = \frac{p_1 \cdot x + p_2}{x + q_1}, \]
\[ f(x) = \frac{p_1 \cdot x^2 + p_2 \cdot x + p_3}{x + q_1}, \]
\[ f(x) = \frac{p_1 \cdot x^3 + p_2 \cdot x^2 + p_3 \cdot x + p_4}{x + q_1}, \]
\[ f(x) = \frac{p_1 \cdot x^5 + p_2 \cdot x^4 + p_3 \cdot x^3 + p_4 \cdot x^2 + p_5 \cdot x + p_6}{x^3 + q_1 \cdot x^2 + q_2 \cdot x + q_3}, \]
\[ f(x) = \frac{p_1 \cdot x^3 + p_2 \cdot x^2 + p_3 \cdot x + p_4}{x^2 + q_1 \cdot x + q_2}, \]

3. Results

Acquired data have 829 time samples, which are not uniformly sampled. There are obtained from single ship (5.10.2008 – 28.8.2012) and consists of fuel consumption in 24 hours. We used first 415 samples to obtain fit function. Following 414 samples were used to conclude which fitting function is better for prediction of future fuel consumption.

For fitting function described by (8), it is found that coefficients (with 95% confidence bounds) are: \( a = 29.41 \) (or alternatively 28.64, 30.18), \( b = -5.57 \cdot 10^{-5} \) (or alternatively -0.0001654, 5.395 \cdot 10^{-5} ), and the test function is therefore:

\[ f(x) = 28.64 \cdot e^{-0.0001654 x}, \]

Coefficients (with 95% confidence bounds) for function (9) are: \( a = -1.968 \cdot 10^5 \) (or alternatively -1.731 \cdot 10^5, 1.731 \cdot 10^5), \( b = -0.001531 \) (or -131.3, 131.3), \( c = 1.968 \cdot 10^5 \) (or alternatively -1.731 \cdot 10^5, 1.731 \cdot 10^5), \( d = -0.00153 \) (or -131.3, 131.3), and we have:

\[ f(x) = -1.96 \cdot 10^5 \cdot e^{-0.001531 x} + 1.968 \cdot 10^5 \cdot e^{-0.00153}, \]

Coefficients (with 95% confidence bounds) for function (10) are: \( a_0 = 29.07 \) (or 28.7, 29.44), \( a_1 = 1.968 \cdot 10^5 \) (or 1.731 \cdot 10^5, 1.731 \cdot 10^5), \( b_1 = 0.9755 \) (or 0.0644, 1.887), \( a_0 = 0.07593 \) (or 0.07328, 0.07858). Hence, (10) is:
\[ f_{\text{f1}} = 29.07 + 1.338 \cdot \cos(0.07593t) + 0.9755 \cdot \sin(0.07593x), \]  
(27)

Coefficients (with 95% confidence bounds) for function (11) are: \( a_0 = 29.01 \) (or 28.65, 29.38), \( a_1 = 0.4596 \) (or -0.24, 1.159), \( b_1 = 1.503 \) (or 0.9766, 2.03), \( a_2 = -0.8715 \) (or alternatively -1.497, -0.2459), \( b_2 = 0.5525 \) (or -0.1905, 0.8027). So, (28) is obtained:

\[ f_{\text{f2}} = 29.01 + 0.4596 \cdot \cos(0.07871t) + 1.503 \cdot \sin(0.07871t) - 0.8715 \cdot \cos(2 \cdot 0.07871t) \cdot 1.503 \cdot \sin(2 \cdot 0.07871t), \]  
(28)

Coefficients (with 95% confidence bounds) for function (12) are: \( a_1 = 29.61 \) (or 29.05, 30.18), \( b_1 = 18.58 \) (or 13.84, 23.32), \( c_1 = 891.9 \) (or 527.2, 1257), and it is obtained:

\[ f_{\text{g1}} = 29.61 \cdot e^{\left[\frac{1}{(t - 185.8)^2}ight]} + 29.5 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]}, \]  
(29)

Coefficients (with 95% confidence bounds) for function (13) are: \( a_1 = 9.13 \) (or 5.881, 12.16), \( b_1 = 336.1 \) (or 334.3, 337.9), \( c_1 = 6.184 \) (or 3.581, 8.788), \( a_2 = 29.5 \) (or 28.97, 30.02), \( b_2 = 160.2 \) (or 109, 211.4), \( c_2 = 6.184 \) (or 3.581, 8.788), \( a_3 = 29.66 \) (or 29.12, 30.2), \( b_3 = 206.5 \) (or 184, 229), \( c_3 = 612.6 \) (or 487.2, 738.1). Hence, it is obtained:

\[ f_{\text{g2}} = 9.13 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]} + 29.5 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]}, \]  
(30)

