CHALMERS

Prediction of coast-down test results

A statistical study of environmental influences

Master of Science Thesis

Peter Norrby

Department of Product and Production Development
Division of Product Development
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2012

A THESIS FOR THE DEGREE OF MASTER OF SCIENCE

Prediction of coast-down test results

A statistical study of environmental influences

Peter Norrby

Department of Product and Production Development
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2012

Prediction of coast-down test results

A statistical study of environmental influences

Peter Norrby

Department of Product and Production Development

Chalmers University of Technology

SE-412 96 Gothenburg

Sweden

Telephone +46 (0)31-772 1000

Printed by Chalmers Reproservice

Gothenburg, Sweden, 2012

Prediction of coast-down test results
A statistical study of environmental influences
Peter Norrby
Department of Product and Production Development
Division of Product Development
Chalmers University of Technology

Abstract

In order to measure a car's total road load and thereby form the basis of the determination of the car's certified fuel consumption, Volvo Car Corporation (VCC) performs coast-down tests. Despite thorough checks, fine adjustment of the car and well documented weather conditions there is a great inconsistency in the results. Large differences in road load for the same car model means that it is possible to obtain a lower load with a car with theoretically higher road load, which in turn creates problems in the internal development. In a scenario when a car is performing very well in a coast down test and the cause is not known, it may require very large and costly improvements for the next model to reach the same low result, although it in theory easily would perform better than the old car. This is because it is not known what factors influenced the first model in such a way that it suddenly delivered a very low road load.

The purpose of this master's thesis is to find and understand the parameters that affect the coast-down result and predict the most accurate road load at the given circumstances, so that coast-down expeditions can be done with as few and effective test runs as possible and thereby make the expedition quicker, cheaper and with more precise and reliable result. The goal is to fulfil this with a stable, mathematical model that describes the true road load within +/- 5% with 95% confidence.

The goal was achieved by collecting and compiling data from three coast-down expeditions and performing multiple regressions analyses on the dataset with a model developed by literature studies and expertise at VCC. The dataset and model were analyzed and further developed by residual analyses, F-tests, t-tests, correlation analyses and VIF-tests. The final regression model was used on three different subsets of data, one for each coast-down expedition, in order to study the stability of the regression models.

With the final regression model

$$F = \underbrace{N \cdot (C_{temp} \cdot T + C_{time} \cdot Time)}_{F_{RR}} + \underbrace{(C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot AIR}_{F_{AIR}} + \underbrace{Direc + D1 + ... + D11}_{Dummies}$$

the goal of this master thesis was met by explaining just over 96% of the total variation in the coast-down results and thus describing the true road load within less than +/- 4%. The coefficients (C_{temp}, C_{time} etc.) were found significant and rather stable. The model can be used to normalize the boundary conditions at a coast-down expedition in order to investigate whether the obtained results are representative in relation to the circumstances or not.

Keywords: Coast-down expedition; Boundary conditions; Multiple regression analysis; Vehicle dynamics; Statistics.

Preface

This thesis is a part of the requirements for the master's degree at Chalmers University of Technology, Gothenburg, and has been carried out at the Division of Product Development, Department of Product and Production Development, Chalmers University of Technology and at the Division of Fuel Economy Analysis, Volvo Car Corporation during the spring of 2012.

I would like to acknowledge and thank my examiner and supervisor, Dr. Lars Lindkvist at the Department of Product and Production Development, for supportive feedback and my co-supervisor, Björn Lindenberg at the Division of Fuel Economy Analysis, Volvo Car Corporation for his help and guidance during the thesis. I would also like to thank the Coast-down team at the Division of Fuel Economy Analysis, Volvo Car Corporation and especially Erik Carlsson for being very helpful with the input data and analysis of the results. Finally, I would also like to thank Kristina Wärmefjord at the Department of Product and Production Development for statistical support and for enduring all my questions.

Gothenburg, June, 2012
Peter Norrby

Nomenclature

APG – Arizona Proving Ground
AWD – All Wheel Drive
Chalmers – Chalmers University of Technology
Eq. – Equation
Fig. – Figure
FWD – Front-Wheel Drive
U.S. – Unites States of America
VCC – Volvo Car Corporation

Table of contents

1. Introduction

This chapter gives an introduction to this master thesis by providing a description of the problem's background and what purpose and goal that was set up in order to solve the problem.

1.1. Background

In order to measure a car's total road load, coast-down tests are performed. The result from the coast-down tests form the basis of the determination of the car's certified fuel consumption. At Volvo Car Corporation (VCC) these tests are today performed at the Arizona Proving Ground (APG) in the southern U.S.A.. Fig. 1 shows the long, straight road and the black road with the hexagonal area were the tests are performed.

Fig. 1. *The Arizona Proving Ground*

Despite thorough checks, fine adjustment of the car and well documented weather conditions there is a large inconsistency in the results. Large differences in road load for the same car model means that it is possible to obtain a lower load with a car with theoretically higher road load, which in turn creates problems in the internal development. In a scenario when a car is performing very well in a coast down test and the cause is not known, it may require very large and costly improvements for the next model to reach the same low result, although it in theory easily would perform better than the old car. This is because it is not known what factors influenced the first model in such a way that it suddenly delivered a very low road load.

Random checks are made by authorities in the U.S in order to check if the test results stated by the car manufactures correspond to the reality and thereby be able to determine whether there has been unallowable actions or not. Knowing the parameters that affect the coast-down result makes it easier for VCC to argue that the test results are really representative. That is, if it can be proved that the certificated coast-down result done by VCC was performed during different circumstances than for the authorities' test.

Knowing what parameters that affects the coast-down result and the extent to which they do, provide many benefits in additions to the above mentioned. Today's tests are carried out until you believe the right result is achieved, but it is not possible to be certain. By knowing how the environment affects the result, the expeditions may be finished earlier since it is then possible to predict what result you can expect during the current circumstances. The coast-down expeditions can then be made more efficient in terms of both time and money. To fly the prototype cars to the U.S for a four week long test period is not only very expensive, it also keeps the prototypes away from other development departments that need to do other tests as well.

1.2. Purpose

The purpose is to find and understand the parameters that affect the coast-down result and predict the most accurate road load at the given circumstances, so that coast-down expeditions can be made with as few and effective test runs as possible and thereby make the expedition quicker, cheaper and with more precise and reliable result.

1.3. Aim and goal

The thesis shall result in a stable, mathematical model that describes the true road load within +/- 5% with 95% confidence.

1.4. Delimitations

How well a car performs in a coast-down test depends on two main areas: Partly car specific parameters such as aerodynamics, weight and power train drag and partly boundary conditions such as weather conditions and the character of the test track. Since the car specific parameters are relatively well known and can be considered constant during a test with the same car, only the boundary conditions impact on the coast-down result will be studied within this master's thesis.

1.4.1. Secrecy

Due to secrecy, some results, conclusions and discussions are removed, expressed as *"XXX"* or mentioned in general terms in this thesis. Also the number of significant figures varies and the real names of the dummy variables are hidden.

2. Coast-down and vehicle theory

This chapter covers the theory and regulations of coast-down tests and the vehicle dynamics that were used as a base for the models presented in Section 5.

