

Article

Predicting External Influences to Ship's Average Fuel Consumption Based on Non-Uniform Time Set

Igor Vujović , Joško Šoda , Ivica Kuzmanić  and Miro Petković

Faculty of Maritime Studies, University of Split, Ruđera Boškovića 37, 21000 Split, Croatia; jsoda@pfst.hr (J.Š.); ikuzman@pfst.hr (I.K.); mpetkovic@pfst.hr (M.P.)

* Correspondence: ivujovic@pfst.hr

Received: 3 July 2020; Accepted: 14 August 2020; Published: 19 August 2020



Abstract: Nowadays, the impact of the ships on the World economy is enormous, considering that every ship needs fuel to sail from source to destination. It requires a lot of fuel, and therefore, there is a need to monitor and predict a ship's average fuel consumption. However, although there are much models available to predict a ship's consumption, most of them rely on a uniform time set. Here we show the model of predicting external influences to ship's average fuel consumption based on a non-uniform time set. The model is based on the numeric fitting of recorded data. The first set of recorded data was used to develop the model, while the second set was used for validation. Statistical quality measures have been used to choose the optimal fitting function for the model. According to statistical measures, the Gaussian 7, Fourier 8, and smoothing spline fitting functions were chosen as optimal algorithms for model development. In addition to extensive data analysis, there is an algorithm for filter length determination for the preprocessing of raw data. This research is of interest to corporate logistics departments in charge of ensuring adequate fuel for fleets when and where required.

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

1. Introduction

The effects of the marine environment and other causes on fuel consumption can be examined by various parameters and various approaches. Approaches could be model-based, which is usual in references or signal-based, which is considered in this paper. Model-based approaches tend to estimate the exact fuel consumption in the exact engine operating mode, examples of this are [1,2]. Such models are useful in, e.g., dual-fuel engines, when it is possible to calculate how much of fuels are being spent [1]. ANNs (Artificial Neural Network) in [2] need to be trained and re-trained for specific ships over time. However, there is no exact real-time monitoring of the hull status, which could be a problem that increases the error in the calculation during time between maintenances. Namely, during exploitation and over time, the ship hull is affected by fouling—a natural phenomenon where marine/aquatic vegetation and microorganisms attach to the hull, creating bio-layers that have an impact on the ship's speed and fuel consumption. Other contributing factors may include weather conditions, cargo, propulsion, engine conditions, etc. Data trend analysis is a suitable approach for predicting fuel consumption, depending on the biofouling layers as environmental contributors. In light of current increased efforts to improve energy efficiency, the above-mentioned topic is current, especially since it includes the use of hybrid technologies [3,4]. As stated in [4], fuel consumption reduction cannot be established without first exploring standard fuel consumption prediction models.

Instead of focusing on a specific ship, there are a number of references to fuel demand prediction [5–8] in which the focus is on energy efficiency and ecology [9–11], or predicting global

demand and port demand. The dependence of fuel consumption on vessel design was explored in [12,13], and on vessel speed in [14]. There are new trends in addressing the fuel consumption issue that could be divided into several research branches. While the effect of biofouling on ship resistance using CFD (Computational Fluid Dynamics) was the research topic in [15], vessel fuel consumption prediction was examined in [16], and the authors in [17] used a fuel consumption model based on the Vehicle Specific Power distribution. Traffic condition prediction was linked with the fuel consumption model to predict fuel consumption. “Fuel consumption data provided by the On-Board Diagnostic tool was used to verify the proposed application, with a prediction error under 20%” [17]. A statistical approach to the ship’s fuel consumption was presented in [18]. This research could be extended to a speculative approach—can the calculated consumption of the model be subtracted from the total consumption to obtain its environmental impact? Learning approaches to ship fuel consumption with ANNs are the most popular, and include [19–23]. On-line fuel consumption prediction was obtained by machine learning in [19]. Shallow and deep learning was combined in [22]. An outstanding result was the correlation matrix in [20], which correlated various causes of fuel consumption increase (e.g., wind, trim, currents, cargo, etc.). An interesting case study was published in [24], which explored AI (Artificial Intelligence) driven tools to identify fuel usage anomalies across the company’s entire fleet. The gray box model was applied to optimize trim and identify the possibility of decreasing fuel consumption [25]. It was shown that this optimization could decrease fuel consumption up to 2.3% [25].

The subject of this paper is to monitor the daily ships fuel consumption seasonal. The main variable used is the daily fuel consumption, which is the main indicator of fuel consumption. Concerning the previously mentioned research problem of ships fuel consumption, the following hypotheses is that the average daily consumption could be predicted seasonally and yearly. The aim of this paper is to find the fitting function to establish the dynamics of the ships daily fuel consumption, using a simple and processing cost-efficient model. The goal of the paper is to find the fitting function that approximates the ship’s daily fuel consumption.

The paper is organized as follows: the second section defines quality measures and fitting curves used in the research; the third section explains the methodology, while the fourth presents the results. The latter are obtained using the known data that do not pertain to consecutive days, but rather cover an irregular day sequence. The results are produced at yearly and seasonal levels, which is a novelty of the paper. Finally, the discussion and conclusions are presented.

2. Mathematical Background of Curve Fitting and Prediction

This section presents the mathematical foundations used in the paper, together with related references. There are several ways to fit the data. As the dataset in the paper is non-uniform, we used several fitting functions, Equations (1)–(20). First, the data were fitted using the Matlab function “linear fitting”, which can be described with the following equation [26]:

$$f_{LF}(x) = a \cdot \sin(x - \pi) + b \cdot (x - 10)^2 + c \tag{1}$$

where x is an independent variable, and a , b , and c are constants that have to be determined. The next data fitting function is the so-called exponential of the 1st order, which can be defined as [26]:

$$f_{e1}(x) = a_{e1} \cdot e^{b_{e1} \cdot x} \tag{2}$$

where x is an independent variable, a_{e1} and b_{e1} are constant coefficients. The data were tested using the so-called exponential of the 2nd order, defined as follows [26]:

$$f_{e2}(x) = a_{e2} \cdot e^{b_{e2} \cdot x} + c_{e2} \cdot e^{d_{e2} \cdot x} \tag{3}$$

where a_{e2} , b_{e2} , c_{e2} , and d_{e2} are coefficients. The next curve fitting functions considered for research are taken from the Fourier series [26]:

$$f_{F\infty}(x) = a_0 + \sum_{i=1}^{\infty} [a_i \cdot \cos(i \cdot \omega \cdot x) + b_i \cdot \sin(i \cdot \omega \cdot x)] \tag{4}$$

where coefficients from (4) are: $a_0 = (2/T) \cdot \int_0^T f(x) \cdot dt$, $a_i = (1/T) \cdot \int_0^T f(x) \cdot \cos(i \cdot \omega \cdot x) \cdot dt$, and $b_i = (1/T) \cdot \int_0^T f(x) \cdot \sin(i \cdot \omega \cdot x) \cdot dt$, $\omega = (2\pi)/T$, and T is the period or data width. In the analysis, the Fourier series of the 1st, 2nd, and 8th order were considered, presented with Equations (5)–(7) as follows:

Fourier series of the first order:

$$f_{F1}(x) = a_0 + a_1 \cdot \cos(\omega \cdot x) + b_1 \cdot \sin(\omega \cdot x) \tag{5}$$

Fourier series of the second order:

$$f_{F2}(x) = a_0 + \sum_{i=1}^2 [a_i \cdot \cos(i \cdot \omega \cdot x) + b_i \cdot \sin(i \cdot \omega \cdot x)] \tag{6}$$

and Fourier series of the eighth order:

$$f_{F8}(x) = a_0 + \sum_{i=1}^8 [a_i \cdot \cos(i \cdot \omega \cdot x) + b_i \cdot \sin(i \cdot \omega \cdot x)] \tag{7}$$

We likewise considered the Gaussian function [26], defined as:

$$f_G(x) = \sum_{i=1}^{\infty} a_i \cdot e^{-[(x-b_i)/c_i]^2} \tag{8}$$

where a_i , b_i , and c_i are Gaussian function coefficients. From (8), a Gaussian function with one term (defined in Matlab as Gaussian 1) was considered in the analysis, which could be defined with the following equation [26]:

$$f_{G1}(x) = a_{11} \cdot e^{-[(x-b_{11})/c_{11}]^2} \tag{9}$$

where a_{11} , b_{11} , and c_{11} are coefficients of the Gaussian function with one term. Next, a Gaussian function with two terms (Gaussian 2 in Matlab) was considered, defined as [26]:

$$f_{G2}(x) = a_{12} \cdot e^{-[(x-b_{12})/c_{12}]^2} + a_{22} \cdot e^{-[(x-b_{22})/c_{22}]^2} \tag{10}$$

where a_{12} , b_{12} , c_{12} , a_{22} , b_{22} , and c_{22} are coefficients of the Gaussian function with two terms. Polynomials-based fitting functions were also considered. First, the polynomial of the first order was defined as follows [26]:

$$f_{P1}(x) = p_{01} + p_{11} \cdot x \tag{11}$$

where p_{01} and p_{11} are first-order polynomial coefficients. Next, the second-order polynomial was defined as [26]:

$$f_{P2}(x) = p_{02} + p_{12} \cdot x + p_{22} \cdot x^2 \tag{12}$$

where p_{02} , p_{12} , and p_{22} are second-order polynomial coefficients. Third-order polynomials were also used in the analysis, that could be defined as [26]:

$$f_{P3}(x) = p_{03} + p_{13} \cdot x + p_{23} \cdot x^2 + p_{33} \cdot x^3 \tag{13}$$

where p_{03} , p_{13} , p_{23} , and p_{33} are third-order polynomial coefficients. All coefficients from (11), (12), and (13) could be calculated by performing the least square method [26]. In the next two equations, (14) and (15), the so-called first and second power model fitting functions were used, which could be described with the following equations [26]:

$$f_{p1} = \alpha_{11} \cdot x^{\beta_{11}} \tag{14}$$

where α_{11} and β_{11} are coefficients from the so-called first-order power model, and [26]:

$$f_{p2} = \delta_{21} + \alpha_{21} \cdot x^{\beta_{21}} \tag{15}$$

where α_{21} , β_{21} , and δ_{21} are coefficients of the so-called second-order power model. Furthermore, Equations (16)–(20) use rational fitting functions, and can be defined as follows; the rational fitting function 1/1 could be expressed as [26]:

$$f_{1/1}(x) = \frac{z_{01} + z_{11} \cdot x}{q_{01} + x} \tag{16}$$

where z_{01} , z_{11} , and q_{01} are coefficients of the rational 1/1 fitting function. The rational fitting function 2/1 can be expressed as follows [26]:

$$f_{2/1}(x) = \frac{z_{02} + z_{12} \cdot x + z_{22} \cdot x^2}{q_{02} + x} \tag{17}$$

where z_{02} , z_{12} , z_{22} , and q_{02} are coefficients of the rational 2/1 fitting function. The rational fitting function 3/1 could be expressed as [26]:

$$f_{3/1}(x) = \frac{z_{03} + z_{13} \cdot x + z_{23} \cdot x^2 + z_{33} \cdot x^3}{q_{04} + x} \tag{18}$$

where z_{03} , z_{13} , z_{23} , z_{33} , and q_{03} are coefficients of the rational 2/1 fitting function. Equation (19) describes the 3/2 rational fitting function as follows [26]:

$$f_{3/2}(x) = \frac{z_{05} + z_{15} \cdot x + z_{25} \cdot x^2 + z_{35} \cdot x^3}{q_{05} + q_{15} \cdot x + x^2} \tag{19}$$

where z_{05} , z_{15} , z_{25} , z_{35} , q_{05} , and q_{15} are coefficients of the rational 3/2 fitting function. Finally, the rational 5/3 fitting function could be expressed as [26]:

$$f_{5/3}(x) = \frac{z_{08} + z_{18} \cdot x + z_{28} \cdot x^2 + z_{38} \cdot x^3 + z_{48} \cdot x^4}{q_{08} + q_{18} \cdot x + q_{28} \cdot x^2 + x^3} \tag{20}$$

where z_{08} , z_{18} , z_{28} , z_{38} , z_{48} , q_{08} , q_{18} , and q_{28} are coefficients of the rational 5/3 fitting function.

In the end, the data were fitted using the smoothing spline s , as in (21), for the specified smoothing parameter p and specified weights w_i [26]. The smoothing spline minimizes the expression:

$$p \cdot \sum_i w_i \cdot (y_i - s(x_i))^2 + (1 - p) \cdot \int \left(\frac{d^2s}{dx^2} \right)^2 \cdot dx \tag{21}$$

If the weights are not specified, they are assumed to be 1 for all the data points. Parameter p is defined between 0 and 1. For $p = 0$, a least-squares straight-line is produced that fits the data. For $p = 1$, a cubic spline interpolant is obtained.

The sheer multitude of fitting functions to be tested with the data requires them to be quantified, and quality measures to be introduced. The following quality measures have been used—the first

quality measure is the root mean square error (RMSE)—a measure that determines the differences between samples or population values predicted by the fitting function—that can be described as follows [27,28]:

$$RMSE = \sqrt{\frac{\|X - Y\|^2}{N}} \tag{22}$$

where X is the observed vector, Y the predicted vector, and N the number of data measured in the observed vector space.

The following three quality measures, as in (23), (24), and (25), are statistical measures that describe the number of variations in the dependent variable. The first quality measure is the sum of the squared estimate errors (SSE), which can be expressed as [27,28]:

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \tag{23}$$

where y_i is the i -th value of the variable to be predicted, and \hat{y}_i the predicted value of y_i . SSE shows the measure of discrepancy between the data and the estimation model. The second quality measure is the total sum of squares (SST), that can be defined as [27–29]:

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2 \tag{24}$$

where y_i is the i^{th} value of the variable to be predicted, and \bar{y} the mean value. SST is defined as a quality measure that equals the sum of the squared differences between observations and their overall mean value. The third quality measure used here was the sum of squares due to regression, that could be defined as [27–29]:

$$SSR = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \tag{25}$$

The following expression is known to be true [27–29]:

$$SST = SSE + SSR \tag{26}$$

Furthermore, the R-square measure could be devised from (23), (24), (25), and (26), and defined as:

$$r^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{27}$$

It is known in the literature [20–22] as the coefficient of determination that describes the proportion of dependent variable variance predictable from the independent variable. In addition, there is an adjusted R-square quality measure, which corrects the possible error in the R-square measure by increasing the number of samples, described as [27–29]:

$$\bar{r}^2 = 1 - (1 - r^2) \cdot \frac{N - 1}{N - k - 1} = 1 - \left(\frac{SSE}{SST}\right) \cdot \frac{N - 1}{N - k - 1} \tag{28}$$

where N is the number of data samples, and k is the number, which explains the independent variable.

In the next paragraph of this section, the random variable theory used throughout the paper will be presented. If we consider random variables \mathbf{X}_{su} , \mathbf{X}_a , \mathbf{X}_w , \mathbf{X}_{sp} in the vector notation:

$$\begin{aligned} \mathbf{X}_{su} &= [x_{su1}, x_{su2}, \dots, x_{sui}] & sui &= 1, 2, 3, \dots, N & sui &\in N \\ \mathbf{X}_a &= [x_{a1}, x_{a2}, \dots, x_{ai}] & ai &= 1, 2, 3, \dots, N & ai &\in N \\ \mathbf{X}_w &= [x_{w1}, x_{w2}, \dots, x_{wi}] & wi &= 1, 2, 3, \dots, N & wi &\in N \\ \mathbf{X}_{sp} &= [x_{sp1}, x_{sp2}, \dots, x_{spi}] & spi &= 1, 2, 3, \dots, N & spi &\in N \end{aligned} \tag{29}$$

where $\mathbf{X}_{su}, \mathbf{X}_a, \mathbf{X}_w, \mathbf{X}_{sp}$ represent random variables with N data samples measured in the summer (su), autumn (a), winter (w), and spring (sp) for four years. Standard statistical metrics such as expectation or average value, standard deviation, standard error, and correlation coefficient were used to study random variables. The average value of the random variable x or expectation could be represented by equation [27,29]:

$$E[x] = \bar{x} = \frac{1}{N} \cdot \sum_{i=1}^N x_i = \frac{1}{N} \cdot (\mathbf{X} \times \mathbf{X}^T) \tag{30}$$

where E represents the expectation for a random variable, is the sign for random variable x , and N is the number of samples measured. The standard deviation of the random variable x could be represented by equation [27–29]:

$$\sigma_x = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (x_i - \bar{x})^2} = \sqrt{E[x^2] - E[x]^2} \tag{31}$$

where σ_x represents the standard deviation of the random variable x . The transformation of coordinates from one coordinate system to another is described by the following equation [27–29]:

$$\begin{aligned} x_a^* &= \frac{(x_{ia} - \bar{x}_a)}{\sigma_{xa}} = \frac{1}{\sigma_{xa}} \cdot \left((\mathbf{X}_{ia} - \bar{\mathbf{X}}_a) \times (\mathbf{X}_{ia} - \bar{\mathbf{X}}_a)^T \right) \\ x_w^* &= \frac{(x_{iw} - \bar{x}_w)}{\sigma_{xw}} = \frac{1}{\sigma_{xw}} \cdot \left((\mathbf{X}_{iw} - \bar{\mathbf{X}}_w) \times (\mathbf{X}_{iw} - \bar{\mathbf{X}}_w)^T \right) \end{aligned} \tag{32}$$

where x_a^* i x_w^* are transformed coordinates of random variables. The statistical metric used to quantify the similarity and dependence between variables, x_a and x_w is the correlation coefficient between the random variables. The correlation coefficient could be calculated using the following equation [27–29]:

$$r = \frac{1}{N-1} \cdot \frac{\sum_{i=1}^N ((x_{ia} - \bar{x}_a) \cdot (x_{iw} - \bar{x}_w))}{\sigma_{xa} \cdot \sigma_{xw}} = \frac{E[(x_a - E[x_a]) \cdot (x_w - E[x_w])]}{\sigma_{xa} \cdot \sigma_{xw}} = \frac{1}{N-1} \times (\mathbf{X}_a^* \times \mathbf{X}_w^{*T}) \tag{33}$$

where r represents the correlation coefficient, N is the number of measurements, while σ_{xa} and σ_{xw} represent standard deviations of the random variables x_a and x_w . Furthermore, a model matrix A could be created using the following equation:

$$\mathbf{A} = \begin{bmatrix} \mathbf{X}_{su} \\ \mathbf{X}_a \\ \mathbf{X}_w \\ \mathbf{X}_{sp} \end{bmatrix} = \begin{bmatrix} x_{su1} & x_{su2} & \cdots & x_{suN} \\ x_{a1} & x_{a2} & \cdots & x_{aN} \\ x_{w1} & x_{w2} & \cdots & x_{wN} \\ x_{sp1} & x_{sp2} & \cdots & x_{spN} \end{bmatrix}_{(4,N)} \tag{34}$$

where $\mathbf{X}_{su}, \mathbf{X}_a, \mathbf{X}_w, \mathbf{X}_{sp}$ are vectors of random independent variables, together with Equations (29)–(33), transformed in the correlation matrix C_x into the equation:

$$C_x = \frac{1}{N-1} \times (\mathbf{A} \times \mathbf{A}^T) = \frac{1}{N-1} \cdot \left(\begin{bmatrix} x_{su1} & x_{su2} & \cdots & x_{suN} \\ x_{a1} & x_{a2} & \cdots & x_{aN} \\ x_{w1} & x_{w2} & \cdots & x_{wN} \\ x_{sp1} & x_{sp2} & \cdots & x_{spN} \end{bmatrix}_{(4xN)} \cdot \begin{bmatrix} x_{su1} & x_{a1} & x_{w1} & x_{sp1} \\ x_{su2} & x_{a2} & x_{w2} & x_{sp2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{suN} & x_{aN} & x_{wN} & x_{spN} \end{bmatrix}_{(Nx4)} \right) \tag{35}$$