Coefficients (with 95% confidence bounds) for function (14) are: \( a_1 = 9.019 \) (or 5.852, 12.41), \( b_1 = 336.1 \) (or 334.3, 337.9), \( c_1 = 6.184 \) (or 3.581, 8.788), \( a_2 = 29.5 \) (or 28.97, 30.02), \( b_2 = 18.58 \) (or 16.87, 20.3), \( c_2 = 6.184 \) (or 3.581, 8.788), \( a_3 = 29.66 \) (or 29.12, 30.2), \( b_3 = 206.5 \) (or 184, 229), \( c_3 = 612.6 \) (or 487.2, 738.1). Using numbers, we obtain:

\[ f_{\text{g3}} = 9.019 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]} + 29.5 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]} + 29.66 \cdot e^{\left[\frac{1}{(t - 336.1)^2}\right]}, \]  
(31)

Coefficients (with 95% confidence bounds) for (15) are: \( p_1 = -0.001643 \) (or -0.00483, 0.001545), \( p_2 = 29.41 \) (or 28.65, 30.18). So, the prediction function should be:

\[ f_{\text{poly1}} = -0.001643 \cdot t + 29.41, \]  
(32)

Coefficients (with 95% confidence bounds) for (16) are: \( p_1 = -3.64 \cdot 10^{-5} \) (or alternatively -6.598 \cdot 10^{-5}, -6.829 \cdot 10^{-6}), \( p_2 = 0.0135 \) (or 0.007967, 0.02621), \( p_3 = 28.36 \) (or 27.22, 29.5). Hence, obtained function is:

\[ f_{\text{poly2}} = -3.64 \cdot 10^{-5} \cdot t^2 + 0.0135 \cdot t + 28.36, \]  
(33)

Coefficients (with 95% confidence bounds) for (17) are: \( p_1 = -2.505 \cdot 10^{-7} \) (or -5.309 \cdot 10^{-7}, 2.983 \cdot 10^{-6}), \( p_2 = 0.0001199 \) (or -8.549 \cdot 10^{-5}, 0.0002974), \( p_3 = -0.01254 \) (or -0.001, 0.01924), \( p_4 = 29.27 \) (or 27.74, 30.8). So, it is obtained:

\[ f_{\text{poly3}} = -2.505 \cdot 10^{-7} \cdot t^3 + 0.0001199 \cdot t^2 - 0.01254 \cdot t + 29.27, \]  
(34)

Coefficients (with 95% confidence bounds) for (18) are: \( a = 29.17 \) (or 27.14, 31.19), \( b = -0.006518 \) (or -0.01418, 0.01287). So, it is obtained:

\[ f_{\text{pow1}} = 29.17 \cdot t^{-0.006518}, \]  
(35)

Coefficients (with 95% confidence bounds) for (19) are: \( a = -3.114 \cdot 10^{-15} \) (-1.787 \cdot 10^{-15}, 1.725 \cdot 10^{-15}), \( b = 5.659 \) (or -3.779, 15.1), \( c = 29.38 \) (or 28.85, 29.9), and (19) becomes:

\[ f_{\text{pow2}} = -3.114 \cdot 10^{-15} \cdot t^{5.659} + 29.38, \]  
(36)

Coefficients (with 95% confidence bounds) for (20) are: \( p_1 = 29.08 \) (or 28.7, 29.47), \( p_2 = -57.03 \) (or -83.42, -30.64), \( q_1 = -1.954 \) (or -3.031, -0.8765), and we have:

\[ f_{\text{rat1}} = \frac{29.08 \cdot t - 57.03}{t - 1.954}, \]  
(37)
Coefficients (with 95% confidence bounds) calculated for (21) are: \(p_1 = -0.002987\) (or \(-0.00688, 0.000905\)), \(p_2 = 29.83\) (or 28.78, 30.88), \(p_3 = 12.44\) (or -129.1, 154), \(q_1 = 0.756\) (or -4.772, 6.284). Hence, the prediction can be obtained by:

\[
f_{\text{rat}21} = \frac{-0.002987 \cdot t^2 + 29.83 \cdot t + 12.44}{t + 0.756},
\]

(38)