2.1. Relation between road load and fuel consumption

The road load of a vehicle is defined as the force needed to push the vehicle forward in neutral gear in constant speeds on a flat road. VCC and many other car companies use a real world test, a coast-down test, to determine this force. The basic principle behind the coast-down test, illustrated in Fig. 2, is the following: accelerate the car to a predetermined speed, let it decelerate in Neutral Gear down to another predetermined speed and measure the time for the process. The road load is then calculated from Newton's second law using the vehicle mass and the difference in speed and time: (Hilmersson, 2010)

$$F = m \cdot \frac{\Delta v}{\Delta t} \tag{2.1}$$

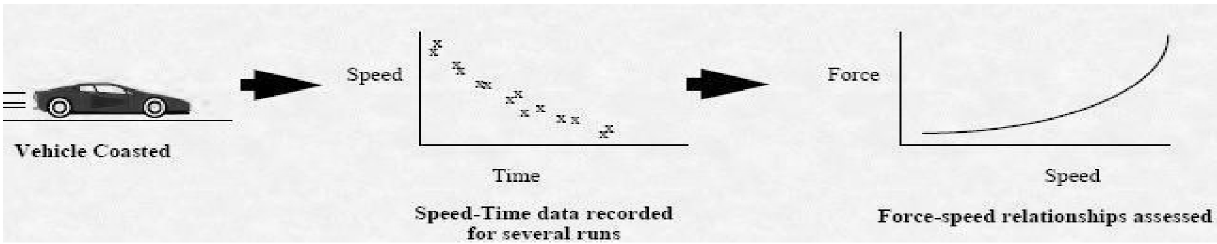

Fig. 2. *Schematic view over the coast-down test procedure (Hilmersson, 2010).*

By calculating a second order polynomial fit to the drag force as a function of vehicle speed, the vehicle specific coefficients f_0, f_1 and f_2 are produced:

$$F_{vehicle}(v) = f_0 + f_1 \cdot v + f_2 \cdot v^2 \tag{2.2}$$

The cars' certified emission level and fuel consumption is determined by performing predefined driving cycles on a roller test bench (Fig. 3). This bench is equipped with a dynamometer that simulates driving on a real road. The dynamometer load is acquired by running a coast down test on the roller test bench and calculates F_{dyno}:

$$F_{dyno}(v) = F_0 + F_1 \cdot v + F_2 \cdot v^2 \tag{2.3}$$

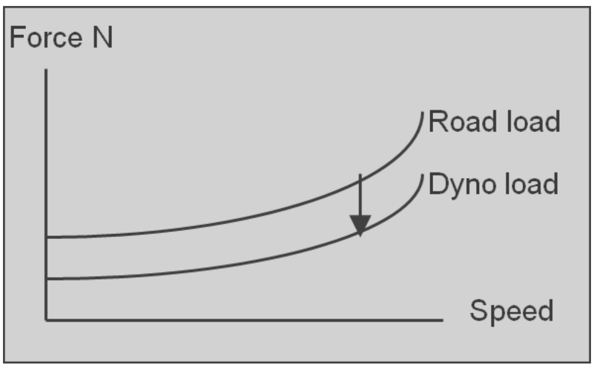

Fig. 3. *The difference in road load and dynamometer load. (Hilmersson, 2010)*

The coefficients in Eq. 2.3 are then adjusted so that F_{dyno} generates the same time-speed trace as the real world coast down test represented by f_0, f_1 and f_2 in Eq. 2.2.
Important is that F_{dyno} is not equal to $F_{vehicle}$. The dynamometer load is equal to all forces that are not acting on the car during the test, such as aerodynamic forces and difference in rolling resistance between the dynamometer and asphalt. Resistance that is already acting on the car at roller test bench are taken away to not have those forces twice. (Hilmersson, 2010)

2.2. Laws and regulations

Since the coast-down result is directly related to the car´s certified fuel consumption, those tests are governed by laws and regulations. The rules which are presented below, concerning how the tests shall be performed and under what environmental conditions the results are valid, are taken from Regulations No. 83-05. (2009).

2.2.1. Environmental conditions

- The road shall be level and the slope shall be constant within $\pm 0.1\%$ and not exceed 1.5%.
- The wind speeds shall be measured 0.7 m above the road surface and the wind speeds shall not exceed 3m/s in average and 5m/s in wind peak speeds. The vector component of the wind speed across the road shall be less than 2m/s.
- The road shall be dry
- The air density shall not deviate more than $\pm 7.5\%$ from the reference conditions P = 100 kPa and T = 293.2 K.

2.2.2. Test procedure

- The vehicle shall be accelerated up to a speed 10km/h higher than the chosen test speed v.
- The gearbox shall then be placed in Neutral
- The time taken, t_1, for the vehicle to decelerate from speed $v_2 = v + \Delta v$ to $v_1 = v - \Delta v$ shall be measured.
- The same procedure shall be performed again, but in the opposite direction.
- The average time T of the two test runs shall be calculated.
- The procedure must be repeated several times such that the statistical accuracy, p, of the average

$$T = \frac{1}{n} \cdot \sum_{i=1}^{n} T_i \qquad (2.4)$$

is no more than 2% ($p \leq 2\%$)

2.2.3. Correction formula

Since the temperature and the air density are considered to influence the outcome of the test, there is a correction factor defined in Regulations No. 83-05. (2009),

$$K = \frac{R_R}{R_T} \cdot (1 + K_R \cdot (t - t_0) + \frac{R_{Aero}}{R_T} \cdot \left(\frac{\rho_0}{\rho} \right) \qquad (2.5)$$

where R_R is the rolling resistance at speed v, R_{Aero} is the aerodynamic drag at speed v, R_T is total driving resistance (R_R + R_{Aero}), K_r is the temperature correction factor of rolling resistance, equal to $8.64\times10^{-3}/^{\circ}C$, t is the ambient temperature at the test, t_0 is the ambient reference temperature (20° C), ρ is the air density at the test and ρ_0 is the air density at the reference conditions (see Section 2.2.1). This correction factor K is multiplied by the power P determined on the track

$$P = \frac{m \cdot v \cdot \Delta v}{500 \cdot T} \tag{2.6}$$

where m is the vehicle reference mass, v is the speed of the test, Δv is the speed deviation from speed v (see section 2.2.1) and T is the time. The corrected power, P_{corr} is then calculated by

$$P_{corr} = K \cdot P \tag{2.7}$$

2.3. Vehicle dynamics

In Eq. 2.1 the road load F is defined by Newton's second law. This total force acting on a vehicle can also be described as the sum of the gravitational force and the drag force. The drag force can divided into following components: air resistance, side force resistance, rolling resistance and losses in transmission and bearings (Karlsson, Hammarström, Sörensen, & Eriksson, 2011). The road load can therefore also be defined as:

$$F = m \cdot \frac{\Delta v}{\Delta t} = F_{grav} + F_{air} + F_{side} + F_{RR} + F_{trm} \tag{2.8}$$

The force components in Eq. 2.8 are described below.

2.3.1. Gravitation force

The gravitational force is defined as:

$$F_{grav} = m \cdot g \cdot \sin(\theta) \tag{2.9}$$

where m is the vehicle mass, g the gravitational acceleration and θ is the longitudinal slope of the road (Karlsson et al, 2011). The gravitational acceleration is not constant but varies across the globe. Following formula can be used to determine the value of g:

$$g_0 = 9{,}7803267714 \cdot \left(\frac{1 + 0{,}00193185138639 \cdot \sin^2(\lambda)}{\sqrt{1 - 0{,}00669437999013 \cdot \sin^2(\lambda)}} \right) \tag{2.10}$$

where λ is the geographic latitude (Ahern, 2004).

2.3.2. Aerodynamic drag

Aerodynamic drag occur as when the airflow around and through the vehicle is being moved. When air flows over and past a solid form, vortices are created at the rear causing the flow to deviate from the smooth streamline flow. The air flow pressure in the front of the solid object will be higher than the surrounding pressure while the pressure behind will be lower, therefore the vehicle will be dragged in the direction of air movement. This effect is created in addition to the skin friction drag, which is the viscous resistance generated within the boundary layer when air flows over a solid surface (Heisler, 2002). The skin friction drag is not studied in this thesis. The air resistance can be written as

$$F_D = \frac{1}{2} \cdot \rho \cdot v_{rel}^2 \cdot C_D \cdot A \tag{2.11}$$

where ρ is the density of the air, v_{rel} is the relative air velocity striking the surface, C_D is the drag coefficient and A is the cross section area (Heisler, 2002). The air density is a function of total air pressure, temperature and relative humidity (Shelquist, 2011) that can be seen in Appendix A. The dimensionless drag coefficient depends upon the shape of the body exposed to the airstream (Heisler, 2002) and its value for different vehicles is determined in wind tunnels. Typical C_D-values for private cars are between 0.22-0.4 (Heisler, 2002). C_D is sensitive even for relatively small changes in the shape of the vehicle; a general rule is that one centimetre change of the car body height change the C_D-value with 0.01 steps and an open window or sunroof can increase C_D with five to seven percent (Bauer, 1996). As can be seen in Eq. 2.11 and Fig.4, the drag force is strongly depending on the relative air velocity.