In the example from Equation (29), the correlation matrix C_x has dimensions (4×4) , as follows:

$$C_X = \begin{bmatrix} C_{x_{su},x_{su}} & C_{x_{su},x_a} & C_{x_{su},x_w} & C_{x_{su},x_{sp}} \\ C_{x_a,x_{su}} & C_{x_a,x_a} & C_{x_a,x_w} & C_{x_a,x_{sp}} \\ C_{x_w,x_{su}} & C_{x_w,x_a} & C_{x_w,x_w} & C_{x_w,x_{sp}} \\ C_{x_{sp},x_{su}} & C_{x_{sp},x_a} & C_{x_{sp},x_w} & C_{x_{sp},x_{sp}} \end{bmatrix}_{(4,4)} \tag{36}$$

The correlation matrix, C_x , shows correlations between variables.

To conclude, this section covers the mathematical background of the fitting functions, from (1) to (21) that will be used in the prediction. Equations from (22) to (28) describe the quality measures, which will be used to grade the fitting functions and determine which fitting function is best suited to predicting the average fuel consumption based on the non-uniform data time set. Finally, the theory of random variables is introduced in (29) to (36) that will be useful for identifying the optimal fitting function for prediction.

3. Methodology, Setup, and Preprocessing

This section will cover the data smoothing methodology, otherwise known as data preprocessing. Owing to the non-uniform set of data and the need to identify the best-suited fitting function, choosing the “right” window to perform the moving average operation or filtering is of paramount importance. There are a number of moving average algorithms available that can be used for smoothing, such as a simple moving average (SMA), a weighted moving average (WMA), an exponential moving average (EMA), and a weighted exponential moving average method (WEMA). All the mentioned moving average algorithms have their advantages and disadvantages, but they all require choosing optimal filter window size, i.e., the number of past and future points that will determine the current point. The preprocessing procedure can be explained in steps, as shown in Figure 1.

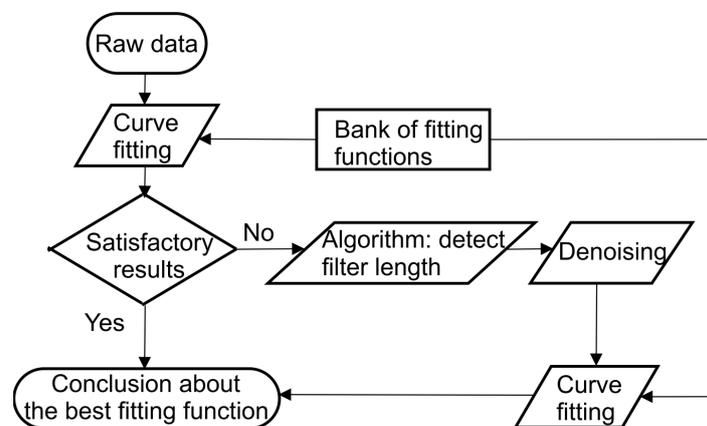


Figure 1. Preprocessing and overall process flow.

The first step in the signal processing is noise removal. Although the noise does not exist physically within the average daily fuel consumption data, it exists from the signal processing point of view. Namely, as there are sudden spikes in fuel consumption, comparing raw data would be somewhat misleading. Instead, the moving average or filtering operation is performed on raw data. The first step is to choose the size of the moving average filter (filter length). It has to be noted that ship data are not uniformly collected. Hence, there is no year, season, or month with an equal number of samples. Smoothing at the season level has been performed. The optimal filter length was identified by developing and performing an algorithm.

First, the R-square analysis of the raw data was performed. In the raw data, R-square is, e.g., 0.01636 for the Gauss 1 fitting function. As a low R-square measure implies a high noise level,

raw data should be preprocessed by eliminating the noise. Since we failed to obtain satisfactory results, we proceeded according to the flow diagram given in Figure 1. First, we detected the filter length necessary for moving the average operation, and then performed the moving average to denoise the raw data and obtain enhanced input to the curve fitting operation.

The influence of filter length can be seen in the correlation matrix. The results for the length of 52 points over four years of the moving average filter are as follows (including trends):

$$\begin{bmatrix} 1.0000 & 0.3987 & 0.3963 & 0.4210 \\ -0.3987 & 1.0000 & 0.3532 & 0.3311 \\ -0.3963 & 0.3532 & 1.0000 & 0.6988 \\ -0.4210 & 0.3311 & 0.6988 & 0.9999 \end{bmatrix} \tag{37}$$

The results for the length of 12 points of the moving average filter are:

$$\begin{bmatrix} 0.9999 & -0.0235 & -0.2615 & -0.1486 \\ -0.0235 & 0.9999 & 0.1758 & 0.2552 \\ -0.2615 & 0.1758 & 1.0000 & 0.2785 \\ -0.1486 & 0.2552 & 0.2785 & 0.9999 \end{bmatrix} \tag{38}$$

The results for the length of 24 points of the moving average filter are:

$$\begin{bmatrix} 1.0000 & -0.2436 & -0.4057 & -0.2353 \\ -0.2436 & 1.0000 & 0.2725 & 0.4084 \\ -0.4057 & 0.2725 & 1.0000 & 0.5211 \\ -0.2353 & 0.4084 & 0.5211 & 1.0000 \end{bmatrix} \tag{39}$$

The closer the correlation is to 1, the higher the similarity, and the closer the correlation is to 0, the lower the similarity. As various results for the correlation matrix are obtained by using different moving average filter lengths, we implemented an algorithm for identifying the best solution (see below: Algorithm 1 for optimum moving average filter length identification).

Algorithm 1 for optimum moving average filter length identification.

```

for kk from 3 to 52
    data1_filtered=movemean(data1,kk)
    data2_filtered=movemean(data2,kk)
    A = [data1(1:length(min(data1, data2)))'
        data2(1:length(min(data1, data2)))'];
    % alternatively zero padding can be used that all vectors have the same length.
    d = (A*A')/(N - 1);
    e = d/max(max(d));
    zb(k) = sum(sum(dist(e-ones(size(e)))));
end;
find(zb == min(zb))
    
```

where N = length(min(data1, data2)).

Different results were obtained for different seasons (years). An optimum filter length was used for corresponding data in further analysis.

Data are presented by years (in the first part of the research) and by seasons (in the second part of the research). Half of the data were used to generate the equation and the other half to test the prediction hypothesis.

All the calculations were performed in the Matlab application Curve fitting tool [26]. The variables of interest were input as y-data.

4. Results

The data obtained consist of 829-time samples that are not uniformly sampled. They are obtained from a single ship (5 October 2008–28 August 2012), and refer to fuel consumption in 24 h. We used the first 434 samples (two years) to obtain the fit function, while the remaining 395 samples (two years) were used for the identification of the fitting function best suited to predict future fuel consumption. Similar analysis was performed by seasons. The analysis was divided into two parts: analysis by years known as the vertical analysis and horizontal analysis, where one year was divided into four seasons, namely, summer, autumn, winter, and spring.

4.1. Analysis by Year

The correlation matrix for the first two years is:

$$\begin{bmatrix} 1.0000 & 0.9661 \\ 0.9661 & 0.9361 \end{bmatrix} \tag{40}$$

From (40), it can be seen that the correlation between the first, and the second year was 96.61%. For four years, the smoothed data had the following correlation:

$$\begin{bmatrix} 1.0000 & 0.9600 & 0.9595 & 0.8379 \\ 0.9600 & 0.9237 & 0.9231 & 0.8064 \\ 0.9595 & 0.9231 & 0.9235 & 0.8064 \\ 0.8379 & 0.8064 & 0.8064 & 0.7048 \end{bmatrix} \tag{41}$$

As the lowest correlations were between the second and the fourth year, the third and the fourth, and within the fourth year itself, there was obviously a fundamental occurrence in the 4th year, as can be seen from (41). The results obtained suggest that denoised data are suitable for further research.

Table A1 indicates the coefficients obtained. As in the case of the fitting function called Rational 1/1 (16), the fit computation did not converge, Matlab stopped fitting, because the number of iterations or function evaluations exceeded the specified maximum. The results for the other fitting functions are presented in Table A2. Table A2 shows the results for years 1 and 2, where the fit was examined by SSE, R-square, RMSE, and the adjusted R-square. The table is sorted by R-square (the best on the top). Negative results denoted with “*” imply the invalidity of the model. It can be seen that the best fit (other than the smoothing spline curves) was obtained by Fourier 8. Table A3 shows the results of the prediction for various fitting functions. Table A3 is sorted by R-square in the prediction interval (the best on the top). The measure of quality is the R-square coefficient. The best result after smoothing splines was obtained by Gaussian 7.

Figures 2 and 3 show examples of the results. The dots are real samples, and the full lines are the estimated curve. Since the data are obtained by non-uniform daily sampling, the *x*-axes are marked as “indexed time” in all figures in the Results section. Hence, the numbers on the *x*-axes are not time units, but the number of daily samples, which have no units. Figures 2a and 3a present the first two years, denoised (y_{12m}), and the fitted data ($y_{years12_fit}$). Figures 2b and 3b show the same for years 3 and 4, and Figures 2c and 3c the same for all data (four years).

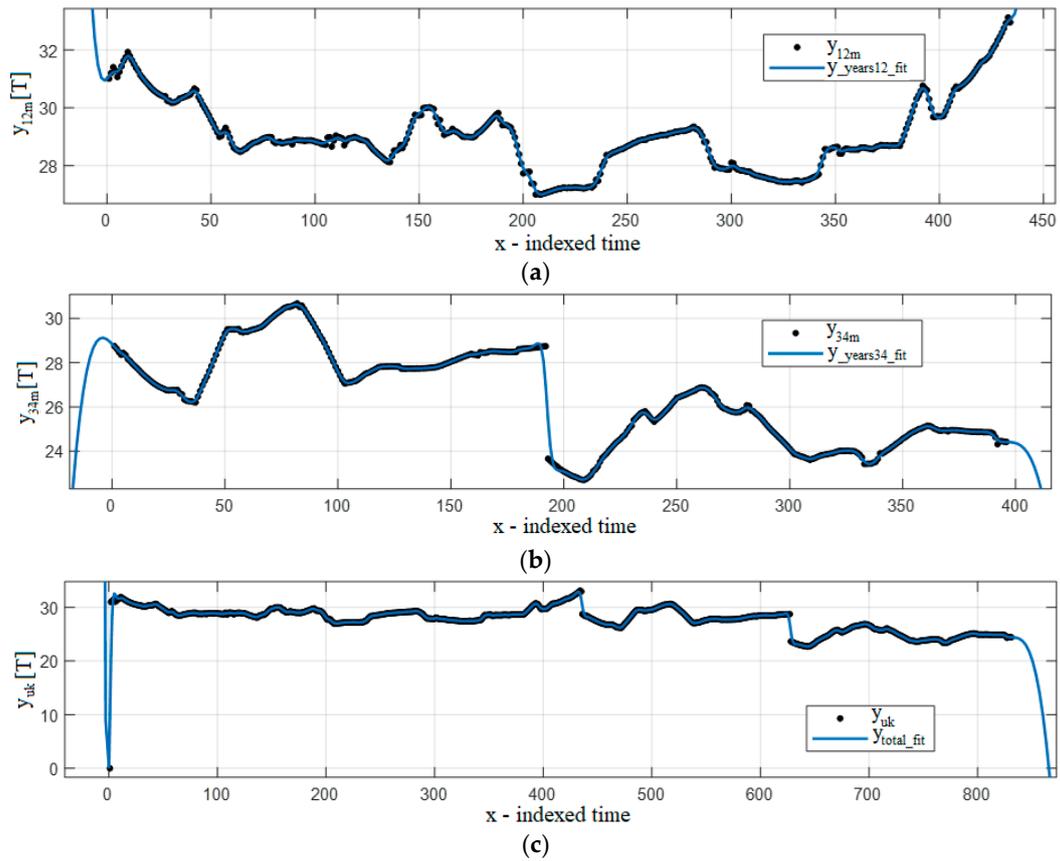


Figure 2. Smoothing spline for $p = 0.309432$: (a) years 1–2, (b) years 3–4, (c) years 1–4.

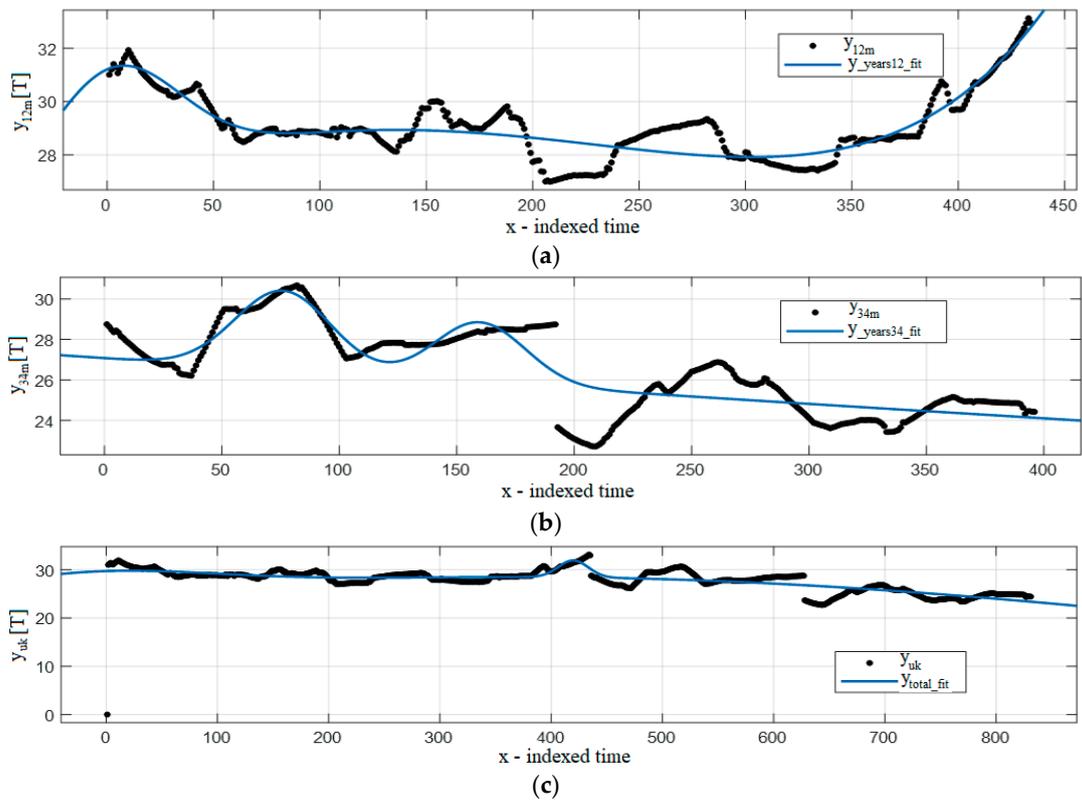


Figure 3. Gaussian 3 fitting function: (a) years 1–2, (b) years 3–4, (c) years 1–4.