Coefficients (with 95% confidence bounds) for (22) are: \(p_1 = -3.957 \times 10^{-5}\) (or \(-6.909 \times 10^{-5}, -1.006 \times 10^{-5}\)), \(p_2 = 0.01591\) (or 0.002675, 0.02915), \(p_3 = 27.88\) (or 26.52, 29.24), \(p_4 = -533.5\) (or -560.3, -506.6), \(q_1 = -18.96\) (or -19.36, -18.55). This type of prediction function is:

\[
f_{\text{rat}31} = \frac{-3.957 \times 10^{-5} \cdot t^3 + 0.01591 \cdot t^2 + 27.88 \cdot t - 533.5}{t - 18.96},
\]

(39)

Coefficients (with 95% confidence bounds) for (23) are: \(p_1 = -0.001815\) (or \(-0.005116, 0.001486\)), \(p_2 = 29.47\) (or 28.64, 30.31), \(p_3 = -271.4\) (or -310.9, -231.8), \(p_4 = 554.9\) (or 342.3, 767.4), \(q_1 = -9.246\) (or -10.85, -7.641), \(q_2 = 18.97\) (or 10.54, 27.41). Prediction function is:

\[
f_{\text{rat}32} = \frac{-0.001815 \cdot t^3 + 29.47 \cdot t^2 - 271.4 \cdot t + 554.9}{t^2 - 9.246 \cdot t + 18.97},
\]

(40)

Coefficients (with 95% confidence bounds) for (24) are: \(p_1 = -6.959 \times 10^{-5}\) (or \(-0.000102, -3.715 \times 10^{-5}\)), \(p_2 = 0.03339\) (or 0.0173, 0.04949), \(p_3 = 25.22\) (or 23.13, 27.3), \(p_4 = -1152\) (or -1296, -1008), \(p_5 = 1.472 \times 10^4\) (or 1.136 \times 10^4, 1.808 \times 10^4), \(p_6 = -3.457 \times 10^4\) (or -5.92 \times 10^4, -9944), \(q_1 = -43.6\) (or -47.65, -39.55), \(q_2 = 542.5\) (or 423.6, 661.4), \(q_3 = -1266\) (or -2241, -291). So, prediction is obtained with:

\[
f_{\text{rat}53} = ( -6.959 \times 10^{-5} \cdot t^5 + 0.03339 \cdot t^4 + 25.22 \cdot t^3 - 1152 \cdot t^2 + 
+ 1.472 \cdot 10^4 \cdot t - 3.457 \cdot 10^4 ) / (t^3 - 43.6 \cdot t^2 + 542.5 \cdot t - 1266),
\]

(41)

Matlab’s piecewise polynomial was also used as the fitting function. It is computed from \(p\)-order smoothing parameter, where \(x\) is normalized by mean 208 and standard deviation 119.9. Smoothing parameter was \(p = 0.99997739\), and other Matlab data: form 'pp', breaks: [1x415 double], coefficients: [414x4 double], pieces: 414, order: 4, dimension: 1. Table 1 shows results of fit for the first half of total data (415 samples).

<table>
<thead>
<tr>
<th>Fitness function</th>
<th>SSE</th>
<th>R-square</th>
<th>RMSE</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential of 1st order</td>
<td>6468</td>
<td>0.002443</td>
<td>3.958</td>
<td></td>
</tr>
<tr>
<td>Exponential of 2nd order</td>
<td>6393</td>
<td>0.01413</td>
<td>3.944</td>
<td></td>
</tr>
<tr>
<td>Fourier 1</td>
<td>5915</td>
<td>0.08784</td>
<td>3.794</td>
<td></td>
</tr>
<tr>
<td>Fourier 2</td>
<td>5733</td>
<td>0.1159</td>
<td>3.744</td>
<td></td>
</tr>
<tr>
<td>Gauss 1</td>
<td>6378</td>
<td>0.01636</td>
<td>3.935</td>
<td></td>
</tr>
<tr>
<td>Gauss 2</td>
<td>5785</td>
<td>0.1078</td>
<td>3.761</td>
<td></td>
</tr>
<tr>
<td>Gauss 3</td>
<td>5215</td>
<td>0.1958</td>
<td>3.584</td>
<td>The runner up</td>
</tr>
<tr>
<td>Polynomial 1</td>
<td>6468</td>
<td>0.002478</td>
<td>3.957</td>
<td></td>
</tr>
<tr>
<td>Polynomial 2</td>
<td>6378</td>
<td>0.01645</td>
<td>3.934</td>
<td></td>
</tr>
<tr>
<td>Polynomial 3</td>
<td>6330</td>
<td>0.02378</td>
<td>3.924</td>
<td></td>
</tr>
<tr>
<td>Power 1</td>
<td>6484</td>
<td>2.212 \times 10^5</td>
<td>3.962</td>
<td>SSE and R-square are the worst</td>
</tr>
<tr>
<td>Power 2</td>
<td>6382</td>
<td>0.01573</td>
<td>3.936</td>
<td></td>
</tr>
</tbody>
</table>
Examples of fitting are shown in Fig. 1. Table 1 suggest that the best choice is smoothing spline. However, Fig. 1 shows that this function does not results in some regular and understandable prediction, as i.e. linear case.
c) Fitting by polynomial of the first degree for all data