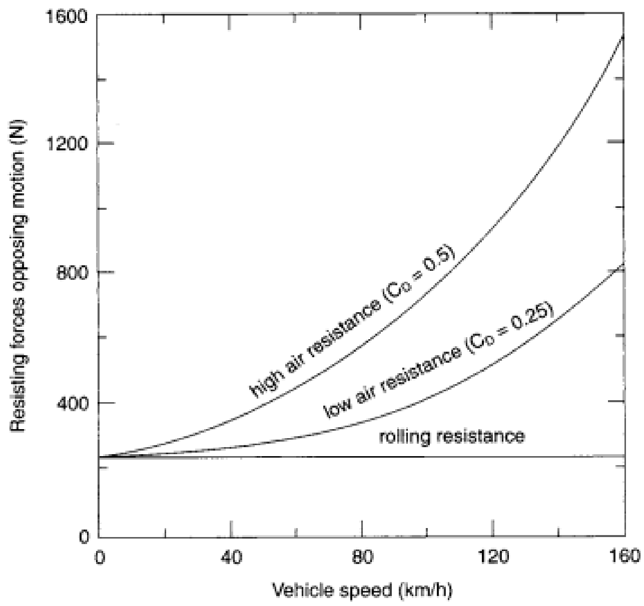

Fig. 4. *Comparison of aerodynamic drag forces with rolling resistance (Heisler, 2002).*

This air velocity is in favourable conditions equal to the vehicle speed as in Fig 4, but for real world test such as a coast-down test the meteorological wind affects the relative air velocity. The relative air velocity can be expressed as:

$$v_{rel} = \sqrt{v^2 + 2 \cdot v \cdot w \cdot \cos(\alpha) + w^2} \tag{2.12}$$

where v is the speed of the vehicle, w is the wind speed and α is the wind direction relative to the velocity vector of the vehicle. This v_{rel} strikes the vehicle with an angle β (Karlsson et al., 2011):

$$\beta = \arccos((v + \cos(\alpha) \cdot w) / v_{rel}) \tag{2.13}$$

The force F_D from Eq. 2.11 is then projected on an axis parallel with the vehicle's direction of travel, using the β-angle:

$$F_{air} = (C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot \cos(\beta) \cdot C_D \cdot A \cdot \rho \cdot v_{rel}^2 / 2 \tag{2.14}$$

The unknown coefficients C_L and C_{Lbeta} can be determined by regression.

2.3.3. Side force resistance

The side force resistance can be calculated as:

$$F_{side} = C_{side} \cdot (m \cdot (\cos(\sigma) \cdot v^2 / R - g \cdot \sin(\sigma) \cdot \cos(\theta)))^2 \tag{2.15}$$

where σ is the crossfall of the road, R is the radius of curvature of the road and C_{side} is a constant. (Karlsson et al., 2011)

2.3.4. Rolling resistance

Rolling resistance is the force acting on a vehicle caused by the interaction between the vehicle and the road surface (Karlsson et al., 2011). It is the tire deformation that causes the rolling resistance and as illustrated in Fig.5 the resultant normal force is shifted forward causing a torque around the wheel centre that resists rolling. The horizontal force that is required to keep the wheel at constant speed and thus overcome this torque is called the rolling resistance and is defined as (Nielsen & Sandberg, 2002):

$$F_{RR} = C_{RR} \cdot N \tag{2.16}$$

where N is the normal force and C_{RR} is the rolling resistance coefficient. C_{RR} is depends on many variables such as inflation pressure, tire temperature, vehicle speed, road conditions and wheel adjustments (Nielsen & Sandberg, 2002).

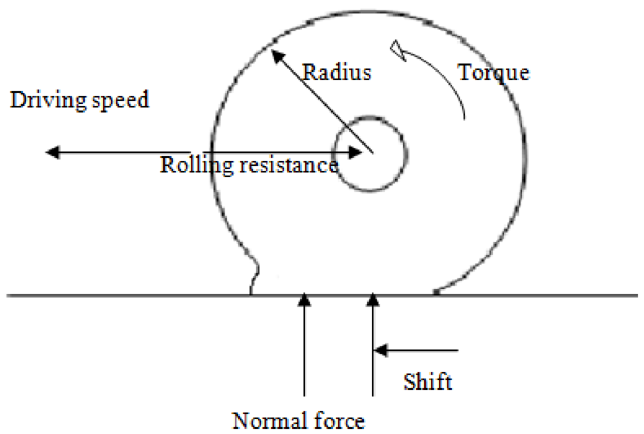

Fig. 5. *Illustrates the horizontal shift of the resultant normal force during tire deformation (Nielsen & Sandberg, 2002).*

In the report by Karlsson, et al. (2011) the following equation for the rolling resistance is suggested

$$F_{RR} = N \cdot (C_{RR\_00} + C_{RR\_temp} \cdot T + C_{RR\_MPD} \cdot MPD + Cr_{RR\_IRI} \cdot IRI + Cr_{RR\_IRI\_v} \cdot v) \quad (2.17)$$

where T is the ambient temperature, MPD is a measure for the macrotexture of the road, IRI is a measure of the unevenness of the road, v is the vehicle speed and C_{RR_00}, C_{RR_temp} etc are constants that can be determined by multiple regression. Another model suggested by Nielsen & Sandberg (2002) is

$$F_{RR} = C_r(T, v) \cdot N \quad (2.18)$$

where T in this model is the tire temperature. This model with a rolling resistance coefficient depending only on the tire temperature and the vehicle speed require a driving scenario with a given tire and constant or slow varying external conditions. It is established in the report that the tire temperature is the dominant parameter for the above described driving scenario and that the tire temperature is dependent on the ambient temperature, road surface temperature, the vehicle speed and how long time the vehicle is driven at various speeds. The main reason for the strong correlation between tire temperature and rolling resistance is that the tire pressure raises when the inflated air gets warmer and thereby reducing the rolling resistance. Even if the tire pressure is held constant, the rolling resistance is known to depend also on the velocity. The effect of the vehicle speed is however rather small, around 20-30% of the tire pressure effect. (Nielsen & Sandberg, 2002)

2.3.5. Transmission drag

Due to the limitations of this thesis, the transmission resistance and losses in bearings are not discussed since they are considered to not vary substantially for different boundary conditions. The contribution to the road load that the transmission losses constitute is instead covered by dummy variables, which is explained in Section 3.7.

2.4. Weather station

The weather station measures the properties of the ambient air, which are: pressure, humidity, temperature, wind speed and wind direction. Those properties are measured each 10 second, which means that the weather-data obtained from the coast-down tests does not need to be exactly what occurred during the test. The weather station is located at the side of the test track and hence, it could be some differences between what is measured and what is affecting the car at the test track.

3. Regression theory

Regression analysis is a statistical tool for the investigation of relationships between two or more variables that are functionally linked (Petterson, 2003), (Grandin, 2012). In this chapter a review of the regression methods and statistical tests used in this thesis, are described.

3.1. Simple linear regression model

The basic principle in linear regression is that a straight line, a regression line, is adjusted to a statistical material consisting of *n* pairs of observations (x_i, y_i) using the least square method. This means determine the values of the variables *a* and *b* in the equation of a straight line

$$y = a + b \cdot x \tag{3.1}$$

so that the sum of squares

$$\sum_{i=1}^{n} (y_i - a - b \cdot x_i)^2 \tag{3.2}$$

is as small as possible. The solution to this can be written as

$$a = \bar{y} - b \cdot \bar{x} \tag{3.3}$$

$$b = \frac{\sum y_i \cdot x_i - \frac{\sum \bar{y}_i \sum x_i}{n}}{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}} \tag{3.4}$$

where *a* indicates where the straight line intersect with the y-axis and *b* indicates the average change in *y* for one unit change in *x*. (Petterson, 2003)
It is very rare that this minimization problem has an exact solution, especially when the data are from real world measurements or experiments (Umeå University, 2004). The deviation from the dependent variable's conditional expectation for the level *x* is called the residual, ε. The variation is described with this model:

$$y = a + b \cdot x + \varepsilon \tag{3.5}$$

The residual is a random variable with expectation zero and the value of the residual can be interpreted as the total effect of other factors that influence the dependent variable *y* (Wahlgren & Körner, 2006). More about residuals and how to analyse those can be found in Section 3.6.