4.2. Analysis by Seasons

Optimal filter length for the available data set in the summer is 21. By using the result from the algorithm, the correlation matrixes for summers 2 and 4 are:

$$\begin{bmatrix} 1.0000 & 0.8915 \\ 0.8915 & 0.7984 \end{bmatrix} \quad (42)$$

$$\begin{bmatrix} 1.0000 & 0.9007 & 0.9859 & 0.8006 \\ 0.9007 & 0.8132 & 0.8931 & 0.7200 \\ 0.9859 & 0.8931 & 0.9858 & 0.7863 \\ 0.8006 & 0.7200 & 0.7863 & 0.6419 \end{bmatrix} \quad (43)$$

The obtained coefficients for the considered fitting functions are given in Table A4 (see Appendix A). As the fit computation for the Gaussian 7 function did not converge, the fitting was discontinued since the number of iterations or function evaluations exceeded the specified maximum.

The obtained measures of similarity with the fitting functions are given in Table A5, sorted by the R-square (the best on the top row). In spite of the calculation problems, Gaussian 7 yielded the best fit under the R-square criterion. Table A6 shows R-square results for the domain used to extrapolate the fitting function parameters, predicted curve, and total range, sorted by the predicted R-square domain. The Gaussian 7 fitting function could be observed to be the best fit in the prediction interval. Figures 4 and 5 show examples of the results for the summer.

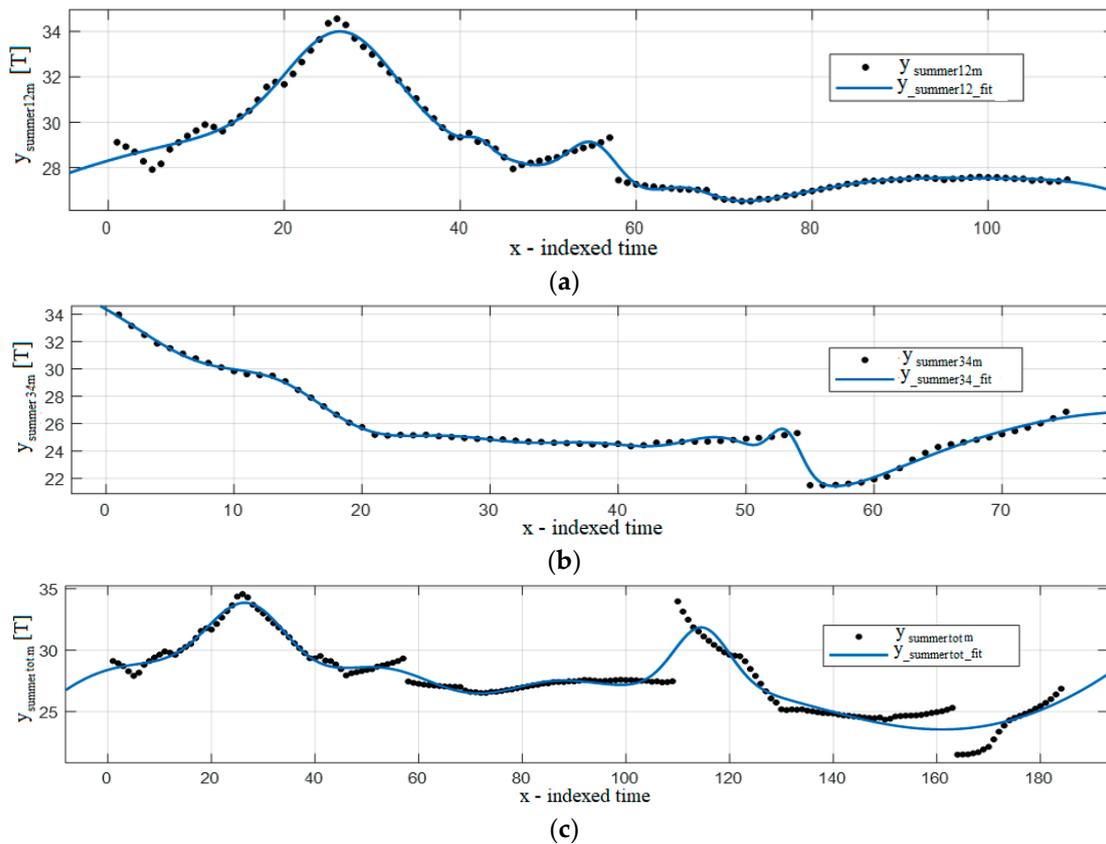


Figure 4. Gaussian 7 fitting function: (a) summers 1–2, (b) summers 3–4, (c) summers 1–4.

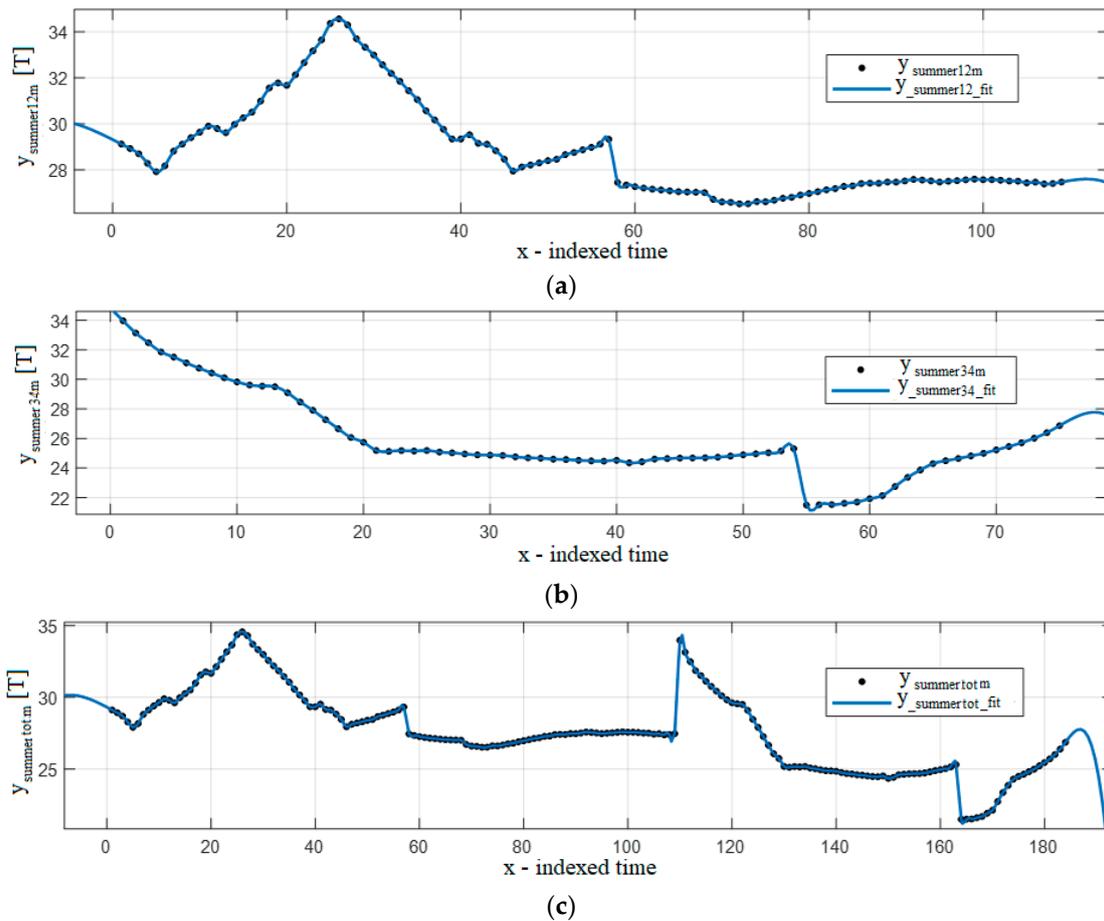


Figure 5. Smoothing spline for $p = 0.9987$: (a) summers 1–2, (b) summers 3–4, (c) summers 1–4.

After the summers, the analysis by season turned to autumns. The algorithm for the optimal length of the moving average filter gave the length of 42 in case of autumn. The correlation matrix for the first two autumns is:

$$\begin{bmatrix} 0.9734 & 0.9828 \\ 0.9828 & 1.0000 \end{bmatrix} \tag{44}$$

In case of all four autumns, the correlation matrix is:

$$\begin{bmatrix} 0.9734 & 0.9828 & 0.9432 & 0.8338 \\ 0.9828 & 1.0000 & 0.9557 & 0.8436 \\ 0.9432 & 0.9557 & 0.9155 & 0.8088 \\ 0.8338 & 0.8436 & 0.8088 & 0.7147 \end{bmatrix} \tag{45}$$

The obtained coefficients are given in Table A7 (see Appendix A). There was a problem with the Gaussian 7 function, where the fit computation did not converge and Matlab stopped fitting because the number of iterations or function evaluations exceeded the specified maximum.

Table A8 shows the results for level of fit. Data were sorted by the R-square parameter. The best results (after the smoothing splines) were obtained by the Gaussian 8 and Fourier 7. Table A9 shows R-square for autumns 1 and 2, autumns 3 and 4, and autumns 1–4. Data are sorted by prediction interval, with the Gaussian 7 fitting function giving the best prediction (after the smoothing spline functions).

Figures 6 and 7 show examples of the results, with real data illustrated with dots, and fitted curves with full lines. In this case, Gaussian 7 is chosen for Figure 6, and Rational 3/2 function for Figure 7. Figure 6b,c and Figure 7b,c show discontinuities in the middle of the graph.

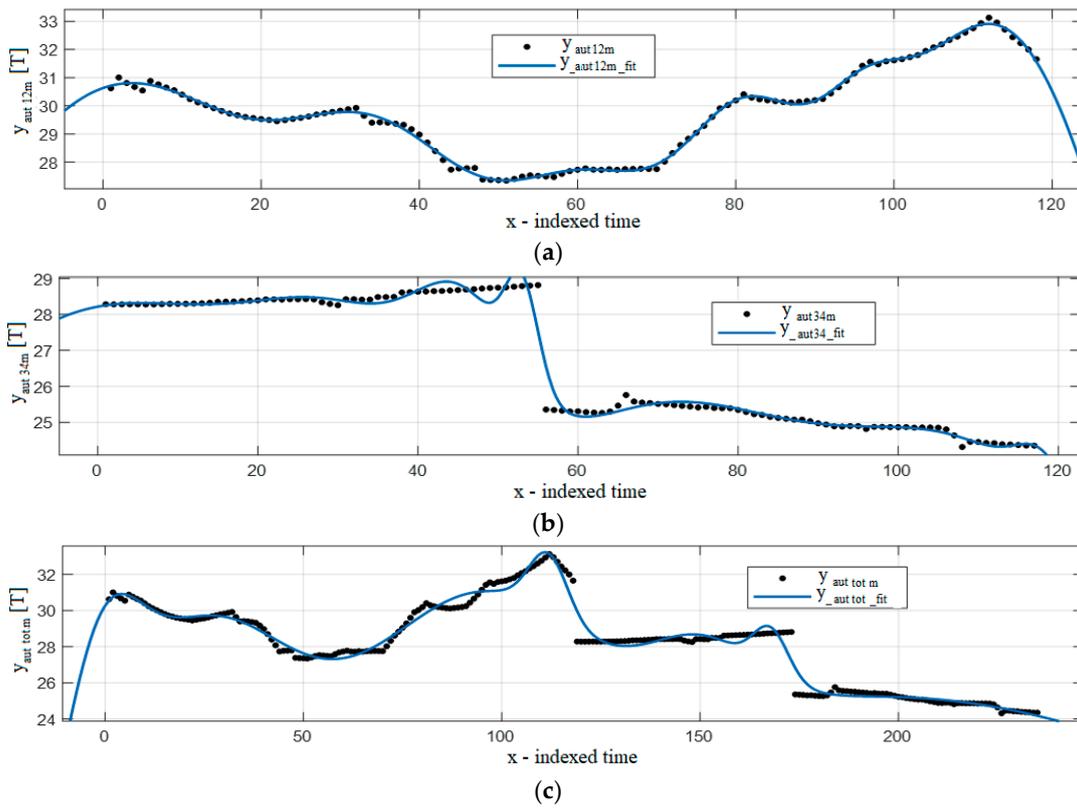


Figure 6. Gaussian 7 fitting function results: (a) autumns 1–2, (b) autumns 3–4, (c) autumns 1–4.

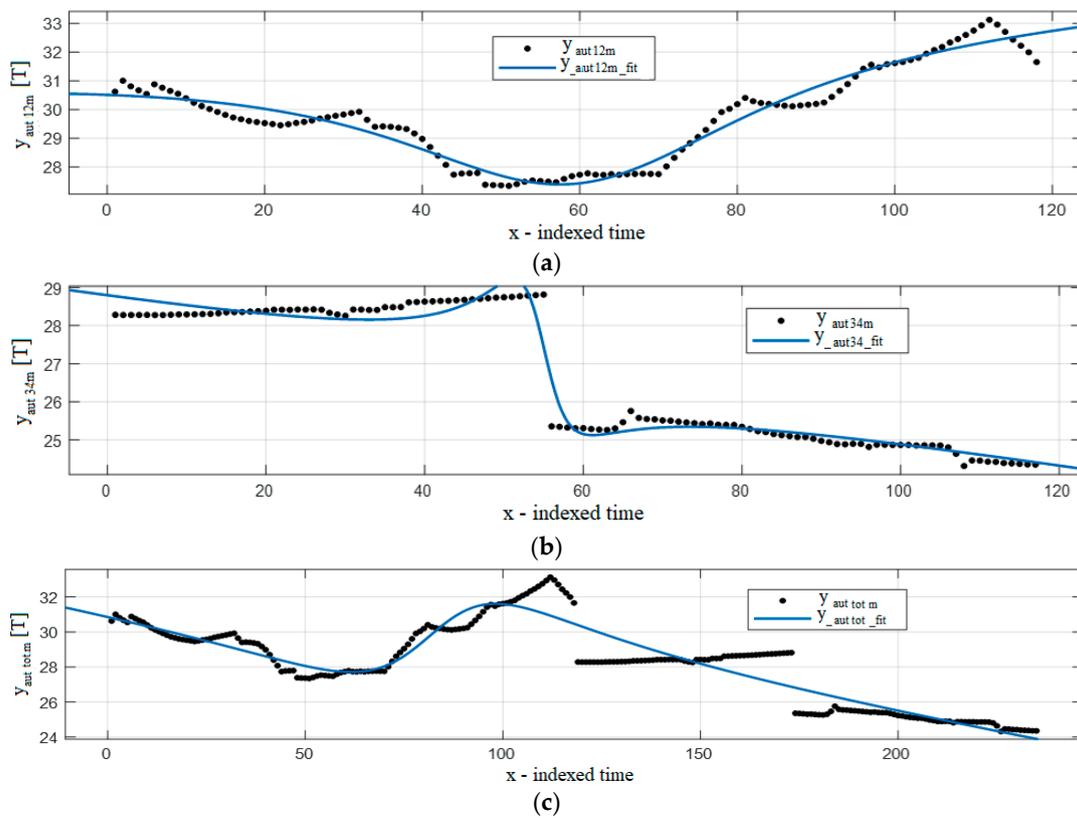


Figure 7. Rational3/2 fitting function results: (a) autumns 1–2, (b) autumns 3–4, (c) autumns 1–4.

Correlations for two winters are:

$$\begin{bmatrix} 1.0000 & 0.9513 \\ 0.9513 & 0.9049 \end{bmatrix} \tag{46}$$

and for all winters:

$$\begin{bmatrix} 1.0000 & 0.9513 & 0.9647 & 0.8458 \\ 0.9513 & 0.9049 & 0.9176 & 0.8045 \\ 0.9647 & 0.9176 & 0.9306 & 0.8160 \\ 0.8458 & 0.8045 & 0.8160 & 0.7157 \end{bmatrix} \tag{47}$$

The obtained coefficients (for fitting functions) for winters 1 and 2 are given in Table A10 (see Appendix A). For the Gaussian 7 fitting function, the fit computation did not converge and the fitting was halted. The results in the table are the last obtained in this case.

Table A11, sorted by the R-square parameter, shows quality measures for winters 1 and 2. The best fit was obtained with Gaussian 7 (if the smoothing splines were excluded). Table A12, sorted by the predicted domain, shows R-squares for winters 1–2, 3–4, and 1–4. Gaussian 7 is the best for the prediction (when smoothing splines are excluded).

Figure 8 shows an example of the results in the case of the Gaussian 7 fitting function.