d) Fitting by Gaussian function for the first 415 samples
Figure 1. Examples of the results for the data fitting: (a) smoothing spline for the first half of data, (b) smoothing spline for the entire dataset, (c) linear fitting for the first half of data, (d) linear fitting for the entire dataset, (e) Gaussian fitting for the first half of data, (f) Gaussian fitting for all data.

Table 2 shows MSE for data used for fitting (the first column), predicted “future” values (the second column), and total for all samples (the third column). Comments to some results are given in the last column of the table.

Table 2. MSE (obtained by Matlab function “mse”) for domain used to fit, prediction interval, and total range.

<table>
<thead>
<tr>
<th>Fitness function</th>
<th>Fit-domain (samples 1 – 415)</th>
<th>Prediction interval (samples 416 – 829)</th>
<th>Total range (samples 1 – 829)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential of 1st order</td>
<td>2.8873 e-103</td>
<td>2.2868e-103</td>
<td>2.5874e-103</td>
<td>Almost 0</td>
</tr>
<tr>
<td>Exponential of 2nd order</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Fourier 1</td>
<td>845.9130</td>
<td>837.9584</td>
<td>841.9405</td>
<td></td>
</tr>
<tr>
<td>Fourier 2</td>
<td>854.2480</td>
<td>839.5154</td>
<td>846.8906</td>
<td></td>
</tr>
<tr>
<td>Gauss 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Gauss 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Gauss 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Polynomial 1</td>
<td>1.3844e+06</td>
<td>1.3871e+06</td>
<td>1.3858e+06</td>
<td>Obvious computational problem in function</td>
</tr>
<tr>
<td>Polynomial 2</td>
<td>3.8426e+14</td>
<td>3.8574e+14</td>
<td>3.8500e+14</td>
<td>Obvious computational problem in function</td>
</tr>
<tr>
<td>Polynomial 3</td>
<td>9.8025e+21</td>
<td>9.8592e+21</td>
<td>9.8308e+21</td>
<td>Obvious computational problem in function</td>
</tr>
<tr>
<td>Power 1</td>
<td>836.0386</td>
<td>836.0375</td>
<td>836.0380</td>
<td></td>
</tr>
</tbody>
</table>
4. Discussion and Conclusions

This research is induced by a maritime company, which would like to know how to predict expenses and supply of their ships. We expected a practical result. By academic measures, there are a lot of points for discussion.

If the smoothing spline is discarded, which fits from point to point, it can be concluded that fuel consumption is almost linear. Such data can be fitted with many curves. The best fit of the forecasted data (samples 416 - 829) are for Gauss functions, second order exponential function, and rational functions generally tending to be shortly represented by the first order polynomial (which is actually linear dependence at limited time span).

In this research, intuition was telling that fuel consumption will grow over time by linear law. However, we obtain decrease in consumption, which could be in problems regarding:

- non-uniform time sampling (which leads to wrong curve angle between interpolating points), and
- average (which depends on route ship was sailing at time of data acquisition and sailing hours in exact day – such precise data is almost impossible to obtain without strong financial support).

Further research should include identification of factors that causes unplanned fuel consumption and modelling fuel consumption by artificial neural networks (ANN) and/or hybrid model of velocity and fuel consumption.

Author Contributions: conceptualization, Igor Vujović; methodology, Igor Vujović and Ivica Kuzmanić; software, Igor Vujović and Miro Petković; validation, Joško Šoda and Igor Vujović; formal analysis, Ivica Kuzmanić, and Igor Vujović; investigation, Igor Vujović and Miro Petković; writing—original draft preparation, Igor Vujović; visualization, Igor Vujović and Miro Petković; supervision Igor Vujović.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References