3.2. P-value

The p-value is the probability to obtain at least as large difference as between the sample value and the value under the null hypothesis. For regression, the null hypothesis is that there is no linear relationship between the dependent variable and explanatory variables (Körner & Wahlgren, 2006):

$$H_0 : \beta_1 = 0 \tag{3.6}$$

The alternative hypothesis is that each of the regression coefficients is different from zero:

$$H_1 : \beta_1 \neq 0 \tag{3.7}$$

The p-value for which the null hypothesis is rejected is determined by the level of significance. A common value for the level of significance is *α = 5 %*, which means that if the p-value is larger than 0.05 the null hypothesis cannot be rejected and the smaller the p-value, the greater support for the null hypothesis. (Körner & Wahlgren, 2006)

3.3. Analysis of Variance

With the above formulas it is possible to obtain an equation for the linear relationship between the dependent variable *y* and the independent variable *x* (Grandin, 2012). To analyse the strength of the relationship, i.e. how much of the variation in *y* is due to the change in *x*, the ANOVA (*An*alysis *O*f *Va*riance) table can be used. In Table1 an example of an ANOVA table from MS Excel can be seen.

Table 1. *ANOVA table from MS Excel*

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	15	7736627	7736627	20448,36	2,2167E-199
Residual	195	73778,17	378,3496		
Total	196	7810406			

The cell at the intersection of the second column and third row in Table 1 is called Total Sum of Squares (SST); this is the total variation of *Y* and it is defined as

$$SST = \sum (y_i - \bar{y}_i)^2 \tag{3.8}$$

In Table 1 it is also possible to see the Sum of Squares for the residuals (SSE), which is the variation around the adapted regression line that is independent of the variation of *x*. This unexplained variation is calculated as

$$SSE = \sum (y_i - \hat{y}_i)^2 \tag{3.9}$$

where

$$\hat{y}_i = a + b \cdot x_i \tag{3.10}$$

Eq. 3.9 is in other words the squared and summed difference between the actual value y_i and the expected value $\hat{y}_i$ calculated from the regression model at x_i . The difference between the total variation, *SST*, and the unexplained variation, *SSE*, is called *SSR* and is the explained variation (See Table 1 first row, second column); the variation that is described by the regression model. (Körner & Wahlgren, 2006). To sum it up, the total variation can be written as

$$SST = SSR + SSE \tag{3.11}$$

and it is also illustrated below in Fig.6

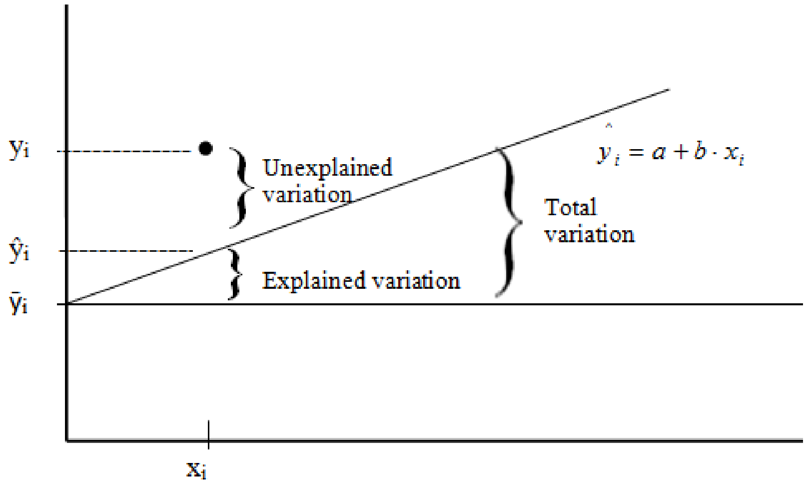

Fig. 6. *How the division of the total variation, SST, can be seen. (Illustration done after model by Karlsson et al., (2011)*

3.3.1. Coefficient of determination

As a measure of how strong the linear correlation is, the coefficient of determination, R^2, is used

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{3.12}$$

The R^2-value tells how much of the total variation of the dependent variable is explained by the linear relationship between the variables (Körner & Wahlgren, 2006). When using more than one explanatory variable in the model, the R^2-value can be misleading since it increases with the number of terms whether the terms are significant or not (Grandin, 2012). Therefore a R^2-value that is adjusted based on the residuals degrees of freedom must be used in multiple regression. The residual degrees of freedom is defined as

$$v = n - m \tag{3.13}$$

where *n* is the number of response values and *m* is the number of fitted coefficients estimated from the response values. The adjusted R^2-value is defined as (UNSW, 2011)

$$adj - R^2 = 1 - \frac{SSE(n-1)}{SST(v)} \tag{3.14}$$

3.3.2. F-value

The F-value can be found in the ANOVA table (Table1) and is a test function for the null hypothesis that there is no linear relationship between the variables. F is calculated as

$$F = \frac{MSR}{MSE} \tag{3.15}$$

where *MSR* is the mean sum of squares for the regression and *MSE* is the mean sum of squares for the residuals. The value of *MSR* and *MSE* can be found under *MS* in the ANOVA table. The F-distribution is skewed to the right and the critical area, illustrated in Fig.7, in the right tail depends on the degrees of freedom for *MSR* and *MSE* and the level of significance. The probability that F is in the accepted area and the null hypothesis is accepted can be seen in the ANOVA table under *Significance F*. If a significance level of 5% is selected, the *Significance F* must be lower than 0.05 if the null hypothesis should be rejected. (Körner & Wahlgren, 2006)

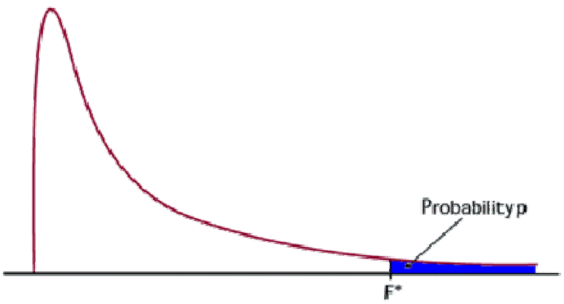

Fig. 7. *The shape of the F-distribution with the critical area to the right.*

3.4. Multiple regression analysis

The multiple regression is a natural extension of the simple linear model described in Section 3.1. The main difference is that a multiple regression model contains two or more explanatory variables in order to better explain the variation of *y*. When using two x-variables, the model can look like this:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 \tag{3.16}$$

The interpretation of b_1 and b_2 is now as follows:

b_1 = The average change in the variable *y* if x_1 increases one unit and x_2 remain unchanged.
b_2 = The average change in the variable y if x_2 increases one unit and x_1 remain unchanged.

The value for those variables is determined in the same way as for the simple regression in Section 3.1 (Petterson, 2003).

Building a suitable multiple regression model can be a balancing act between using as many interesting and essential variables as possible and the risk for overdetermination. For a small amount of data with many explanatory variables there is a risk that the regression model indeed provides a good description of the variation in *y*, but that only applies to that particular sample and is not valid for the entire population.Therefore, the number of variables in the model should be carefully considered. Only the most essential variables that together provides as high coefficient of determination, R^2 (see Section 3.3.1), as possible should be used. (Körner & Wahlgren, 2006)

3.5. Multicollinearity

Linear relations between the explanatory variable is called multicollinearity (Nationalencyklopedin, 2012). A risk when using many variables in a multiple regression model is that some of them might be strongly correlated and thus provides almost the same information (Grandin, 2012).The evaluation of the regression coefficients and its standard errors is not independent and at multicollinearity the standard errors therefore becomes larger in the valuations. It is thus possible to obtain an inferior model with two explanatory variables instead for only one, even though the coefficient of determination slightly increases with two variables. It is then better to only use one of the correlated variables, preferably that one that provides the highest R^2-value (Wahlgren & Körner, 2006). To avoid strongly correlated variables in the model, a correlation analysis with all the variables should be done before performing the regression analysis. There is no strict value when the regression cannot be made, but as a rule of thumb variables with correlation stronger than 0.8 should not be used (Sundell, 2010). Since multicollinearity between two or more variables often indicates that they provide about the same information to the model, one way to get around the problem is to drop one of the variables from the model and keep that one that provides the highest value of R^2 (Körner & Wahlgren, 2006). If all the correlated variables are considered important to the model and cannot be taken away, another way to solve multicollinearity is to combine the correlated variables to a single variable with which the regression is performed (UKY, 2010).