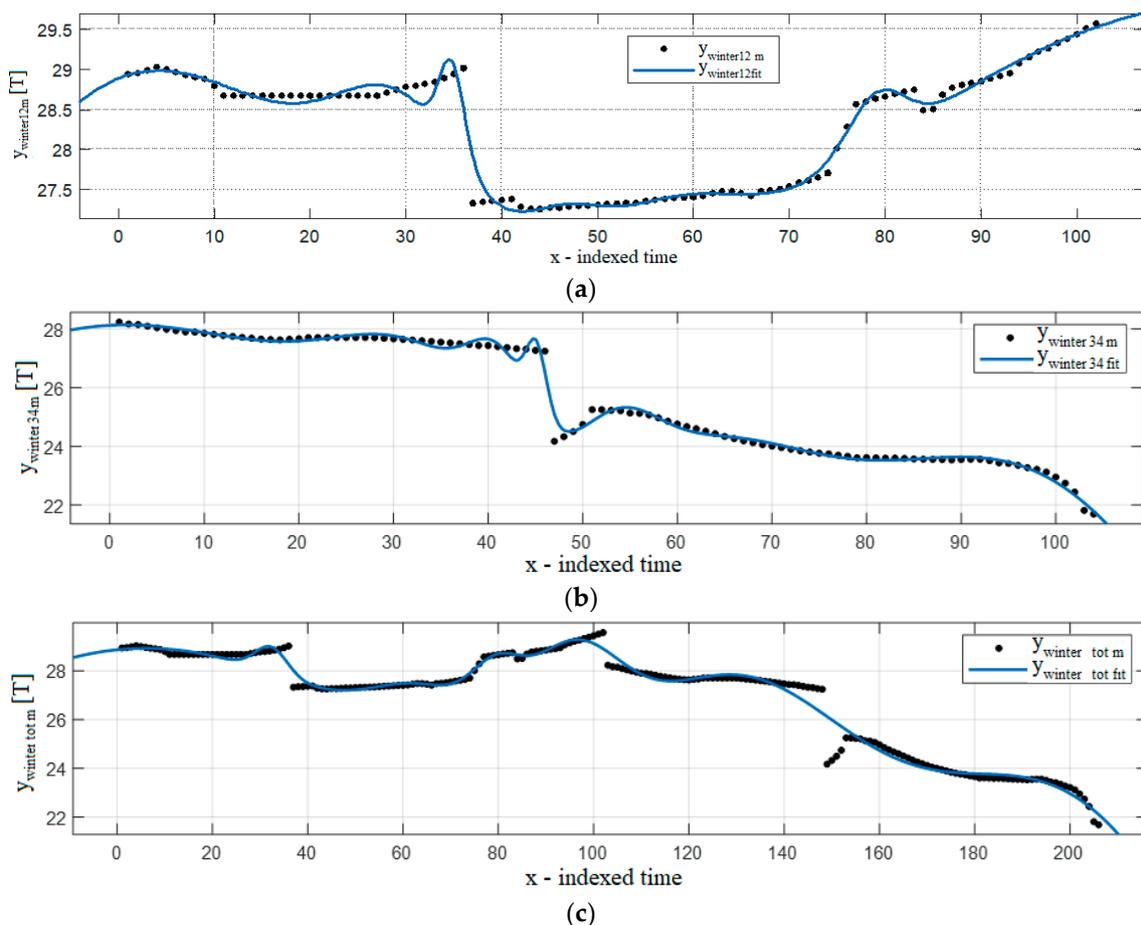


Figure 8. Gaussian 7 fitting function results: (a) winters 1–2, (b) winters 3–4, (c) winters 1–4.

In the case of the two springs, the optimal length for the moving average filter was 47. Correlations for the two springs are:

$$\begin{bmatrix} 0.9780 & 0.9885 \\ 0.9885 & 1.0000 \end{bmatrix} \tag{48}$$

and for four springs:

$$\begin{bmatrix} 0.8157 & 0.8332 & 0.9025 & 0.6972 \\ 0.8332 & 0.8515 & 0.9226 & 0.7113 \\ 0.9025 & 0.9226 & 1.0000 & 0.7700 \\ 0.6972 & 0.7113 & 0.7700 & 0.5974 \end{bmatrix} \quad (49)$$

The results for two springs are given. The obtained coefficients for the fitting functions are presented in Table A13 (see Appendix A).

The results for quality measures in case of springs 1 and 2 are shown in Table A14, which is sorted by the best R-square. Better results are for rougher smoothing spline (p closer to 1), which is not favorable. On the contrary, smoother splines are welcome and the achieved results are not as good. Apart from the smoothing splines, the best result was obtained with Gaussian 7. Table A15, sorted by the best fits in the prediction interval, shows the results for springs 1–2, 3–4, 1–4. Fourier 8 was shown as the best for the prediction purposes in the case of springs.

Figures 9 and 10 show examples of the results for springs. Fourier 8 clearly exhibits discontinuity around the middle of the dataset.

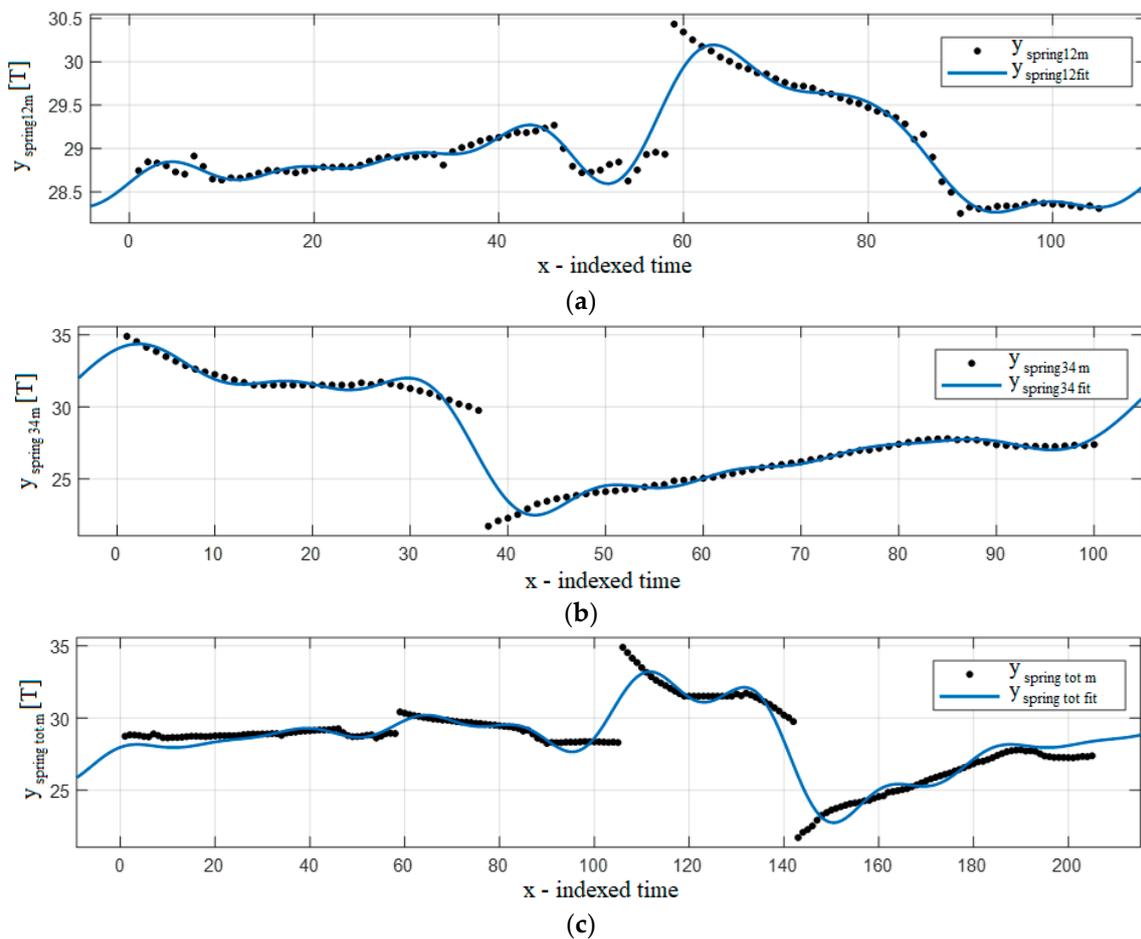


Figure 9. Fourier 8 fitting function results: (a) springs 1–2, (b) springs 3–4, (c) springs 1–4.

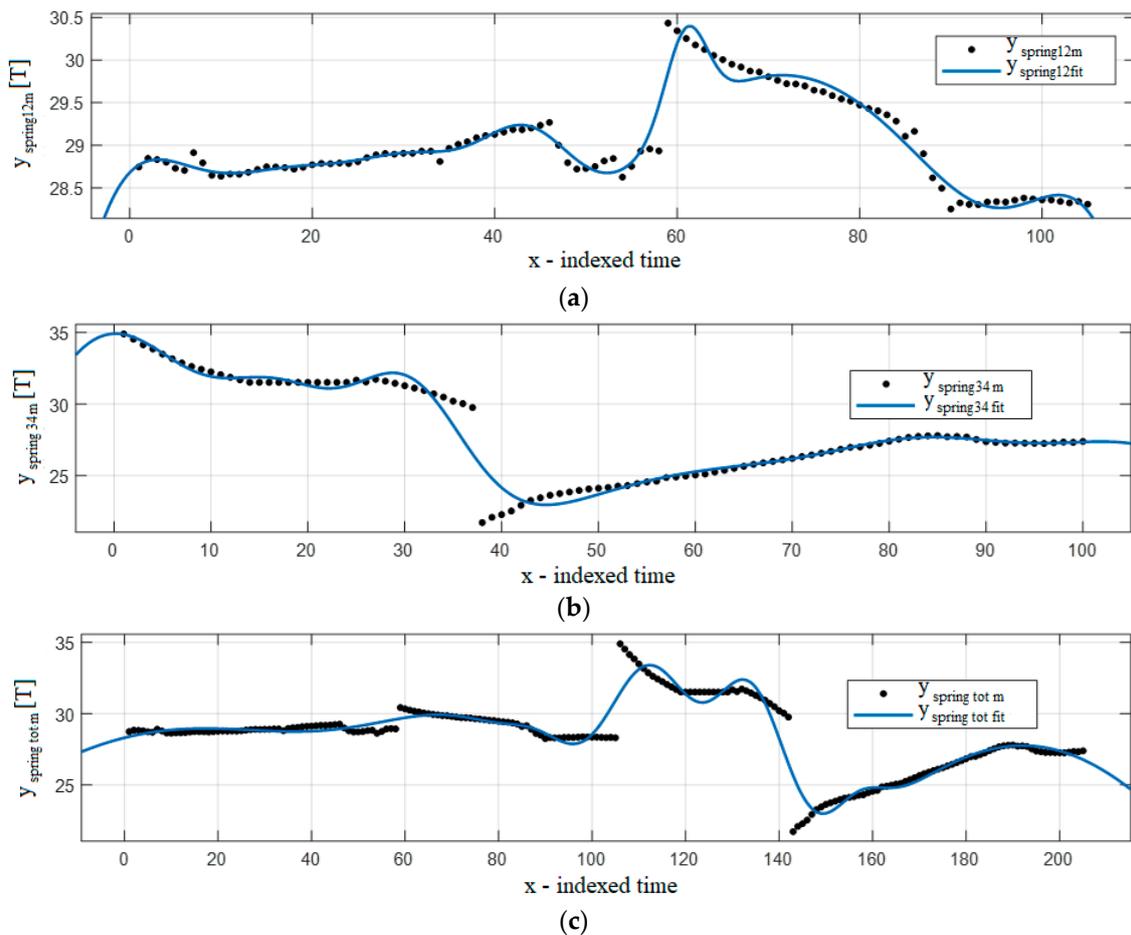


Figure 10. Gaussian 7 fitting function results: (a) springs 1–2, (b) springs 3–4, (c) springs 1–4.

5. Discussion and Conclusions

The paper presents a black-box approach to ship fuel consumption based on signal (daily fuel consumption) analysis. The real data were analyzed. The fitting function should enable the easy build in of the possible tools/software in the future, which is simple and requires low processor power. This fitting function should reflect the coupling between data and sources. Due to nonlinearities, various data were transformed to the same coordinates. This is important, because monthly, seasonally, or yearly data do not have the same length. The explored approach should include aspects of de-trending, coupling, nonlinearities, unreliable data, and different lengths of data within the same time period. Finally, the fitting function would enable the prediction of fuel consumption and differentiate it from the actual consumption. If an anomaly is detected, the company could investigate whether the cause is the weather, fouling, or even fraud. During the research, it was assumed that fuel consumption would grow linearly over time. Such dependence would make possible to use some linear regression methods (e.g., [20,29]). However, the results did not show that consumption had actually decreased, which could be associated with the following factors:

- non-uniform time sampling (leading to wrong curve angle between the interpolating points), and
- average (which depends on the route the ship was sailing at the time of data acquisition, and the sailing hours on a specific day).

It was obvious from the results of the R-square and the adjusted R-square measures that the data were not well conditioned for the correct analysis. Hence, different approaches were used. Firstly, the data were preprocessed with filters, which resulted in better R-square and adjusted R-square metrics. The next step was to use a bank of fitting functions to find the best match. Higher-order

functions were observed to result in a better fit. The best functions to choose for the model were the smoothing splines, but in the so-called rougher (not smoother) versions (parameter p closer to 1), as rougher splines fit better sampling points, although they were not suitable for real-world applications. Instead, the Gaussian, Fourier or rational functions were chosen for the prediction model. In this case, the best fit of the predicted data (samples for 3rd and 4th year/season) was obtained by Gaussian, Fourier, and rational functions that generally tend to be represented by the first-order polynomial. Higher order functions tend to oscillate around the sampled data more tightly. That reduces the error, and it could lead to the conclusion that such functions (i.e., Gaussian 7) could be used in any ship by simply adjusting the function parameters, which is the advantage in our signal-based approach. The result was obtained easily and fast, which is an advantage over methods that attempt to estimate real fuel consumption by examining engine characteristics and parameters.

In the analysis, we explored comparisons between years, and comparisons between seasons. The worst correlations were obtained between the second and the fourth year, the third and the fourth, and within the fourth year itself, suggesting a relevant event of some sort in the fourth year. Likewise, the correlation between summers, autumns, and winters was above 80%. On the other hand, springs were less correlated, presumably due to the state of the ship's hull after wintertime, and the great oscillations in weather conditions that typically occur in springtime.

Finally, there were two contributions of this paper. The main contribution, as far as the authors are aware, was that this was the first time that seasons (so called horizontal analysis) were considered for prediction purposes. A minor contribution of the paper was the algorithm for identifying the optimum moving average filter length.

The prediction formula takes into account environmental and biological influences, as well as cargo mass, and all other possible fuel consumption factors. We strongly believe that it could be used to detect potential frauds (fuel theft), which may be of interest to various authorities. Furthermore, the proposed prediction model is simple and fast to use, and can be used to check deviations in fuel consumption by comparing the predicted and deviated consumption. A deviation could be caused by fraud or by environmental factors such as fouling.

There are additional differences between this paper and references. For example, the results in [19] are obtained by excluding suspicious data in the preprocessing stage. We used all available data to obtain our results. Most of the references used a deep learning approach with ANNs [19–23]. On the other hand, it was not necessary to use ANNs always. It was popular, but not necessary. Hence, further research could include the identification of factors that cause unplanned fuel consumption and modeling fuel consumption by artificial neural networks (ANN) and hybrid models of velocity and fuel consumption.

Author Contributions: Conceptualization, I.V.; methodology, I.V., J.Š. and I.K.; software, I.V. and M.P.; validation, J.Š. and I.V.; formal analysis, I.K., and I.V.; investigation, I.V. and M.P.; writing—original draft preparation, I.V.; Writing—Review & Editing, I.V. and J.Š.; visualization, I.V. and M.P.; supervision, I.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Appendix A

The obtained coefficients of the fitting functions are given in this Appendix. To simplify the tables, several additional equations are used. The expression for Fourier 8 is:

$$\begin{aligned}
 f(x) = & a_0 + a_1 \cdot \cos(x \cdot \omega) + b_1 \cdot \sin(x \cdot \omega) + a_2 \cdot \cos(2x \cdot \omega) + b_2 \cdot \sin(2x \cdot \omega) + \\
 & + a_3 \cdot \cos(3x \cdot \omega) + b_3 \cdot \sin(3x \cdot \omega) + a_4 \cdot \cos(4x \cdot \omega) + b_4 \cdot \sin(4x \cdot \omega) + \\
 & + a_5 \cdot \cos(5x \cdot \omega) + b_5 \cdot \sin(5x \cdot \omega) + a_6 \cdot \cos(6x \cdot \omega) + b_6 \cdot \sin(6x \cdot \omega) + \\
 & + a_7 \cdot \cos(7x \cdot \omega) + b_7 \cdot \sin(7x \cdot \omega) + a_8 \cdot \cos(8x \cdot \omega) + b_8 \cdot \sin(8x \cdot \omega)
 \end{aligned}
 \tag{A1}$$

The expression for Gaussian 3 is:

$$f(x) = a_1 e^{-((x-b_1)/c_1)^2} + a_2 e^{-((x-b_2)/c_2)^2} + a_3 e^{-((x-b_3)/c_3)^2} \tag{A2}$$

The expression for Gaussian 7 is:

$$f(x) = a_1 e^{-((x-b_1)/c_1)^2} + a_2 e^{-((x-b_2)/c_2)^2} + a_3 e^{-((x-b_3)/c_3)^2} + a_4 e^{-((x-b_4)/c_4)^2} + a_5 e^{-((x-b_5)/c_5)^2} + a_6 e^{-((x-b_6)/c_6)^2} + a_7 e^{-((x-b_7)/c_7)^2} \tag{A3}$$

If we take Table A1, the first row, for the fitting function described by (1), it was found that the coefficients (with 95% confidence bounds) are: $a = 0.01021$ ($-0.1528, 0.1732$), $b = 2.519 \cdot 10^{-6}$ (confidence bounds: $3.792 \cdot 10^{-7}, 4.658 \cdot 10^{-6}$), $c = 28.91$ ($28.74, 29.08$) and the test function is therefore:

$$f(t) = 0.01021 * (\sin(t - \pi)) + 2.519 \cdot 10^{-6} \cdot (t - 10)^2 + 28.91 \tag{A4}$$

for the first two years. Variable t denotes samples, not physical time.

For the fitting exponential function described by (2), the coefficients are: $a = 29.25$ (bounds: $29.02, 29.48$), $b = -3.011 \cdot 10^{-5}$ (95% confidence bounds: $-6.18 \cdot 10^{-5}, 1.573 \cdot 10^{-6}$):

$$f(t) = 29.25 e^{-3.011 \cdot 10^{-5} t} \tag{A5}$$

For the fitting function in (3), the coefficients are: $a = 30.39$ ($30.22, 30.57$), $b = -0.00035$ ($-0.0003961, -0.0003039$), $c = 0.002414$ ($-0.0005264, 0.005354$), $d = 0.01832$ ($0.01554, 0.02109$):

$$f(t) = 30.39 e^{-0.00035 t} + 0.002414 e^{-0.01832 t} \tag{A6}$$

For Fourier 1 (12), the fitting function coefficients are: $a_0 = 6.872 \cdot 10^7$ ($-1.162 \cdot 10^{15}, 1.162 \cdot 10^{15}$), $a_1 = -6.872 \cdot 10^7$ ($-1.162 \cdot 10^{15}, 1.162 \cdot 10^{15}$), $b_1 = 2.12 \cdot 10^4$ ($-1.792 \cdot 10^{11}, 1.792 \cdot 10^{11}$), $\omega = -1.377 \cdot 10^{-6}$ ($-11.64, 11.64$). The final fitting function is:

$$f(t) = 6.872 \cdot 10^7 - 6.872 \cdot 10^7 \cos(t \cdot \omega) + 2.12 \cdot 10^4 \sin(t \cdot \omega) \tag{A7}$$

The detailed results on the coefficients and fitting are presented in the appendix. Tables A2, A3, A5, A6, A8, A9, A11, A12, A14 and A15 present the results of the fitting, and Tables A1, A4, A7, A10 and A13 results for the obtained coefficients.

Table A1. Obtained coefficients for the considered functions (year 1–2 interval), values in brackets are confidence bounds for 95%.

Linear fitting (1)	$a = 0.01021 (-0.1528, 0.1732)$, $b = 2.519 \cdot 10^{-6} (3.792 \cdot 10^{-7}, 4.658 \cdot 10^{-6})$, $c = 28.91 (28.74, 29.08)$
Exponential of 1st order (2)	$a = 29.25 (29.02, 29.48)$, $b = -3.011 \cdot 10^{-5} (-6.18 \cdot 10^{-5}, 1.573 \cdot 10^{-6})$
Exponential of 2nd order (3)	$a = 30.39 (30.22, 30.57)$, $b = -0.00035 (-0.0003961, -0.0003039)$, $c = 0.002414 (-0.0005264, 0.005354)$, $d = 0.01832 (0.01554, 0.02109)$
Fourier 1 (5)	$a_0 = 6.872 \cdot 10^7 (-1.162 \cdot 10^{15}, 1.162 \cdot 10^{15})$, $a_1 = -6.872 \cdot 10^7 (-1.162 \cdot 10^{15}, 1.162 \cdot 10^{15})$, $b_1 = 2.12 \cdot 10^4 (-1.792 \cdot 10^{11}, 1.792 \cdot 10^{11})$, $\omega = -1.377 \cdot 10^{-6} (-11.64, 11.64)$
Fourier 2 (6)	$a_0 = 4.758 \cdot 10^9 (-2.346 \cdot 10^{14}, 2.346 \cdot 10^{14})$, $a_1 = -6.344e+09 (-3.128 \cdot 10^{14}, 3.127 \cdot 10^{14})$, $b_1 = -5.466 \cdot 10^7 (-2.021 \cdot 10^{12}, 2.021 \cdot 10^{12})$, $a_2 = 1.586 \cdot 10^9 (-7.818 \cdot 10^{13}, 7.818 \cdot 10^{13})$, $b_2 = 2.733 \cdot 10^7 (-1.011 \cdot 10^{12}, 1.011 \cdot 10^{12})$, $\omega = 4.293 \cdot 10^{-5} (-0.529, 0.5291)$
Fourier 8 (7)	$a_0 = -8.62 \cdot 10^6 (-5.631 \cdot 10^7, 3.907 \cdot 10^7)$, $a_1 = 6.305 \cdot 10^6 (-3.368 \cdot 10^7, 4.629 \cdot 10^7)$, $b_1 = 1.433 \cdot 10^7 (-6.25 \cdot 10^7, 9.116 \cdot 10^7)$, $a_2 = 7.899 \cdot 10^6 (-2.917 \cdot 10^7, 4.497 \cdot 10^7)$, $b_2 = -8.618 \cdot 10^6 (-6.156 \cdot 10^7, 4.432 \cdot 10^7)$, $a_3 = -6.728 \cdot 10^6 (-4.564 \cdot 10^7, 3.218 \cdot 10^7)$, $b_3 = -2.288 \cdot 10^6 (-7.388 \cdot 10^6, 2.812 \cdot 10^6)$, $a_4 = 2.967 \cdot 10^5 (-6.255 \cdot 10^6, 6.848 \cdot 10^6)$, $b_4 = 3.439 \cdot 10^6 (-1.443 \cdot 10^7, 2.131 \cdot 10^7)$, $a_5 = 1.139 \cdot 10^6 (-3.655 \cdot 10^6, 5.932 \cdot 10^6)$, $b_5 = -6.212 \cdot 10^5 (-5.908 \cdot 10^6, 4.665 \cdot 10^6)$, $a_6 = -2.82 \cdot 10^5 (-2.176 \cdot 10^6, 1.612 \cdot 10^6)$, $b_6 = -2.183 \cdot 10^5 (-7.059 \cdot 10^5, 2.694 \cdot 10^5)$, $a_7 = -1.572 \cdot 10^4 (-9.789 \cdot 10^4, 6.645 \cdot 10^4)$, $b_7 = 6.237 \cdot 10^4 (-2.793 \cdot 10^5, 4.041 \cdot 10^5)$, $a_8 = 5689 (-1.772 \cdot 10^4, 2.909 \cdot 10^4)$, $b_8 = -940.6 (-2.177 \cdot 10^4, 1.988 \cdot 10^4)$, $\omega = 0.00524 (0.003601, 0.006878)$
Gaussian 1 (9)	$a_1 = 3.427 \cdot 10^{14} (-3.9 \cdot 10^{20}, 3.9 \cdot 10^{20})$, $b_1 = -2.005 \cdot 10^6 (-7.582 \cdot 10^{10}, 7.581 \cdot 10^{10})$, $c_1 = 3.655 \cdot 10^5 (-6.909 \cdot 10^9, 6.91 \cdot 10^9)$
Gaussian 2 (10)	$a_1 = 2.511 \cdot 10^{13} (-7.073 \cdot 10^{16}, 7.078 \cdot 10^{16})$, $b_1 = 5051 (-4.541 \cdot 10^5, 4.642 \cdot 10^5)$, $c_1 = 861.5 (-4.259 \cdot 10^4, 4.431 \cdot 10^4)$, $a_2 = 31.26 (23, 39.52)$, $b_2 = -239 (-1875, 1397)$, $c_2 = 1316 (-3279, 5910)$
Gaussian 3 (A2)	$a_1 = 4.832 \cdot 10^{13} (-2.725 \cdot 10^{17}, 2.726 \cdot 10^{17})$, $b_1 = 6316 (-1.167 \cdot 10^6, 1.18 \cdot 10^6)$, $c_1 = 1096 (-1.102 \cdot 10^5, 1.124 \cdot 10^5)$, $a_2 = 4.068 (0.9324, 7.204)$, $b_2 = 1.883 (-16.1, 19.86)$, $c_2 = 43.76 (23.13, 64.38)$, $a_3 = 28.29 (9.087, 47.5)$, $b_3 = 104 (-323, 531.1)$, $c_3 = 473.6 (-422.7, 1370)$
Gaussian 7 (A3)	$a_1 = 67.24 (-1483, 1617)$, $b_1 = 1033 (-1.778 \cdot 10^4, 1.985 \cdot 10^4)$, $c_1 = 705.8 (-1.017 \cdot 10^4, 1.159 \cdot 10^4)$, $a_2 = 23.55 (-64.26, 111.4)$, $b_2 = -9.979 (-35.28, 15.32)$, $c_2 = 138.8 (-90.6, 368.3)$, $a_3 = 10.09 (-10.02, 30.2)$, $b_3 = 171.9 (140.2, 203.6)$, $c_3 = 58.59 (9.888, 107.3)$, $a_4 = -7.617 (-2314, 2298)$, $b_4 = 284.8 (-70.82, 640.5)$, $c_4 = 43.15 (-808.4, 894.7)$, $a_5 = 2.474 (-0.5536, 5.502)$, $b_5 = 100.5 (93.52, 107.4)$, $c_5 = 29.94 (17.49, 42.39)$, $a_6 = 1.201 (-19.26, 21.66)$, $b_6 = 349.6 (252.2, 447)$, $c_6 = 30.59 (-61.28, 122.5)$, $a_7 = 14.42 (-2241, 2269)$, $b_7 = 277.2 (-838.6, 1393)$, $c_7 = 48.2 (-165.1, 261.5)$
Polynomial 1 (11)	$p_1 = -0.0008497 (-0.001771, 7.112 \cdot 10^{-5})$, $p_2 = 29.25 (29.01, 29.48)$
Polynomial 2 (12)	$p_1 = 6.516 \cdot 10^{-5} (5.972 \cdot 10^{-5}, 7.061 \cdot 10^{-5})$, $p_2 = -0.0292 (-0.03164, -0.02675)$, $p_3 = 31.31 (31.07, 31.54)$
Polynomial 3 (13)	$p_1 = 1.807 \cdot 10^{-7} (1.343 \cdot 10^{-7}, 2.272 \cdot 10^{-7})$, $p_2 = -5.277 \cdot 10^{-5} (-8.352 \cdot 10^{-5}, -2.202 \cdot 10^{-5})$, $p_3 = -0.008652 (-0.01441, -0.002893)$, $p_4 = 30.56 (30.27, 30.85)$
Power 1 (14)	$a = 31.41 (30.8, 32.01)$, $b = -0.01529 (-0.01903, -0.01156)$
Power 2 (15)	$a = 5.338 (4.078, 6.597)$, $b = -0.3466 (-0.5568, -0.1364)$, $c = 28.08 (27.17, 28.98)$
Rational 1/1 (16)	$p_1 = 26.56 (25.81, 27.31)$, $p_2 = -14.02 (-19.49, -8.555)$, $q_1 = -2.457 (-2.51, -2.404)$
Rational 2/1 (17)	$p_1 = -0.0007214 (-0.001646, 0.0002029)$, $p_2 = 29.21 (28.97, 29.44)$, $p_3 = -80.3 (-95.63, -64.96)$, $q_1 = -2.77 (-3.257, -2.283)$
Rational 3/1 (18)	$p_1 = 6.501 \cdot 10^{-5} (5.95 \cdot 10^{-5}, 7.051 \cdot 10^{-5})$, $p_2 = -0.02954 (-0.03205, -0.02703)$, $p_3 = 31.49 (31.24, 31.73)$, $p_4 = -208.4 (-255.8, -161.1)$, $q_1 = -6.665 (-8.17, -5.161)$
Rational 3/2 (19)	$p_1 = -0.0006458 (-0.001574, 0.0002824)$, $p_2 = 29.19 (28.94, 29.43)$, $p_3 = -279.9 (-305.2, -254.6)$, $p_4 = 645.2 (522.9, 767.4)$, $q_1 = -9.623 (-10.42, -8.825)$, $q_2 = 22.25 (18.36, 26.14)$
Rational 5/3 (20)	$p_1 = 6.325 \cdot 10^{-5} (5.593 \cdot 10^{-5}, 7.057 \cdot 10^{-5})$, $p_2 = -0.02889 (-0.03347, -0.02431)$, $p_3 = 31.51 (30.55, 32.47)$, $p_4 = -371.8 (-993.9, 250.3)$, $p_5 = 1140 (-5307, 7586)$, $p_6 = 184.8 (-1.71 \cdot 10^4, 1.747 \cdot 10^4)$, $q_1 = -11.96 (-31.33, 7.415)$, $q_2 = 37.02 (-166.3, 240.4)$, $q_3 = 5.183 (-543.4, 553.8)$

Table A2. Quality measures for the considered functions (year 1–2 interval).

Fitness Function	SSE	R-Square	RMSE	Adjusted R-Square
Smoothing spline, $p = 0.99876718875$	0.0003582	1	0.006983	1
$p = 0.9$	0.1907	0.9997	0.03309	0.9993
$p = 0.309432$	0.9227	0.9986	0.05477	0.998
Fourier 8	39.79	0.9389	0.3093	0.9364
Gaussian 7	44.33	0.9319	0.3276	0.9286
Gaussian 3	152.1	0.7663	0.5983	0.7619
Fourier 2	159.1	0.7556	0.6097	0.7527
Exponential of 2nd order	205.2	0.6848	0.6907	0.6826
Gaussian 2	217.1	0.6665	0.7121	0.6626
Polynomial 3	248.9	0.6176	0.7608	0.6149
Rational 5/3	281.3	0.5679	0.8135	0.5598
Rational 3/1	282.4	0.5661	0.8113	0.5621
Fourier 1	282.7	0.5656	0.8109	0.5626
Polynomial 2	282.7	0.5656	0.8099	0.5636
Power 2	549.4	0.1559	1.129	0.152
Power 1	569.2	0.1255	1.148	0.1235
Rational 3/2	634	0.02602	1.217	0.01464
Rational 2/1	638.2	0.01947	1.218	0.01263
Linear fitting	642.9	0.01231	1.221	0.007731
Exponential of 1st order	645.8	0.007776	1.223	0.005479
Gaussian 1	645.8	0.007772	1.224	0.003168
Polynomial 1	646	0.007557	1.223	0.00526
Rational 1/1	2.66×10^4	-39.87^*	7.856	-40.06^*

* model results are not well conditioned by Matlab because the model is not good for this data.

Table A3. R-square for the domain used to fit, prediction interval, and total range.

Fitness Function	Fit-Domain (Samples 1–434)	Prediction Interval (Samples 435–829)	Total Range (Samples 1–829)	Comment
Smoothing spline, $p = 0.99876718875$	1	1	1	The best fit
$p = 0.9$	0.9997	0.9987	0.9919	Near the best fit
$p = 0.309432$	0.9986	0.9955	0.9501	Near the best fit
Gaussian 7	0.9319	0.9599	0.6899	The best fit if smoothing splines are not taken into account
Fourier 8	0.9389	0.934	0.7713	
Gaussian 3	0.7663	0.7583	0.5583	
Rational 5/3	0.56	0.6949	0.5008	
Rational 3/2	0.5656	0.6666	-3.154^*	Not good for calculations
Fourier 2	0.7556	0.6513	0.6933	
Gaussian 2	0.6665	0.6462	0.504	
Polynomial 3	0.6176	0.6247	0.4774	
Fourier 1	0.5656	0.6236	0.5708	
Exponential 2	0.6848	0.5763	0.5848	
Power 2	0.1559	0.5335	0.4715	
Gaussian 1	0.007772	0.5297	0.5667	
Rational 3/1	0.5659	0.529	0.4613	
Rational 2/1	0.01768	0.5283	0.5817	Too low
Polynomial 2	0.5656	0.5282	0.4597	
Polynomial 1	0.007557	0.5272	0.3788	Too low for column1
Exponential 1	0.007776	0.5256	0.486	Too low for column1
Linear fitting	0.01231	0.5023	0.5635	Too low for column1
Power 1	0.1255	0.3073	0.1848	
Rational 1/1	0.008808	0.003205	0.1977	Too low

* Matlab warning: A negative R-square is possible if the model does not contain a constant term, and the fit is poor (worse than just fitting the mean). Try changing the model or using a different start point.

Table A4. Obtained coefficients for the considered functions (summers 1–2 interval), values in brackets are confidence bounds for 95%.