3.5.1. Variance Inflation Factor

Another way of detecting multicollinearity is to use the Variance Inflation Factor (VIF).
The VIF quantifies how much the variance of the estimated coefficients is inflated and thus the severity of the multicollinearity (Simon, 2004). As for the correlation between variables there is no strict VIF-value when the regression result is considered unreliable but as a general rule of thumb VIF higher than 4 warrant further investigations and over 10 is a sign of severe multicollinearity that must be solved (Simon, 2004). The Variance Inflation Factor is defined as:

$$VIF_j = \frac{1}{1-R_j^2} = \frac{S_{x,j}^2 (n-1) SE_{b,j}^2}{S^2} \tag{3.17}$$

where $S_{x,j}$ is the standard deviation for the regression variable, $SE_{b,j}$ is the standard error for the slope coefficient, S^2 is the mean squared residual and (n-1) is the degree of freedom (ProfTDub, 2010).

3.6. Residual analysis

The residual, ε, is defined as the difference between the actual observed value y and the corresponding value $\hat{y}$ obtained from the regression model:

$$\varepsilon = y - \hat{y} \tag{3.18}$$

There are three prerequisites regarding the residuals so that the equations used in the regression should give accurate values (Körner & Wahlgren, 2006):

- The residuals must be distributed independently

- The residuals' standard deviation is equal for all levels
- The residuals are normally distributed for all levels

By using plots of the residuals, those prerequisites can be controlled quite easily. Fig. 8 shows some examples of this.

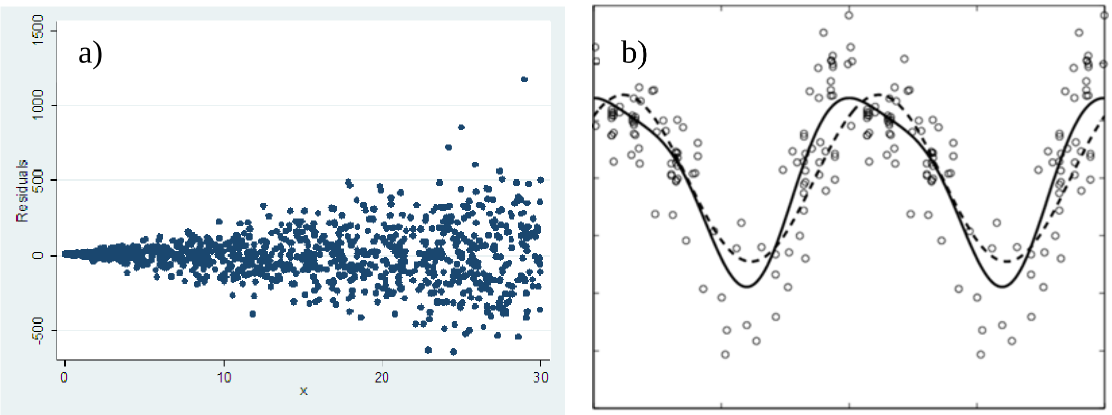

Fig. 8. *a) residuals with standard deviation that is not equal for all levels of x (UCLA, 2007). b) residuals with a systematic pattern, they are not independently distributed (Zejda, 2008).*

The third prerequisite about normal distribution can be illustrated with a histogram. However, this assumption not necessary for large data sets (Körner & Wahlgren, 2006), since the rules for the Central Limit Theorem shows that for a data set larger than 20 observations, one can assume approximately normal distribution (Grandin, 2012).

A lot of information can be obtained by analyzing the residuals, such as the suitability of the model, the data set as a whole and as individual observations. The reason for large residuals is often measuring fault, incorrect regression model or the actual individual variation. Large deviations should be seen as warning signals that something is not right (Körner & Wahlgren, 2006). If the outliers should be removed from the data set must be carefully analyzed so that the real cause is bad data and not part of the natural variation or an incorrect regression model.

If the residuals look like those in Fig. 8, that is a clear sign of an incorrect regression model. To solve the problem, the data may be transformed in a manner such that linear regression can be used. If that is not possible, other types of regression models, for example polynomial regression or non-linear regression, must be used. (Grandin, 2012) In this thesis only linear regression will be discussed.

3.7. Dummy variables

The use of dummy variables is a way to treat qualitative variables in a regression model, for example if there is a difference in road load using a manual or automatic gearbox. The regression model is made to handle quantitative variables but can manage qualitative variables if they are transformed into binary dummy variables with values 0 or 1. (Körner & Wahlgren, 2006) The estimated regression relations can for example now be written as:

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 \tag{3.19}$$

where

x_1 = an ordinary quantitative variable
x_2 = gearbox (with values “manual” = 0 and “automatic”=1)

The value of the regression coefficient b_2 is equal to the difference in $\hat{y}$ when shifting between manual and automatic gearbox. It is possible to have more dummy variables than just one, but it is then necessary to have a reference variable that the other variables can be compared with. Each of the dummy variables uses one degree of freedom, so *n* groups has *n-1* degrees of freedom. What group or variable that is used as reference does not influence the R^2-value (UCLA, 2007).

4. Methodology

This section covers the setup of the thesis and what methods that are used in order to achieve a result that fulfils the aim and purpose.

4.1. Literature study

In order to achieve an understanding for the task and to increase the knowledge in areas related to vehicle dynamics, coast-down tests, statistics and data mining, a literature study was done. The search for literature was primarily done using the search engine at Chalmers Library's homepage, in order to obtain reliable information and get access to articles and reports that otherwise are unavailable. Information about coast-down tests in terms of implementation, measuring methods, accuracy and laws and regulations was captured both by interviews with experts at the Fuel Economy department at VCC and internal information. Also interviews with statistics experts at Chalmers were made in order to verify and discuss methods and results.

4.2. Data collection and preparation

Coast-down expeditions must be performed for every new car model and its variants; the amount of data available for analysis is thus very large. There is however some important information that is not officially documented and only available in terms of personal notes and what is remembered of the team that performed the test. Since such information becomes weaker over time and no one in the team that carried out tests earlier than 2010 are left in today's coast-down team, only data from tests performed during 2010 and later are used in order to compromise between minimize the risk for misinterpret the results and use as much data as possible.

All data used in this thesis is derived from measurements made for internal development at VCC and is not required by the authorities. The documentation of these measurements is therefore not as thoroughly as for the official data.

During coast-down expeditions not only data about the cars, speeds and times are gathered, but also information about weather and road conditions. The information about weather, coast-down times and car specific data were merged into one large file and then matched against the current time for the test. According to the regulations of the coast-down tests (Section 2.2) the coast-down must be performed in both directions of the test track and the average of the two runs is the time that counts. Those two runs were treated as individual runs in order to get as many runs as possible to analyze. Measurements of the road surface roughness were not included in this file, both because they were considered not sufficient and accurate enough, and for the lack of information about the where on the test track the current coast-down tests were made, which in turn made it unfeasible to match the asphalt measurements against the coast-down times.

In order to prepare the dataset for further analysis, obvious errors such as duplicates and tests without weather measurements were removed. Arizona, where APG is located, does not use day light saving time (Prerau, 2006), but the measurement equipment from Sweden does. Therefore, data from the day when the time shift occurred were analysed and cleared from results being registered twice with an hour difference.

4.3. Data analysis and model structure

To be able to describe the variation of the forces acting on a car during a coast-down test, some kind of mathematical model is needed. From the literature study, formulas describing elementary correlations between environmental parameters and forces acting on a car were found and assembled into a first model. The descriptions of these formulas are presented in Section 2.3. With this model a first multiple regression on the dataset was done in order to be able to analyse the quality of both coast-down data and regression model. Naturally, it is important that the input data is correct and contains only natural variation and no significant measurement errors or other sporadic human error that cannot be predicted in a model. Therefore a residual analysis was done after the multiple regression in order to detect such inaccuracies in the coast-down data. The residual is defined as the difference between the actual observed value y and the corresponding value ŷ obtained from the regression model (Körner & Wahlgren, 2006) and it is consequently a good measure of whether something is not right, either the regression model or the input data. The residuals of the variables in the regression model were illustrated in diagrams to visualize possible outliers. Those residuals that were considered as outliers were thoroughly analysed by tracking the coast-down results which led to the error and go through the current testing protocol together with the group in charge of the expedition. Those tests that were considered invalid for some reason were removed from the dataset and for those where no errors were detected, the test results were retained in order not to affect the natural variation. By this method both large groups of invalid coast-down results and single measurement errors can be found much quicker and more accurate than if all test protocols had been examined one by one without knowing what kind of error that is being sought. The residual analysis was used also as a quality control of the regression model by investigate if the prerequisites described in Section 3.6 were met. More about residual analysis and regression models can be found in Section 3.1 and 3.6, respectively.