Linear fitting (1)	$a = 0.01683$ (−0.4147, 0.4484), $b = -0.000439$ (−0.0005423, −0.0003357), $c = 30.13$ (29.7, 30.57)
Exponential of 1st order (2)	$a = 31.25$ (30.63, 31.87), $b = -0.001501$ (−0.001827, −0.001175)
Exponential of 2nd order (3)	$a = 31.57$ (30.81, 32.33), $b = -0.00183$ (−0.002467, −0.001192), $c = 7.661 \cdot 10^{-5}$ (−0.001465, 0.001618), $d = 0.09373$ (−0.09131, 0.2788)
Fourier 1 (5)	$a_0 = 28.8$ (28.6, 29.01), $a_1 = 0.09534$ (−0.4701, 0.6608), $b_1 = 2.519$ (2.217, 2.821), $\omega = 0.05844$ (0.05501, 0.06188)
Fourier 2 (6)	$a_0 = 28.82$ (28.67, 28.96), $a_1 = 0.1301$ (−0.1963, 0.4566), $b_1 = 2.519$ (2.301, 2.736), $a_2 = -1.042$ (−1.245, −0.8392), $b_2 = 0.02157$ (−0.2299, 0.273), $\omega = 0.05837$ (0.05573, 0.061)
Fourier 8 (7)	$a_0 = -1.122 \cdot 10^{12}$ (−7.808·10 ¹³ , 7.584·10 ¹³), $a_1 = 1.699 \cdot 10^{12}$ (−1.173·10 ¹⁴ , 1.207·10 ¹⁴), $b_1 = 1.061 \cdot 10^{12}$ (−6.774·10 ¹³ , 6.986·10 ¹³), $a_2 = -6.233 \cdot 10^{11}$ (−4.929·10 ¹³ , 4.805·10 ¹³), $b_2 = -1.278 \cdot 10^{12}$ (−8.583·10 ¹³ , 8.327·10 ¹³), $a_3 = -8.302 \cdot 10^{10}$ (−1.999·10 ¹¹ , 3.384·10 ¹⁰), $b_3 = 7.891 \cdot 10^{11}$ (−5.367·10 ¹³ , 5.525·10 ¹³), $a_4 = 2.104 \cdot 10^{11}$ (−1.157·10 ¹³ , 1.199·10 ¹³), $b_4 = -2.69 \cdot 10^{11}$ (−2.053·10 ¹³ , 1.999·10 ¹³), $a_5 = -1.03 \cdot 10^{11}$ (−6.63·10 ¹² , 6.424·10 ¹²), $b_5 = 3.73 \cdot 10^{10}$ (−3.696·10 ¹² , 3.771·10 ¹²), $a_6 = 2.414 \cdot 10^{10}$ (−1.672·10 ¹² , 1.72·10 ¹²), $b_6 = 5.209 \cdot 10^9$ (−4.283·10 ⁹ , 1.47·10 ¹⁰), $a_7 = -2.504 \cdot 10^9$ (−2.098·10 ¹¹ , 2.047·10 ¹¹), $b_7 = -2.443 \cdot 10^9$ (−1.241·10 ¹¹ , 1.192·10 ¹¹), $a_8 = 5.484 \cdot 10^7$ (−7.873·10 ⁹ , 7.982·10 ⁹), $b_8 = 2.28 \cdot 10^8$ (−1.38·10 ¹⁰ , 1.426·10 ¹⁰), $\omega = 0.009698$ (−0.03157, 0.05097)
Gaussian 1 (9)	$a_1 = 34.57$ (12.13, 57), $b_1 = -183.4$ (−1057, 689.9), $c_1 = 561.6$ (−470.3, 1594)
Gaussian 2 (10)	$a_1 = 26.05$ (15.26, 36.83), $b_1 = 14.46$ (8.582, 20.34), $c_1 = 44.1$ (35.12, 53.08), $a_2 = 27.11$ (25.91, 28.31), $b_2 = 103.9$ (98.54, 109.3), $c_2 = 70.06$ (27.97, 112.2)
Gaussian 3 (A2)	$a_1 = 5.155$ (4.873, 5.437), $b_1 = 26.16$ (25.76, 26.57), $c_1 = 9.728$ (9.011, 10.44), $a_2 = -1.671$ (−1.947, −1.396), $b_2 = 72.03$ (70.63, 73.44), $c_2 = 12.67$ (9.922, 15.43), $a_3 = 28.78$ (28.59, 28.98), $b_3 = 14.29$ (−15.42, 44), $c_3 = 410.3$ (269.1, 551.6)
Gaussian 7 (A3)	$a_1 = 4.437$ (3.671, 5.202), $b_1 = 26.49$ (25.87, 27.11), $c_1 = 8.597$ (7.324, 9.869), $a_2 = 29.6$ (28.86, 30.34), $b_2 = 22.63$ (13.09, 32.18), $c_2 = 107$ (68.95, 145.1), $a_3 = 0.4699$ (−0.03077, 0.9705), $b_3 = 42.14$ (40.67, 43.6), $c_3 = 1.975$ (−0.6027, 4.552), $a_4 = 1.863$ (0.6962, 3.029), $b_4 = 55.07$ (54.4, 55.75), $c_4 = 3.957$ (2.218, 5.697), $a_5 = 12.85$ (−13.58, 39.29), $b_5 = 119$ (62.36, 175.7), $c_5 = 29.6$ (−249.7, 308.9), $a_6 = 2.819$ (−70.61, 76.25), $b_6 = 85.94$ (23.4, 148.5), $c_6 = 17.51$ (−73.16, 108.2), $a_7 = 0.6927$ (−0.2962, 1.682), $b_7 = 65.57$ (63.04, 68.1), $c_7 =$ 4.351 (0.2728, 8.43)
Polynomial 1 (11)	$p_1 = -0.04331$ (−0.05269, −0.03393), $p_2 = 31.19$ (30.6, 31.78)
Polynomial 2 (12)	$p_1 = -7.924 \cdot 10^{-6}$ (−0.0003428, 0.000327), $p_2 = -0.04244$ (−0.08046, −0.004409), $p_3 = 31.17$ (30.27, 32.08)
Polynomial 3 (13)	$p_1 = 4.488 \cdot 10^{-5}$ (3.633·10 ^{−5} , 5.342·10 ^{−5}), $p_2 = -0.007412$ (−0.008842, −0.005983), $p_3 = 0.2848$ (0.217, 0.3527), $p_4 = 28.11$ (27.24, 28.97)
Power 1 (14)	$a = 32.62$ (31.08, 34.16), $b = -0.03351$ (−0.04599, −0.02103)
Power 2 (15)	$a = -0.0117$ (−0.05463, 0.03122), $b = 1.274$ (0.5054, 2.042), $c = 30.86$ (29.9, 31.81)
Rational 1/1 (16)	$p_1 = 28.8$ (28.39, 29.2), $p_2 = -45.01$ (−143.4, 53.39), $q_1 = -1.57$ (−4.969, 1.828)
Rational 2/1 (17)	$p_1 = -0.04398$ (−0.05356, −0.0344), $p_2 = 31.59$ (30.91, 32.27), $p_3 = -257.8$ (−277.1, −238.4), $q_1 = -8.275$ (−8.957, −7.593)
Rational 3/1 (18)	$p_1 = 4.597 \cdot 10^{-5}$ (−0.0002944, 0.0003864), $p_2 = -0.04994$ (−0.08985, −0.01003), $p_3 = 31.54$ (30.5, 32.58), $p_4 = -91.31$ (−111.2, −71.39), $q_1 = -2.9$ (−3.59, −2.21)
Rational 3/2 (19)	$p_1 = -0.04523$ (−0.05495, −0.03552), $p_2 = 31.92$ (31.18, 32.66), $p_3 = -416.9$ (−443.9, −390), $p_4 = 1157$ (972.2, 1342), $q_1 = -13.28$ (−14.23, −12.32), $q_2 = 37.01$ (30.4, 43.63)
Rational 5/3 (20)	$p_1 = 0.0004295$ (0.0002086, 0.0006504), $p_2 = -0.08024$ (−0.1215, −0.03895), $p_3 = 31.99$ (29.48, 34.49), $p_4 = -1545$ (−1626, −1465), $p_5 = 2.44 \cdot 10^4$ (2.175·10 ⁴ , 2.704·10 ⁴), $p_6 = -4.617 \cdot 10^4$ (−6.472·10 ⁴ , −2.762·10 ⁴), $q_1 = -52.8$ (−55.81, −49.79), $q_2 = 841.7$ (747.9, 935.5), $q_3 = -1592$ (−2210, −973.6)

Table A5. Quality measures for the considered functions (summers 1 and 2).

Fitness Function	SSE	R-Square	RMSE	Adjusted R-Square
Smoothing spline, $p = 0.99876718875$	0.0008646	1	0.02183	0.9999
Smoothing spline, $p = 0.9$	0.4474	0.999	0.1018	0.9976
$p = 0.309432$	1.994	0.9957	0.1614	0.9939
Gaussian 7	5.886	0.9872	0.2586	0.9843
Fourier 8	7.823	0.983	0.2932	0.9799
Gaussian 3	11.01	0.9761	0.3319	0.9742
Rational 5/3	19.51	0.9577	0.4418	0.9543
Fourier 2	59.3	0.8713	0.7588	0.8651
Gaussian 2	64.52	0.86	0.7915	0.8532
Fourier 1	119.1	0.7415	1.065	0.7341
Polynomial 3	127.1	0.7241	1.1	0.7163
Rational 3/2	245.1	0.4683	1.542	0.4425
Exponential of 2nd order	246.9	0.4642	1.533	0.4489
Rational 2/1	251.7	0.4538	1.548	0.4382
Rational 3/1	251.9	0.4533	1.556	0.4323
Power 2	254.9	0.447	1.551	0.4365
Gaussian 1	258	0.4402	1.56	0.4297
Polynomial 1	258.5	0.4392	1.554	0.4339
Polynomial 2	258.5	0.4392	1.561	0.4286
Exponential of 1st order	258.7	0.4387	1.555	0.4334
Linear fitting	275.9	0.4014	1.613	0.3901
Power 1	363.9	0.2103	1.844	0.2029
Rational 1/1	460.4	0.0009077	2.084	-0.01794

Table A6. R-square for the domain used to fit, prediction interval, and total range in case of summers.

Fitness Function	Fit-Domain (Summers 1 and 2)	Prediction Interval (Summers 3 and 4)	Total Range (all 4 Summers)	Comment
Smoothing spline $p = 0.9987671$	1	1	0.9995	The best fit
Smoothing spline $p = 0.9$	0.999	0.9978	0.996	Near best fit
Gaussian 7	0.9872	0.9931	0.9267	Best results when smoothing splines are excluded
Smoothing spline $p = 0.3094$	0.9957	0.9922	0.9858	Near best fit
Fourier 8	0.983	0.9873	0.9442	Best results when smoothing splines are excluded
Gaussian 3	0.9761	0.9555	0.8218	
Fourier 2	0.8713	0.9418	0.8169	
Exponential of 2nd order	0.4642	0.8828	0.5431	
Gaussian 2	0.86	0.8828	0.7003	
Rational 2/1	0.4516	0.8823	0.5421	
Polynomial 3	0.7241	0.8804	0.5431	
Rational 5/3	0.7961	0.8794	-4.337 **	
Rational 3/1	0.4539	0.8785	0.5471	
Fourier 1	0.7415	0.8784	0.5373	
Polynomial 2	0.4392	0.8784	0.5431	
Rational 1/1	0.0009076	0.8244	0.004622	
Power 1	0.2103	0.7884	0.3223	
Power 2	0.447	0.7884	0.5442	
Exponential of 1st order	0.4387	0.5604	0.5334	
Rational 3/2	0.5441	0.5583	0.5466	
Polynomial 1	0.4392	0.5289	0.5376	
Linear fitting	0.4014	0.2454	0.5263	
Gaussian 1	0.4402	N/A *	0.543	

* Infinity computed by the model function; fitting cannot continue. ** Results are not reliable due to data being not suitable for the function.

Table A7. Obtained coefficients for the considered functions (autumns 1–2 interval), values in brackets are confidence bounds for 95%.

Linear fitting (1)	$a = -0.04539$ (−0.3596, 0.2688), $b = 0.0002863$ (0.000223, 0.0003495), $c = 28.74$ (28.42, 29.06)
Exponential of 1st order (2)	$a = 28.59$ (28.07, 29.1), $b = 0.0006785$ (0.0004205, 0.0009364)
Exponential of 2nd order (3)	$a = 20.28$ (13.76, 26.8), $b = -0.01133$ (−0.01737, −0.005289), $c = 11.39$ (4.565, 18.22), and $d = 0.007709$ (0.004173, 0.01125)
Fourier 1 (5)	$a_0 = 30.23$ (29.9, 30.56), $a_1 = 1.47$ (1.117, 1.824), $b_1 = -1.839$ (−2.402, −1.276), $\omega = 0.04373$ (0.03873, 0.04873)
Fourier 2 (6)	$a_0 = 30.1$ (29.93, 30.27), $a_1 = 1.146$ (0.4727, 1.819), $b_1 = -1.75$ (−2.099, −1.401), $a_2 = -0.4874$ (−0.729, −0.2458), $b_2 = 0.3735$ (−0.194, 0.9409), $\omega = 0.04314$ (0.03784, 0.04844)
Fourier 8 (7)	$a_0 = 29.82$ (29.8, 29.84), $a_1 = 1.851$ (1.794, 1.907), $b_1 = -0.9324$ (−1.015, −0.8492), $a_2 = -0.3776$ (−0.4395, −0.3156), $b_2 = -0.3376$ (−0.3814, −0.2937), $a_3 = 0.1491$ (0.07512, 0.2231), $b_3 = -0.4124$ (−0.4374, −0.3874), $a_4 = -0.04963$ (−0.1024, 0.003145), $b_4 = -0.07979$ (−0.1323, −0.02726), $a_5 = -0.2298$ (−0.2614, −0.1982), $b_5 = 0.1685$ (0.101, 0.236), $a_6 = -0.01184$ (−0.06022, 0.03655), $b_6 = -0.1228$ (−0.1505, −0.09501), $a_7 = -0.1947$ (−0.2209, −0.1685), $b_7 = -0.005933$ (−0.09148, 0.07961), $a_8 = -0.07079$ (−0.1274, −0.0142), $b_8 = 0.09574$ (0.05283, 0.1386), $\omega = 0.05152$ (0.05085, 0.05219)
Gaussian 1 (9)	$a_1 = 4.026 \cdot 10^{96}$ (−1.429 · 10 ¹⁰³ , 1.429 · 10 ¹⁰³), $b_1 = 6.461 \cdot 10^5$ (−1.047 · 10 ¹⁰ , 1.047 · 10 ¹⁰), $c_1 = 4.365 \cdot 10^4$ (−3.536 · 10 ⁸ , 3.536 · 10 ⁸)
Gaussian 2 (10)	$a_1 = 25.59$ (17.38, 33.79), $b_1 = 123.3$ (120.4, 126.3), $c_1 = 55.5$ (43.33, 67.67), $a_2 = 30.31$ (29.55, 31.07), $b_2 = 1.598$ (−2.844, 6.039), $c_2 = 95.27$ (56.39, 134.1)
Gaussian 3 (A2)	$a_1 = 32.85$ (32.26, 33.43), $b_1 = 130$ (116.7, 143.3), $c_1 = 155$ (126, 184), $a_2 = 12.88$ (10.6, 15.15), $b_2 = -1.083$ (−3.304, 1.138), $c_2 = 17.47$ (11.68, 23.26), $a_3 = 7.65$ (4.925, 10.38), $b_3 = 27.5$ (21.28, 33.73), $c_3 = 20.62$ (15.55, 25.69)
Gaussian 7 (A3)	$a_1 = 32.69$ (30.32, 35.06), $b_1 = 112.7$ (108.1, 117.4), $c_1 = 27.42$ (18.18, 36.67), $a_2 = 30.76$ (29.99, 31.53), $b_2 = 3.096$ (−2.256, 8.448), $c_2 = 45.15$ (8.354, 81.94), $a_3 = 10.8$ (−35.29, 56.89), $b_3 = 82.33$ (80.03, 84.63), $c_3 = 11.69$ (2.008, 21.37), $a_4 = 1.048$ (−92.15, 94.25), $b_4 = 26.08$ (−82.89, 135.1), $c_4 = 11.07$ (−89.58, 111.7), $a_5 = 20.26$ (−1.597, 42.11), $b_5 = 63.03$ (56, 70.06), $c_5 = 20.46$ (−48.7, 89.62), $a_6 = 7.41$ (−95.68, 110.5), $b_6 = 37.05$ (−22.28, 96.38), $c_6 = 14.19$ (−81.59, 110), $a_7 = 4.07$ (1.256, 6.884), $b_7 = 95.85$ (94.26, 97.43), $c_7 = 7.985$ (6.273, 9.697)
Polynomial 1 (11)	$p_1 = 0.01943$ (0.01172, 0.02714), $p_2 = 28.62$ (28.09, 29.15)
Polynomial 2 (12)	$p_1 = 0.001234$ (0.001121, 0.001347), $p_2 = -0.1274$ (−0.1412, −0.1136), $p_3 = 31.56$ (31.2, 31.91)
Polynomial 3 (13)	$p_1 = -1.947 \cdot 10^{-6}$ (−5.712 · 10 ^{−6} , 1.818 · 10 ^{−6}), $p_2 = 0.001581$ (0.0009001, 0.002263), $p_3 = -0.144$ (−0.179, −0.109), $p_4 = 31.72$ (31.24, 32.21)
Power 1 (14)	$a = 29.11$ (27.92, 30.3), $b = 0.005937$ (−0.004536, 0.01641)
Power 2 (15)	$a = 1.284 \cdot 10^{-10}$ (−9.654 · 10 ^{−10} , 1.222 · 10 ^{−10}), $b = 5.096$ (3.291, 6.901), $c = 28.99$ (28.72, 29.27)
Rational 1/1 (16)	$p_1 = 29.69$ (29.3, 30.08), $p_2 = 17.66$ (−236.1, 271.4), $q_1 = 0.5237$ (−7.731, 8.778)
Rational 2/1 (17)	$p_1 = 1.366$ (−3.127, 5.859), $p_2 = -110.5$ (−574.6, 353.7), $p_3 = 3.269 \cdot 10^4$ (−8.198 · 10 ⁴ , 1.474 · 10 ⁵), $q_1 = 1033$ (−2601, 4666)
Rational 3/1 (18)	$p_1 = 0.001356$ (0.001122, 0.001589), $p_2 = -0.142$ (−0.166, −0.1181), $p_3 = 31.82$ (30.7, 32.94), $p_4 = 111.7$ (−392.1, 615.4), $q_1 = 3.678$ (−12.62, 19.98)
Rational 3/2 (19)	$p_1 = 0.01698$ (0.004851, 0.02911), $p_2 = 29.55$ (27.51, 31.58), $p_3 = -3671$ (95% confidence bounds: −4079, −3263), $p_4 = 1.327 \cdot 10^5$ (1.116 · 10 ⁵ , 1.539 · 10 ⁵), $q_1 = -118.8$ (−129.5, −108.1), $q_2 = 4350$ (3678, 5021)
Rational 5/3 (20)	$p_1 = 0.00135$ (0.001112, 0.001588), $p_2 = -0.1589$ (−0.1877, −0.1301), $p_3 = 33.7$ (31.57, 35.84), $p_4 = -317.1$ (−1029, 394.7), $p_5 = 38.35$ (−6943, 7020), $p_6 = 3728$ (−1.678e+04, 2.424e+04), $q_1 = -9.637$ (−32.02, 12.74), $q_2 = -0.6917$ (−224.9, 223.5), $q_3 = 123$ (−541, 786.9)

Table A8. Quality measures for the considered functions (autumns 1 and 2).

Fitness Function	SSE	R-Square	RMSE	Adjusted R-Square
Smoothing spline, $p = 0.99876718875$	0.0001338	1	0.008246	1
$p = 0.9$	0.07389	0.9997	0.03972	0.9994
$p = 0.309432$	0.4007	0.9986	0.0695	0.9981
Fourier 8	0.9464	0.9968	0.09728	0.9962
Gaussian 7	0.9249	0.9968	0.09765	0.9962
Gaussian 3	9.53	0.9674	0.2957	0.965
Rational 3/2	14.8	0.9494	0.3635	0.9471
Fourier 2	17.1	0.9415	0.3907	0.9389
Gaussian 2	19.35	0.9338	0.4157	0.9308
Fourier 1	32.92	0.8874	0.5374	0.8844
Rational 5/3	45.34	0.8449	0.6449	0.8335
Rational 3/1	45.43	0.8446	0.6341	0.8391
Polynomial 3	46.8	0.8399	0.6407	0.8357
Rational 2/1	46.98	0.8393	0.6419	0.8351
Polynomial 2	47.23	0.8384	0.6409	0.8356
Exponential of 2nd order	47.87	0.8362	0.648	0.8319
Power 2	126.3	0.5678	1.048	0.5603
Linear fitting	172.1	0.4114	1.223	0.4011
Exponential of 1st order	238.6	0.1838	1.434	0.1767
Gaussian 1	238.6	0.1838	1.44	0.1696
Polynomial 1	240.7	0.1767	1.44	0.1696
Rational 1/1	289.1	0.01099	1.586	-0.00621
Power 1	289.3	0.01042	1.579	0.001884

Table A9. R-square for domain used to fit, prediction interval, and total range (autumns).

Fitness Function	Fit-Domain (Autumns 1 and 2)	Prediction Interval (Autumns 3 and 4)	Total Range (all 4 Autumns)	Comment
Smoothing spline, $p = 0.99876718875$	1	1	1	The best fit
$p = 0.9$	0.9997	0.9971	0.9984	
$p = 0.309432$	0.9986	0.9897	0.9942	
Gaussian 7	0.9968	0.9859	0.9742	Best choice when smoothing splices are excluded.
Rational 3/2	0.9494	0.9741	0.6388	
Fourier 8	0.9968	0.9733	0.9494	Near best choice when smoothing splices are excluded.
Gaussian 3	0.9674	0.9484	0.8847	
Rational 5/3	0.8449	0.8914	0.6214	
Fourier 2	0.9415	0.8814	0.791	
Gaussian 2	0.9338	0.8795	0.7352	
Fourier 1	0.8874	0.8713	0.6388	
Polynomial 3	0.8399	0.865	0.6507	
Exponential of 2nd order	0.8362	0.8287	-4.861 *	* Matlab warning
Power 2	0.5678	0.7995	0.6424	
Gaussian 1	0.1838	0.7959	0.6381	
Rational 3/1	0.8446	0.7953	0.6968	
Polynomial 2	0.8384	0.7931	0.6388	
Rational 2/1	0.8393	0.7909	0.4735	
Polynomial 1	0.1767	0.7858	0.473	
Exponential of 1st order	0.1838	0.7812	0.4593	
Linear fitting	0.4114	0.7609	0.5964	
Power 1	0.01042	0.4893	0.2602	
Rational 1/1	0.01099	0.02938	0.4729	

Table A10. Obtained coefficients for the considered functions (winters 1–2 interval), values in brackets are confidence bounds for 95%.

Linear fitting (1)	$a = -0.02188$ (−0.2189, 0.1752), $b = 6.907 \cdot 10^{-5}$ ($1.449 \cdot 10^{-5}$, 0.0001237), $c = 28.12$ (27.92, 28.32)
Exponential of 1st order (2)	$a = 28.31$ (bounds: 28.02, 28.6), and $b = -6.32 \cdot 10^{-6}$ (−0.0001787, 0.0001661)
Exponential of 2nd order (3)	$a = 28.79$ (27.71, 29.87), $b = -0.00243$ (−0.003808, −0.001053), $c = 0.6529$ (−0.5924, 1.898), $d = 0.02388$ (0.01006, 0.0377)
Fourier 1 (5)	$a_0 = 28.16$ (28.09, 28.23), $a_1 = 0.5406$ (0.3579, 0.7233), $b_1 = 0.7318$ (0.5792, 0.8843), $\omega = 0.07365$ (0.06925, 0.07804)
Fourier 2 (6)	$a_0 = 28.44$ (28.31, 28.56), $a_1 = 0.8091$ (0.4632, 1.155), $b_1 = -0.3872$ (−0.8644, 0.09007), $a_2 = -0.3317$ (−0.4104, −0.2531), $b_2 = 0.1166$ (−0.3926, 0.6258), $\omega = 0.05057$ (0.04012, 0.06102)
Fourier 8 (7)	$a_0 = 28.4$ (28.31, 28.49), $a_1 = 0.9113$ (0.755, 1.068), $b_1 = -0.2112$ (−0.4711, 0.04881), $a_2 = -0.2993$ (−0.4587, −0.1399), $b_2 = -0.06535$ (−0.2176, 0.08691), $a_3 = -0.04238$ (−0.1884, 0.1037), $b_3 = -0.1498$ (−0.2659, −0.03362), $a_4 = 0.09473$ (−0.03816, 0.2276), $b_4 = -0.01995$ (−0.1303, 0.09044), $a_5 = -0.1726$ (−0.2623, −0.0828), $b_5 = 0.1299$ (−0.1434, 0.4033), $a_6 = -0.02845$ (−0.09781, 0.04091), $b_6 = 0.0155$ (−0.07253, 0.1035), $a_7 = 0.07436$ (−0.0176, 0.1663), $b_7 = -0.004225$ (−0.09292, 0.08447), $a_8 = -0.06261$ (−0.2146, 0.08941), $b_8 = 0.1033$ (−0.0578, 0.2644), $\omega = 0.05442$ (0.05006, 0.05879)
Gaussian 1 (9)	$a_1 = 28.35$ (−60.17, 116.9), $b_1 = -637.7$ (−1.061·10 ⁶ , 1.06·10 ⁶), $c_1 = 1.529 \cdot 10^4$ (−1.175·10 ⁷ , 1.178·10 ⁷)
Gaussian 2 (10)	$a_1 = 19.57$ (7.279, 31.86), $b_1 = 116.5$ (110.4, 122.6), $c_1 = 51.55$ (32.7, 70.41), $a_2 = 28.71$ (27.94, 29.48), $b_2 = 7.058$ (0.8335, 13.28), $c_2 = 97.94$ (51.13, 144.8)
Gaussian 3 (A2)	$a_1 = 29.18$ (26.94, 31.43), $b_1 = 113$ (106.9, 119.2), $c_1 = 102.8$ (15.34, 190.2), $a_2 = 21.12$ (3.491, 38.75), $b_2 = -13.39$ (−19.73, −7.037), $c_2 = 61.1$ (34.01, 88.18), $a_3 = 1.117$ (0.893, 1.34), $b_3 = 31.73$ (30.94, 32.53), $c_3 = 5.535$ (4.146, 6.924)
Gaussian 7 (A3)	$a_1 = 29.83$ (28.1, 31.57), $b_1 = 117.1$ (64.29, 170), $c_1 = 140.8$ (−85.14, 366.8), $a_2 = 14.57$ (−13.05, 42.2), $b_2 = -9.303$ (−34.14, 15.53), $c_2 = 44.71$ (−6.636, 96.06), $a_3 = 1.096$ (0.7929, 1.4), $b_3 = 34.79$ (34.52, 35.07), $c_3 = 1.847$ (1.272, 2.422), $a_4 = 0.5781$ (−0.6638, 1.82), $b_4 = 46.14$ (42.81, 49.47), $c_4 = 6.095$ (0.2361, 11.95), $a_5 = 1.61$ (−0.7591, 3.98), $b_5 = 29.99$ (26.84, 33.14), $c_5 = 9.895$ (3.158, 16.63), $a_6 = 0.8952$ (−1.199, 2.99), $b_6 = 56.75$ (46.22, 67.27), $c_6 = 11.02$ (−5.995, 28.03), $a_7 = 0.6716$ (0.3439, 0.9993), $b_7 = 79$ (78.26, 79.75), $c_7 = 4.245$ (2.401, 6.089)
Polynomial 1 (11)	$p_1 = -0.0001754$ (−0.005053, 0.004702), $p_2 = 28.31$ (28.02, 28.6)
Polynomial 2 (12)	$p_1 = 0.0007765$ (0.0006731, 0.0008799), $p_2 = -0.08015$ (−0.09115, −0.06916), $p_3 = 29.69$ (29.45, 29.94)
Polynomial 3 (13)	$p_1 = 8.683e-06$ (5.058e−06, 1.231e−05), $p_2 = -0.000565$ (−0.001133, 2.833e−06), $p_3 = -0.02461$ (−0.04985, 0.0006242), $p_4 = 29.2$ (28.9, 29.51)
Power 1 (14)	$a = 28.99$ (28.41, 29.57), $b = -0.006629$ (−0.01191, −0.001349)
Power 2 (15)	$a = 1.526$ (bounds: 0.4475, 2.604), $b = -0.4284$ (−1.337, 0.4801), and $c = 27.95$ (26.86, 29.03)
Rational 1/1 (16)	$p_1 = 28.28$ (28.14, 28.43), $p_2 = -50.27$ (−80.95, −19.59), $q_1 = -1.783$ (−2.842, −0.7249)
Rational 2/1 (17)	$p_1 = 0.0004235$ (bounds: −0.00459, 0.005437), $p_2 = 28.26$ (27.95, 28.57), $p_3 = -50.2$ (−80.05, −20.35), $q_1 = -1.782$ (−2.812, −0.7532)
Rational 3/1 (18)	$p_1 = 0.001457$ (0.0008546, 0.002059), $p_2 = -0.1586$ (−0.2211, −0.09621), $p_3 = 31.29$ (30.13, 32.46), $p_4 = 735.3$ (−433.5, 1904), $q_1 = 25.65$ (−14.72, 66.02)
Rational 3/2 (19)	$p_1 = 0.0005795$ (−0.004503, 0.005662), $p_2 = 28.25$ (27.88, 28.61), $p_3 = -338.4$ (−388.7, −288.1), $p_4 = 894$ (bounds: 596.2, 1192), $q_1 = -11.98$ (−13.69, −10.27), $q_2 = 31.66$ (21.45, 41.86)
Rational 5/3 (20)	$p_1 = 0.001389$ (0.0009086, 0.001869), $p_2 = -0.1762$ (−0.2345, −0.1179), $p_3 = 34.02$ (32.02, 36.01), $p_4 = -70.51$ (−935, 793.9), $p_5 = -2716$ (−1.161·10 ⁴ , 6177), $p_6 = 1.366 \cdot 10^4$ (bounds: −1.202·10 ⁴ , 3.934·10 ⁴), $q_1 = -0.4298$ (−30.31, 29.45), $q_2 = -103.4$ (−412.6, 205.8), $q_3 = 487.4$ (−412.4, 1387)

Table A11. Quality measures for the considered functions (winters 1 and 2).

Fitness Function	SSE	R-Square	RMSE	Adjusted R-Square
Smoothing spline, $p = 0.99876718875$	0.0005719	1	0.01837	0.9994
$p = 0.9$	0.2665	0.995	0.08124	0.9875
$p = 0.309432$	0.9664	0.9819	0.1162	0.9745
Gaussian 7	1.102	0.9794	0.1167	0.9743
Fourier 8	1.997	0.9626	0.1542	0.9551
Fourier 2	7.085	0.8675	0.2717	0.8605
Gaussian 2	8.046	0.8495	0.2895	0.8416
Fourier 1	8.601	0.8391	0.2962	0.8342
Rational 5/3	11.54	0.7841	0.3523	0.7655
Rational 3/1	11.6	0.7831	0.3457	0.7741
Polynomial 3	13.4	0.7494	0.3697	0.7417
Exponential of 2nd order	14.6	0.7268	0.386	0.7184
Polynomial 2	16.49	0.6916	0.4081	0.6853
Linear fitting	50.23	0.06032	0.7123	0.04134
Power 1	50.38	0.05741	0.7098	0.04798
Rational 3/2	52.47	0.01834	0.7393	-0.03279 *
Rational 2/1	52.78	0.01254	0.7339	-0.01768 *
Rational 1/1	52.8	0.01226	0.7303	-0.007692 *
Gaussian 3	3.794	0.929	0.202	0.9229
Power 2	49.4	0.0758	0.7064	0.05713
Exponential of 1st order	53.45	$5.189 \cdot 10^{-5}$	0.7311	-0.009948 *
Polynomial 1	53.45	$5.09 \cdot 10^{-5}$	0.7311	-0.009949 *
Gaussian 1	53.46	-0.0001639 *	0.7348	-0.02037 *

* only calculated data by Matlab without interpretation (negative numbers should be impossible).

Table A12. R-square for the domain used to fit, prediction interval, and total range for winters.

Fitness Function	Fit-Domain (Winters 1-2)	Prediction Interval (Winters 3-4)	Total Range (all 4 Winters)	Comment
Smoothing spline, $p = 0.99876718875$	1	1	1	The best fit
$p = 0.9$	0.995	0.9978	0.9985	
Gaussian 7	0.9794	0.9931	0.9267	
Smoothing spline $p = 0.309432$	0.9819	0.9922	0.9945	
Fourier 8	0.9626	0.9873	0.9761	
Gaussian 3	0.929	0.9555	0.9466	
Fourier 2	0.8675	0.9418	0.9354	
Rational 5/3	0.7841	0.8829	0.9429	
Exponential of 2nd order	0.7268	0.8828	0.8571	
Gaussian 2	0.8495	0.8828	0.9104	
Polynomial 3	0.7494	0.8804	0.8665	
Fourier 1	0.8391	0.8784	0.8458	
Polynomial 2	0.6916	0.8784	0.8458	
Rational 3/1	0.7831	0.881	0.8477	
Power 1	0.05741	0.7884	0.3671	
Power 2	0.0758	0.7884	0.8614	
Exponential of 1st order	$5.189 \cdot 10^{-5}$	0.5604	0.656	
Rational 3/2	0.01834	0.5397	0.9403	
Rational 2/1	0.01254	0.5332	0.6771	
Polynomial 1	$5.09 \cdot 10^{-5}$	0.5289	0.6745	
Linear fitting	0.06032	0.2454	0.816	
Rational 1/1	0.01226	0.01329	0.0109	
Gaussian 1	-0.0001639 *	N/A	0.841	Inf computed by model function, fitting cannot continue.

* Matlab warning: A negative R-square is possible if the model does not contain a constant term and the fit is poor (worse than just fitting the mean). Try changing the model or using a different start point.

Table A13. Obtained coefficients for the considered functions (springs 1–2 interval), values in brackets are confidence bounds for 95%.