When the dataset is cleared from invalid coast-down results, the regression model can be extended with more interesting parameters. Since this thesis aims to determine the external parameters' impact on the test result, the car specific parameters were treated with dummy-variables. In that way the influence of for example different gearboxes and FWD or AWD were kept away from the more important variables. Due to the regulations, the coast-down test must be performed in both direction of the test track; therefore also a dummy variable for the direction of the test track was used in order to determine a possible difference in road load between the various directions. A description of dummy-variables is presented in Section 3.7.

How well the model describes the variation of forces in the dataset was measured by the adjusted coefficient of determination, $adj\text{-}R^2$, which indicates the proportion explained variation of the total variance, with respect to the number of parameters in the model. The definition of the ordinary coefficient of determination, R^2, the adjusted coefficient of determination and difference between them, is discussed in Section 3.3.1.

To prove that there is a significant linear correlation between the variables and hence reject the null hypothesis that there is no linear relationship between them, an F-test was made for the regression model as a whole. The significance level was set to 5%, in accordance with the aim of the thesis. Also a t-test for each of the regressions coefficients was made in order to try the null hypothesis that the regression coefficient is equal to zero. That is to say that the variable has no effect on the regression equation. P-values lower than 0.05 were accepted for the t-test due to the above mentioned significance level at 5%. More about the F-test and p-values can found in Section 3.2 and 3.3.2.

The use of many variables in a regression model increases the risk of multicollinearity, which provides unstable estimates of the regression coefficients. Hence must multicollinearity be detected and adjusted before any conclusions can be drawn from the model. Both a VIF-test and a correlation analysis were done in order to be sure of detecting variables that could make the model unstable. If multicollinearity is detected, the variable that provides the least contribution to the coefficient of determination is removed from the model. In that way a model with better and more reliable estimates of the regression coefficients is obtained, but at the expense of being able to take account of many different variables. The explanation for multicollinearity, VIF-test and correlation analysis can be found in Section 3.5.

To further determine the stability of the model and its variables, regression analyses with the final model were made on the dataset for each of the three coast-down expeditions separately. If there is a good agreement between the regression coefficients for the analyses, then the model can be considered as robust. If the coefficients strongly fluctuate between the expeditions, they must be analyzed deeper in order to understand the deviation.

4.3. Statistical software

All coast-down data were compiled using MS Excel, which is a program that is well suited for processing large data volumes, is easy to use and requires very short learning time. Also the statistical operations on the dataset were made using MS Excel. Due to secrecy concerning the coast-down data and the calculations on this, only computers on VCC were permitted to be used and they contained no more advanced statistical software, such as SPSS or MiniTab. SPSS, Minitab or some other strict statistical program provides more built-in statistical functions and plots that could have speeded up the calculations, since these functions, VIF-tests for example, must be done "by hand" in MS Excel.

5. Results and analyses

This chapter covers the results obtained from the method in the previous section. The results are continuously analysed and at the end of this section a thoroughly analysis and discussion of the final regression model is presented.

5.1. Dataset

Information about the cars, coast-down times and weather were merged into a single Excel file. With all duplicates, weather errors and other obvious errors removed, 22570 unique coast-down results were obtained. The tables below show the headlines of the input data and fictional examples of those.

The weight in Table 2 is the total weight of the car, including driver and liquids. The total weight was used since it is important to use the actual weight of each test in order to understand the variance of the road load. During a test, the total weight will slightly decrease due to the fuel consumption. The effect of this was considered small and was not taken into account in the calculations.

Table 2. *Car related information*

Test nr.	Car model	Transmission	AWD/FWD	Cd*A	Weight [kg]
31	D1	B6	FWD	0,8	1500

Table 3. *Coast-down information*

Coast-down time [s]	Speed [m/s]	Date	Direction (FWD/BWD)
14,2	30,55	2011-10-27 19:17:45	FWD

Table 4. *Weather information*

Wind speed [m/s]	Wind direction [rad]	Humidity [%]	Air pressure [kPa]	Ambient temperature [˚C]
4,6	2,78	17,7	96,1	27,2

5.2 Analysis of the dataset

5.2.1. First regression model

In order to analyse the data set and find errors that are hard to find just by watching it, a first basic model for the forces acting on a car during a coast-down test was constructed using Eq.2.8 as a base

$$F = m \cdot \frac{\Delta v}{\Delta t} = F_{grav} + F_{air} + F_{side} + F_{RR} + F_{trm}$$

where F_{grav} and F_{side}, Eq. 2.9 respective Eq. 2.15, were neglected since the test road at APG was considered very level. Also F_{trm} was ignored, both since that is a car specific variable which is not discussed in this thesis, and because the values of the regression coefficients are not important at this stage. For the air resistance, F_{air}, Eq. 2.14 was used. Eq.2.17 was used to calculate F_{RR}, but the variables for the road surface condition, *MPD* and *IRI*, were excluded since those variables are not sufficiently measured at the coast-down expeditions. The actual road load, *F*, was calculated using Eq. 2.1 and the information in Table 2 and 3:

$$F = m \cdot \frac{\Delta v}{\Delta t} = 1500 \cdot \frac{10/3,6}{9}$$

The values for the mass and coast-down time are just examples. The speed difference, Δv, is always 10 km/h. It was converted into the SI-unit m/s by dividing by 3.6.

Thus, the first regression model looked like this:

$$\underbrace{m \cdot \frac{\Delta v}{\Delta t}}_{F} = \underbrace{N \cdot (C_{temp} \cdot T + C_{RR\_00})}_{F_{RR}} + \underbrace{(C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot AIR}_{F_{AIR}} \qquad (5.1)$$

where

$$AIR = \cos(\beta) \cdot C_D \cdot A \cdot \rho \cdot v_{rel}^2 / 2 \qquad (5.2)$$

A multiple regression analysis with the above model used on the unadjusted dataset, provided a value of *adj-*R^2 at 84.32%. The values of the regression coefficients are at this stage of no interest, since the accuracy of the input data is somewhat dubious.

5.2.2. Residual analysis and clearance of bad data

The associated residual plots for the regression variables were analysed with the aim of finding outliers. Only outliers which could be traced to some sort of error during the test were removed or adjusted, the others were retained in order not to affect the natural variation. In Fig.9 there is a well-defined group of outliers indicating some kind of error in the weight data.

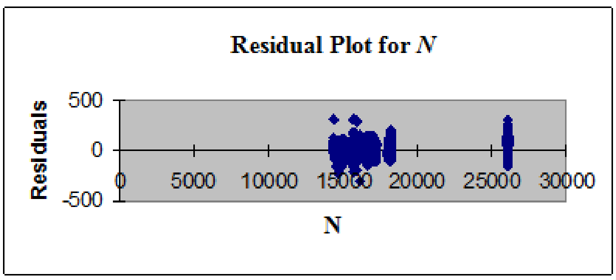

Fig. 9. *Residual plot for N with a distinct group of outliers at N ≈ 26000.*

By tracing those outliers back to the source it was discovered that a group of cars had a total weight of 900 kg more than other similar cars, caused by a human error. For the air resistance variables, *AIR* and $AIR \cdot \sin(\beta)$, both groups and seemingly isolated outliers were found. That is illustrated in Fig.10. Tracing those outliers back the current coast-down expedition protocols resulted in findings of driver mistakes, invalid practice runs, problems with the surveying equipment and other errors that made the coast-down result invalid. Such results were deleted from the dataset. Some of the outliers in Fig. 10 were due to problems that could be adjusted, such as a coast-down test performed in the opposite direction order and a number of tests performed with the wind direction indicator turned 180 degrees. Both of those errors led to that the cars were believed to have head wind when they actually had tailwind, and vice versa. The adjusted dataset used in the further calculations now contains 21300 unique coast-down results.