Linear fitting (1)	$a = -0.001745$ (−0.1453, 0.1418), $b = -2.716 \cdot 10^{-5}$ (−6.444·10 ^{−5} , 1.011·10 ^{−5}), $c = 29.09$ (28.94, 29.23)
Exponential of 1st order (2)	$a = 28.97$ (28.77, 29.18), $b = 2.768 \cdot 10^{-5}$ (−8.86·10 ^{−5} , 0.000144)
Exponential of 2nd order (3)	$a = -0.01843$ (−0.07621, 0.03936), $b = 0.04838$ (0.02118, 0.07558), $c = 28.47$ (28.28, 28.65), $d = 0.0007368$ (0.0003567, 0.001117)
Fourier 1 (5)	$a_0 = 29.09$ (29.02, 29.16), $a_1 = 0.02635$ (−0.1644, 0.2171), $b_1 = -0.5724$ (−0.6691, −0.4757), $\omega = 0.07486$ (0.06923, 0.08049)
Fourier 2 (6)	$a_0 = 29.12$ (29.06, 29.17), $a_1 = 0.0604$ (−0.05226, 0.1731), $b_1 = -0.5426$ (−0.6154, −0.4699), $a_2 = -0.1226$ (−0.2415, −0.003689), $b_2 = -0.3169$ (−0.4064, −0.2273), $\omega = 0.07436$ (0.07171, 0.077)
Fourier 8 (7)	$a_0 = 28.99$ (28.94, 29.04), $a_1 = -0.502$ (−0.5794, −0.4245), $b_1 = -0.1689$ (−0.3016, −0.03619), $a_2 = -0.0574$ (−0.2131, 0.09831), $b_2 = 0.3611$ (0.3103, 0.412), $a_3 = 0.199$ (0.1409, 0.2571), $b_3 = -0.03921$ (−0.1616, 0.08319), $a_4 = -0.03603$ (−0.08402, 0.01196), $b_4 = 0.09401$ (0.05168, 0.1363), $a_5 = 0.08342$ (−0.07823, 0.2451), $b_5 = -0.1118$ (−0.17, −0.05353), $a_6 = -0.1135$ (−0.2218, −0.005219), $b_6 = 0.1204$ (−0.04748, 0.2884), $a_7 = 0.06244$ (0.02226, 0.1026), $b_7 = -0.004121$ (−0.08272, 0.07448), $a_8 = -0.01616$ (−0.09508, 0.06275), $b_8 = 0.07005$ (0.02402, 0.1161), $\omega = 0.05696$ (0.05373, 0.06018)
Gaussian 1 (9)	$a_1 = 29.38$ (29.26, 29.5), $b_1 = 54.13$ (50.75, 57.51), $c_1 = 270$ (236.6, 303.5)
Gaussian 2 (10)	$a_1 = 29.57$ (29.36, 29.77), $b_1 = 68.9$ (59.9, 77.91), $c_1 = 144.9$ (107.5, 182.2), $a_2 = 6.047$ (−1.507, 13.6), $b_2 = -15.25$ (−45.87, 15.36), $c_2 = 36.4$ (8.628, 64.18)
Gaussian 3 (A2)	$a_1 = 1.419$ (0.35, 2.487), $b_1 = 70.92$ (66.05, 75.8), $c_1 = 14.86$ (9.047, 20.67), $a_2 = 78.21$ (−4.145·10 ⁵ , 4.147·10 ⁵), $b_2 = 6155$ (−3.252·10 ⁷ , 3.253·10 ⁷), $c_2 = 6144$ (−1.623·10 ⁷ , 1.625·10 ⁷), $a_3 = -1.338$ (−44.12, 41.44), $b_3 = 101.9$ (−435.5, 639.2), $c_3 = 41.61$ (−449.4, 532.6)
Gaussian 7 (A3)	$a_1 = 1.073$ (0.8171, 1.329), $b_1 = 60.96$ (60.51, 61.42), $c_1 = 3.182$ (2.307, 4.057), $a_2 = 29.82$ (29.74, 29.9), $b_2 = 71.53$ (70.08, 72.99), $c_2 = 79.17$ (66.86, 91.48), $a_3 = 1.615$ (−19.11, 22.34), $b_3 = 43.72$ (27.25, 60.19), $c_3 = 7.56$ (−7.447, 22.57), $a_4 = 15.4$ (−58.75, 89.55), $b_4 = -4.202$ (−90.21, 81.8), $c_4 = 18.05$ (−271.1, 307.2), $a_5 = 6.021$ (−266.5, 278.5), $b_5 = 18.45$ (−40.3, 77.2), $c_5 = 13.43$ (−288.4, 315.2), $a_6 = 3.458$ (1.835, 5.081), $b_6 = 107.5$ (103.7, 111.3), $c_6 = 10.99$ (7.16, 14.81), $a_7 = 3.096$ (−142.1, 148.3), $b_7 = 32.69$ (−91.44, 156.8), $c_7 = 10.67$ (−113.4, 134.8)
Polynomial 1 (11)	$p_1 = 0.0008112$ (−0.002563, 0.004185), $p_2 = 28.97$ (28.77, 29.18)
Polynomial 2 (12)	$p_1 = -0.0003957$ (−0.0004937, −0.0002976), $p_2 = 0.04275$ (0.03203, 0.05348), $p_3 = 28.22$ (27.98, 28.47)
Polynomial 3 (13)	$p_1 = -1.111 \cdot 10^{-5}$ (−1.41·10 ^{−5} , −8.131·10 ^{−6}), $p_2 = 0.001371$ (0.0008905, 0.001852), $p_3 = -0.03252$ (−0.05451, −0.01053), $p_4 = 28.9$ (28.63, 29.17)
Power 1 (14)	$a = 28.65$ (28.23, 29.06), $b = 0.003467$ (−0.0003017, 0.007236)
Power 2 (15)	$a = -0.8686$ (−3.288, 1.551), $b = -0.2256$ (−1.632, 1.181), $c = 29.4$ (26.49, 32.32)
Rational 1/1 (16)	$p_1 = 29.02$ (confidence bounds: 28.91, 29.12), $p_2 = -50.97$ (−125.8, 23.84), $q_1 = -1.755$ (−4.35, 0.8403)
Rational 2/1 (17)	$p_1 = 0.0006562$ (confidence bounds: −0.002789, 0.004101), $p_2 = 28.98$ (28.74, 29.22), $p_3 = -257$ (−303.3, −210.7), $q_1 = -8.865$ (−10.48, −7.253)
Rational 3/1 (18)	$p_1 = -0.000405$ (−0.0005055, −0.0003045), $p_2 = 0.0456$ (0.03414, 0.05707), $p_3 = 28.01$ (27.72, 28.31), $p_4 = -116.4$ (−147.3, −85.38), $q_1 = -4.125$ (−5.196, −3.054)
Rational 3/2 (19)	$p_1 = -18.77$ (−8290, 8252), $p_2 = 2122$ (−9.201·10 ⁵ , 9.244·10 ⁵), $p_3 = 1.299 \cdot 10^6$ (−5.72·10 ⁸ , 5.746·10 ⁸), $p_4 = -3.706 \cdot 10^6$ (−1.638·10 ⁹ , 1.631·10 ⁹), $q_1 = 4.631 \cdot 10^4$ (−2.038·10 ⁷ , 2.048·10 ⁷), $q_2 = -1.316 \cdot 10^5$ (−5.817·10 ⁷ , 5.791·10 ⁷)
Rational 5/3 (20)	$p_1 = -8.931 \cdot 10^{-5}$ (−0.0003049, 0.0001263), $p_2 = 0.01715$ (−0.03251, 0.0668), $p_3 = 27.42$ (23.49, 31.35), $p_4 = -4117$ (−4561, −3673), $p_5 = 1.603e+05$ (1.304·10 ⁵ , 1.902·10 ⁵), $p_6 = -2.4 \cdot 10^5$ (−4.535·10 ⁵ , −2.649·10 ⁴), $q_1 = -145$ (bounds: −158.3, −131.6), $q_2 = 5597$ (4569, 6624), $q_3 = -8388$ (−1.586·10 ⁴ , −918.6)

Table A14. Quality measures for the considered functions (domain used for prediction) for springs 1–2.

Fitness Function	SSE	R-Square	RMSE	Adjusted R-Square
Smoothing spline, $p = 0.99876718875$	0.0005042	1	0.01699	0.999
$p = 0.9$	0.2488	0.9914	0.07735	0.9784
$p = 0.309432$	0.8244	0.9714	0.1058	0.9596
Gaussian 7	1.407	0.9512	0.1294	0.9395
Fourier 8	1.705	0.9408	0.14	0.9293
Gaussian 3	5.626	0.8047	0.2421	0.7885
Rational 5/3	6.391	0.7782	0.258	0.7598
Fourier 2	6.445	0.7763	0.2552	0.765
Gaussian 2	10.64	0.6307	0.3278	0.6121
Polynomial 3	11.46	0.6023	0.3368	0.5905
Fourier 1	12.16	0.5781	0.347	0.5655
Exponential of 2nd order	13.64	0.5265	0.3675	0.5125
Rational 3/1	17.45	0.3945	0.4177	0.3703
Rational 3/2	17.48	0.3933	0.4202	0.3627
Gaussian 1	17.63	0.3882	0.4157	0.3762
Polynomial 2	17.66	0.3872	0.4161	0.3752
Power 2	27.83	0.03406	0.5224	0.01512
Power 1	27.91	0.03153	0.5205	0.02213
Linear fitting	28.24	0.0201	0.5261	0.0008901
Rational 2/1	28.62	0.006746	0.5323	-0.02276
Polynomial 1	28.75	0.002203	0.5284	-0.007485
Exponential of 1st order	28.75	0.002181	0.5284	-0.007507 *
Rational 1/1	28.76	0.001936	0.531	-0.01763

* Matlab warning: A negative R-square is possible if the model does not contain a constant term and the fit is poor (worse than just fitting the mean). Try changing the model or using a different start point.

Table A15. R-square for the domain used to fit, prediction interval, and total range in case of springs.

Fitness Function	Fit-Domain (Springs 1–2)	Prediction Interval (Springs 3–4)	Total Range (all 4 Springs)	Comment
Smoothing spline, $p = 0.99876718875$	1	1	1	The best fit
$p = 0.9$	0.9914	0.995	0.9918	
$p = 0.309432$	0.9714	0.9827	0.9717	
Fourier 8	0.9408	0.9596	0.8896	The best for spring 3–4 when smoothing splines are excluded
Gaussian 7	0.9512	0.9501	0.9062	The best for whole dataset and springs 1–2 when smoothing splines are excluded
Gaussian 3	0.8047	0.9345	0.5874	
Rational 3/2	0.3933	0.9174	0.8254	
Fourier 2	0.7763	0.8365	0.5747	
Gaussian 2	0.6307	0.8252	0.5659	
Fourier 1	0.5781	0.7325	0.3169	
Rational 5/3	0.7782	0.7256	-66.28 *	
Rational 3/1	0.3945	0.7185	0.2409	
Polynomial 3	0.6023	0.7102	0.2664	
Polynomial 2	0.3872	0.71	0.2399	
Exponential of 2nd order	0.5265	0.706	0.2258	
Power 2	0.03406	0.5215	0.2072	
Power 1	0.03153	0.5181	0.04345	
Exponential of 1st order	0.002181	0.3752	0.1342	
Gaussian 1	0.3882	0.3752	0.2458	
Rational 2/1	0.006746	0.3558	0.1383	
Polynomial 1	0.002203	0.3518	0.1375	
Linear fitting	0.0201	0.1475	0.1975	
Rational 1/1	0.001936	0.02794	0.0001715	

* A negative R-square is possible if the model does not contain a constant term and the fit is poor (worse than just fitting the mean). Try changing the model or using a different start point.

References

1. Marques, C.H.; Caprace, J.-D.; Belchior, C.R.P.; Martini, A. An Approach for Predicting the Specific Fuel Consumption of Dual-fuel Two-stroke Marine Engines. *J. Mar. Sci. Eng.* **2019**, *7*, 20. [CrossRef]

2. Tran, T.A. Design the Prediction Model of Low-Sulfur-Content Fuel Oil Consumption for M/V NORD VENUS 80,000 DWT Sailing on Emission Control Areas by Artificial Neural Networks. In Proceedings of the Institution of Mechanical Engineers, Part M. *J. Eng. Marit. Environ.* **2019**, *233*, 345–362.
3. Krčum, M.; Gudelj, A.; Tomas, V. Optimal Design of Ship's Hybrid Power System for Efficient Energy. *Trans. Marit. Sci.* **2018**, *7*, 23–32. [[CrossRef](#)]
4. Geertsma, R.D.; Visser, K.; Negenborn, R.R. Adaptive pitch control for ships with diesel mechanical and hybrid propulsion. *Appl. Energy* **2018**, *228*, 2490–2509. [[CrossRef](#)]
5. Aronietis, R.; Sys, C.; van Hassel, E.; Vanelslander, T. Forecasting Port-level Demand for LNG as a Ship Fuel: The Case of the Port of Antwerp. *J. Shipp. Trade* **2016**, *1*, 2. [[CrossRef](#)]
6. Tsujimoto, M.; Sogihara, N. Prediction of Fuel Consumption of Ships in Actual Seas. Available online: http://www.naoe.eng.osaka-u.ac.jp/kashi/SOEMeeting/PPT/SOE1_Tsujimoto.pdf (accessed on 5 November 2019).
7. Górski, W.; Abramowicz-Gerigk, T.; Burciu, Z. The Influence of Ship Operational Parameters on Fuel Consumption. *Zesz. Nauk.* **2013**, *36*, 49–54.
8. Bialystocki, N.; Konovessis, D. On the Estimation of Ship's Fuel Consumption and Speed Curve: A Statistical Approach. *J. Ocean Eng. Sci.* **2016**, *1*, 157–166. [[CrossRef](#)]
9. Lu, R.; Turan, O.; Boulougouris, E. Voyage Optimisation: Prediction of Ship Specific Fuel Consumption for Energy Efficient Shipping. In Proceedings of the 3rd International Conference on Technologies, Operations, Logistics and Modelling for Low Carbon Shipping, London, UK, 9–10 September 2013.
10. Banawan, A.A.; Mosleh, M.; Seddiek, I.S. Prediction of the Fuel Saving and Emissions Reduction by Decreasing Speed of a Catamaran. *J. Mar. Eng. Technol.* **2014**, *12*, 40–48.
11. Larsen, U.; Pierobon, L.; Baldi, F.; Haglind, F.; Ivarsson, A. Development of a Model for the Prediction of the Fuel Consumption and Nitrogen Oxides Emission Trade-off for Large Ships. *Energy* **2015**, *80*, 545–555. [[CrossRef](#)]
12. Gheriani, E. Fuel Consumption Prediction Methodology for Early Stages of Naval Ship Design. Master's Thesis, MIT, Massachusetts Institute of Technology, Cambridge, MA, USA, 2012.
13. Pétursson, S. Predicting Optimal Trim Configuration of Marine Vessels with Respect to Fuel Usage. Master's Thesis, University of Iceland, Reykjavik, Iceland, 2009.
14. Borkowski, T.; Kasyk, L.; Kowalak, P. Assessment of Ship's Engine Effective Power Fuel Consumption and Emission using the Vessel Speed. *J. KONES Powertrain Transp.* **2011**, *18*, 31–39.
15. Demirel, A.K.; Turan, O.; Incecik, A. Predicting the effect of biofouling on ship resistance using CFD. *Appl. Ocean Res.* **2017**, *62*, 100–118. [[CrossRef](#)]
16. Kan, Z.; Tang, L.; Kwan, M.P.; Zhang, X. Estimating Vehicle Fuel Consumption and Emissions Using GPS Big Data. *Int. J. Environ. Res. Public Health* **2018**, *15*, 566. [[CrossRef](#)] [[PubMed](#)]
17. Zhao, Q.; Chen, Q.; Wang, L. Real-Time Prediction of Fuel Consumption Based on Digital Map API. *Appl. Sci.* **2019**, *9*, 1369. [[CrossRef](#)]
18. Kee, K.K.; Simon, B.Y.L.; Renco, K.H.Y. Prediction of Ship Fuel Consumption and Speed Curve by Using Statistical Method. *J. Comp. Sci. Comput. Math.* **2018**, *8*, 19–24. [[CrossRef](#)]
19. Hu, Z.; Jin, Y.; Hu, Q.; Sen, S.; Zhou, T.; Osman, M.T. Prediction of Fuel Consumption for Enroute Ship Based on Machine Learning. *IEEE Access* **2019**, *7*, 119497–119505. [[CrossRef](#)]
20. Wang, S.; Ji, B.; Zhao, J.; Liu, W.; Xu, T. Predicting Ship Fuel Consumption based on LASSO Regression. *Transp. Res. D Transp. Environ.* **2018**, *65*, 817–824. [[CrossRef](#)]
21. Ahlgren, F.; Mondejar, M.E.; Thern, M. Predicting Dynamic Fuel Oil Consumption on Ships with Automated Machine Learning. *Energy Procedia* **2019**, *158*, 6126–6131. [[CrossRef](#)]
22. Panapakidis, I.; Sourtzi, V.-M.; Dagoumas, A. Forecasting the Fuel Consumption of Passenger Ships with a Combination of Shallow and Deep Learning. *Electronics* **2020**, *9*, 776. [[CrossRef](#)]
23. Uyanik, T.; Arslanoglu, Y.; Kalenderli, O. Ship Fuel Consumption Prediction with Machine Learning. In Proceedings of the 4th International Mediterranean Science and Engineering Congress, Antalya, Turkey, 25–27 April 2019; pp. 757–759.
24. Mosaic Data Science, Predicting Fuel Usage Anomalies, Case Study. Available online: <http://www.mosaicdatascience.com/2018/05/11/predicting-fuel-usage-anomalies/> (accessed on 20 October 2019).
25. Coraddu, A.; Oneto, L.; Baldi, F.; Anguita, D. Vessels Fuel Consumption: A Data Analytics Perspective to Sustainability. In *Soft Computing for Sustainability Science; Studies in Fuzziness and Soft Computing*; Cruz Corona, C., Ed.; Springer: Cham, Switzerland, 2018; pp. 11–48.

26. MathWorks Inc. *Curve Fitting Toolbox™ User's Guide R2018b*; MathWorks Inc.: Natick, MA, USA, 2018.
27. Oppenheim, A.V.; Verghese, G.C. *Signals, Systems and Inference*; Pearson Education Limited: New York, NY, USA, 2016.
28. Wang, Z.; Bovik, A.C. Mean Squared Error: Love It or Leave It? *IEEE Signal Proc. Mag.* **2009**, *26*, 98–117. [[CrossRef](#)]
29. Botchkarev, A. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdiscip. J. Inf. Knowl. Manag.* **2019**, *4*, 45–76. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).