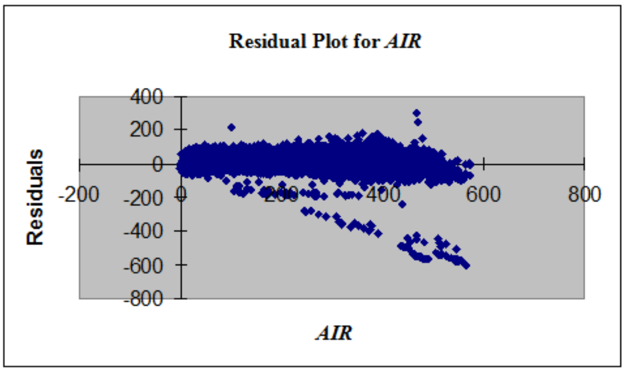

Fig. 10. *Residual plot for the air resistance with both groups of outliers and single extreme values.*

5.2.3. Adjusted dataset

With a new multiple regression analysis performed with the same model as before, Eq. 5.1, but on the new dataset, the proportion explained variation, *ajd-*R^2, was now 95.15%. Thus a large increase was obtained only by clearing the input data from erroneous values. The residual plots for the independent variables can be seen in figures below. With the outliers corrected, the residual plots were used also as a check for the quality of the model with the criteria specified in Sec.3.6. The residuals for the variables N and $T \cdot N$ in Fig.11a respective Fig 11b, meet the criteria; they are independently distributed and the standard deviation is equal for all levels. The groupings in Fig.11 a are due to the difference in weights for the various car models and the variation around each weight group is the difference in driver weight and filling degree of the liquids. In both Fig.11c and Fig.11d small amount of negative forces can be seen which occurs at low vehicle speeds and high meteorological tailwind speeds. The standard deviation seems to decrease for increasing values of $AIR \cdot \sin(\beta)$ in fig.11d), but the number of observations does also strongly decrease for increased values of $AIR \cdot \sin(\beta)$ and it is therefore not possible to determine if the standard deviation actually is equal for all levels or not. The residuals in Fig 11c) indicate a departure of linearity for high vehicle speeds of 110 km/h and more. The reason for this behaviour is further discussed in Section 5.5.3.

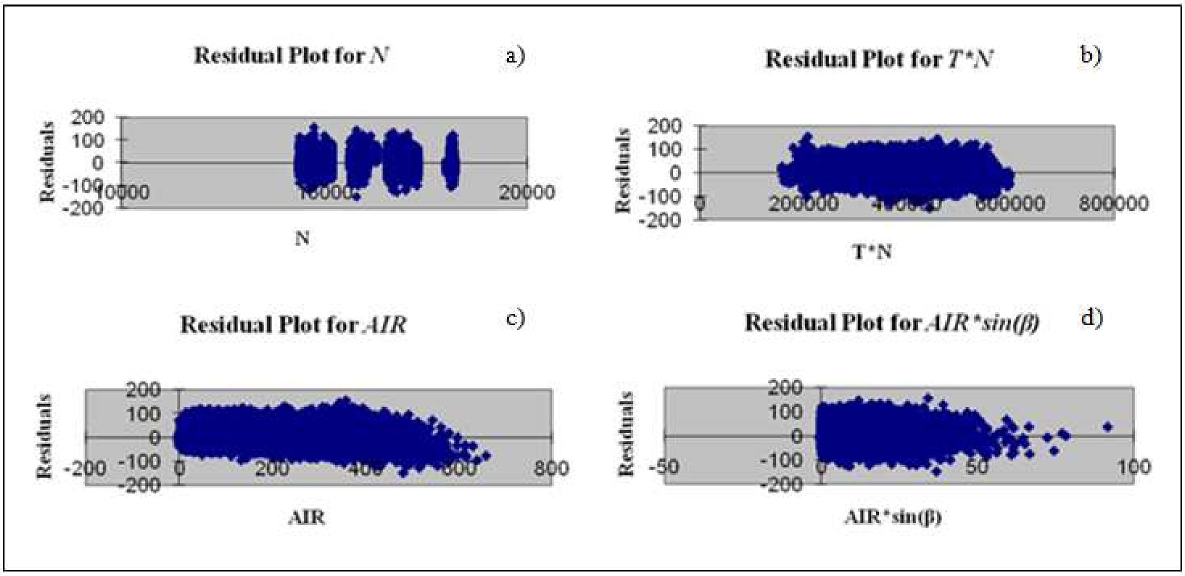

Fig. 11. *Residual plots for the variables in the regression model*

The ANOVA-table and the values of the different regression coefficients and their *p-values* are reported in the table below

Table 5. *ANOVA table and statistics for the regression coefficients.*

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	4,56E+08	1,14E+08	104542,9	0
Residual	21295	23224892	1090,627		
Total	21299	4,79E+08			

	Coef	*SE Coef*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	38,0878	3,2598	11,6840	1,92E-31	31,6983	44,4773
CRR_00	0,0083	0,0002	37,5275	1,6E-298	0,0079	0,0088
C_temp	XXX	0,0000	-37,1285	2E-292	XXX	XXX
CL	1,0219	0,0018	575,6082	0	1,0184	1,0254
CL_beta	0,1669	0,0309	5,3986	6,79E-08	0,1063	0,2275

The *Significance F* in the ANOVA table is equal to 0 and the null hypothesis, that there is no linear relationship between the variables, was therefore rejected. The model as a whole can therefore be considered as significant. The *p-values* of the t-tests for all regression coefficients are also much lower than the level of significance at 0.05 and the alternative hypothesis, that the coefficient is different from zero, was accepted for all coefficients. That is, it is very likely that all the coefficients have a linear relationship to the dependent variable *y*. In this thesis *y* is equal to the road load *F*.

5.3 Dummy variables

Dummy variables were created for each variant of the cars in the data set. There are eight different cars and some of them were tested with different gearboxes and number of driving wheels. There are consequently eleven different variants and hence also eleven car-related dummy variables used in the regression model. The eleventh dummy variable, *D11*, was used as reference and is therefore not included in the table of coefficients.

A twelfth dummy was entered into the model to determine if there are any differences in which direction of the test track the coast-down test is performed. Forward was set to 1 and Backward as 0. The dummy variable was denoted *Direc*.

The regression analysis with the model including dummy variables provides a slightly higher *adj-*R^2-value than the model without them. The adjusted coefficient of determination was now calculated to 96.11 %. As illustrated in Table 6, there is a substantial change in the *p-value* for the *Intercept* and C_{RR_00}, compared to the values in Table 5 without dummy variables. But also the estimation of the coefficients differ in an unreasonably way. Since the weight of the cars to a great extent depends on the car model and the variants of those, there is a high risk for multicollinearity between the normal force variable *N* and the dummy variables. The high VIF-values, which are illustrated to the right in Table 6 and the correlation analysis (See Table B1 in Appendix B) for C_{RR_00} and the car-related dummy variables confirms that multicollinearity actually is present and that is has to be solved in order to obtain a regression model with reliable estimates of the coefficients. The regression coefficient C_{RR_00} was removed from the model to solve the problem. With that removed, the *adj-*R^2-value remains at 96.11 % and the VIF-values were well below 3 for all coefficients. All coefficients, except for the dummy variable *D10*, were also clearly significant (p-value<<0.05). The table for the regression analysis performed with the model without C_{RR_00} can be found in Table C1 in Appendix C.

Table 6. *ANOVA table and statistics for the regression coefficients with dummy variables under the dotted line.*

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	15	4,61E+08	30711384	35099,35	0
Residual	21284	18623168	874,9844		
Total	21299	4,79E+08			

	Coef	*SE Coef*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*VIF*
Intercept	61,6722	23,2528	2,6522	0,008002	16,0949	107,2495	
CRR_00	0,0061	0,0015	3,9422	8,10E-05	0,0031	0,0091	73,60
C_temp	XXX	0,0000	-41,8557	0	XXX	XXX	1,30
CL	1,0160	0,0016	630,5791	0	1,0128	1,0191	1,26
CL_beta	0,2686	0,0286	9,3803	7,22E-21	0,2124	0,3247	1,29
Direc	12,5905	0,4056	31,0436	5,65E-207	11,7956	13,3855	1,00
D1	50,6130	2,1566	23,4692	2,83E-120	46,3860	54,8401	2,68
D2	34,4623	1,5893	21,6841	3,79E-103	31,3472	37,5774	4,06
D3	-9,5033	1,2823	-7,4109	1,30E-13	-12,0168	-6,9898	3,22
D4	-5,6998	1,0711	-5,3217	1,04E-07	-7,7992	-3,6005	2,13
D5	24,6415	1,5058	16,3644	7,97E-60	21,6901	27,5930	5,45
D6	4,6442	2,6751	1,7361	0,082559	-0,5991	9,8876	22,67
D7	11,8388	3,0638	3,8641	0,000112	5,8335	17,8441	23,75
D8	5,4670	4,8344	1,1309	0,258129	-4,0088	14,9427	48,14
D9	0,5504	3,4423	0,1599	0,872972	-6,1969	7,2976	23,72
D10	2,2175	1,1457	1,9354	0,052951	-0,0282	4,4632	2,52

Thus, removing C_{RR_00} can be considered as a good way to solve the multicollinearity problem without reducing the *adj*-R^2-value. The rolling resistance coefficient is consequently now only a function of the ambient temperature.

5.3. Extended regression model

In order to further increase the *adj*-R^2-value and to get a more comprehensive model, more variables were added.

The rolling resistance for the tires used in the coast-down tests is measured at VCC before each coast-down expedition. Those values were entered in the regression model as a part of the rolling resistance coefficient. The variable is called T_{RR}.

According to Section 2.3.4 and Eq. 2.18, the rolling resistance mainly depends on the tire temperature, provided that variables such as road surface, wheel adjustments and tire type are relatively constant. For coast-down tests, the wheel adjustments are carefully controlled and can be considered as constant. The variation in tire types between the coast-down expeditions are rather small and were in addition handled with the variable T_{RR} described above. The influence from the road surface structure can be described with Eq. 2.17, but such surface texture measurements have only been done in a small scale and also with no possibility to connect those results with the other coast-down data. Based on the few measurements that have been made, it can be seen that there appears to be some variations in the asphalt although it is probably small. Assuming that the variations in road surface are small, the tire temperature is the main contribution to the rolling resistance coefficient.
In order to further describe the rolling resistance as function of temperature, a variable called *Time* was entered in the regression model. This variable represents the time of day (1-24) at which the test was performed.
After addition of dummy variables, variables for tire rolling resistance and for time of day and the deletion of the C_{RR_00} coefficient, the model looks like this:

$$F = \underbrace{N \cdot (C_{temp} \cdot T + C_{time} \cdot Time + C_{T\_RR} \cdot T_{RR})}_{F_{RR}} + \underbrace{(C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot AIR}_{F_{AIR}} + \underbrace{Direc + D1 + ... + D11}_{Dummies} \quad (5.3)$$

A regression analysis with the model specified in Eq 5.3 generated the values illustrated in Table 7. The *adj*-R^2-value was calculated to 96.16%.

According to the extremely high VIF-values in Table 7 for C_{T_RR} and the dummy variables, the multicollinearity is very strong between those variables. This is partly because the car weight is correlated to the dummy variables, as for the earlier model containing the coefficient C_{RR_00}, and partly because the variation in T_{RR} is very small. The problem was solved by removing the variable T_{RR}.

Table 7. *Statistics for the regression coefficients in eq. 5.3.*

	Coef	*SE Coef*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*VIF*
Intercept	-2,5450	23,0969	-0,1102	0,912260039	-47,8168	42,7267	
C_time	0,0000	0,0000	-17,4951	4,68E-68	-0,0001	0,0000	1,22
CT_RR	0,0018	0,0003	7,0047	2,55E-12	0,0013	0,0023	344,11
C_temp	XXX	0,0000	-37,7013	3,61E-301	XXX	XXX	1,36
CL	1,0129	0,0016	629,4677	0	1,0098	1,0161	1,27
CL_beta	0,2511	0,0284	8,8289	1,14E-18	0,1954	0,3069	1,29
Direc	12,5328	0,4027	31,1232	5,30E-208	11,7435	13,3221	1
D1	60,8990	1,5114	40,2942	0	57,9366	63,8614	1,33
D2	43,8829	1,1445	38,3412	0	41,6395	46,1263	2,14
D3	8,0040	2,8615	2,7971	0,005160939	2,3952	13,6129	16,26
D4	8,6658	2,1456	4,0389	5,39E-05	4,4603	12,8713	8,68
D5	34,5777	1,0815	31,9713	4,10E-219	32,4578	36,6975	2,85
D6	-7,0977	3,4638	-2,0491	0,040466289	-13,8871	-0,3083	38,56
D7	-3,2429	3,8853	-0,8347	0,403910831	-10,8584	4,3725	38,75
D8	-38,0451	9,3746	-4,0583	4,96E-05	-56,4201	-19,6701	183,66
D9	-39,1692	7,7376	-5,0622	4,18E-07	-54,3354	-24,0029	121,6
D10	4,5850	1,1417	4,0160	5,94E-05	2,3472	6,8228	2,54

The vehicle speed v was entered in the model as a part of the rolling resistance coefficient in order to capture its contribution to tire temperature and rolling resistance. The vehicle speed was however found to be very strongly correlated to the air resistance which causes problems with multicollinearity between those variables. The correlation, illustrated in Table 8, was close to 96 % between the variables $N \times v$ and *AIR*. The VIF-test provided values over 13 (See Table C2 in Appendix C). Also a test to reintroduce the T_{RR}-values, this time as a combined variable with the speed v, was done since T_{RR} are known from earlier tests at VCC to be slightly speed-dependent. But also this variable caused multicollinearity problems. Although the adj-R2-value was increased from 96.16 % to 97.57% when the speed-variable v was entered into the model, it had to be removed in order to eliminate the multicollinearity and thus improve the estimates of the regression variables.

Table 8. *Correlation analysis for the variables included in the model. The dummies are out of picture, the total analysis can be found in Table B2 in Appendix B.*

	*N*Time*	*N*v*	*N*T*	*AIR*	*AIR*sin(β)*
*N*Time*	1,00				
*N*v*	-0,12	1,00			
*N*T*	0,27	-0,04	1,00		
AIR	-0,13	0,96	-0,08	1,00	
*AIR*sin(β)*	-0,08	0,41	0,08	0,40	1,00

5.4. Final regression model

The model that provides the highest value of *adj-*R^2 without causing problems with multicollinearity is described in Eq. 5.4.

$$F = \underbrace{N \cdot (C_{temp} \cdot T + C_{time} \cdot Time)}_{F_{RR}} + \underbrace{(C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot AIR}_{F_{AIR}} + \underbrace{Direc + D1 + ... + D11}_{Dummies} \qquad (5.4)$$

The *adj-*R^2-value and ANOVA table from the regression analysis is illustrated in Table 9 and the statistics for the regression coefficients can be found in Table 10.

Table 9. *Summary output and ANOVA-table*

SUMMARY OUTPUT

Regression Statistics	
R Square	0,961613
Adjusted R Square	0,961586
Standard Error	29,40146
Observations	21300

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	15	4,61E+08	30726338	35544,56	0
Residual	21284	18398861	864,4457		
Total	21299	4,79E+08			

Table 10. *Statistics for the regression coefficients*

	Coef	*SE Coef*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*VIF*
Intercept	159,0079	1,2462	127,5897	0	156,5652	161,4507	
C_time	-0,0001	0,0000	-16,5895	2,02E-61	-0,0001	0,0000	1,19
C_temp	XXX	0,0000	-37,5534	6,70E-299	XXX	XXX	1,36
CL	1,0130	0,0016	628,7740	0	1,0098	1,0161	1,27
CL_beta	0,2463	0,0285	8,6536	5,34E-18	0,1906	0,3021	1,29
Direc	12,5329	0,4031	31,0884	1,49E-207	11,7427	13,3231	1
D1	59,9321	1,5067	39,7760	0	56,9788	62,8854	1,32
D2	40,7072	1,0521	38,6913	0	38,6450	42,7694	1,8
D3	-10,7814	0,9993	-10,7893	4,55E-27	-12,7401	-8,8228	1,98
D4	-4,5659	1,0187	-4,4822	7,43E-06	-6,5626	-2,5692	1,95
D5	31,1567	0,9661	32,2512	7,69E-223	29,2631	33,0503	2,27
D6	16,3041	0,9159	17,8012	2,24E-70	14,5088	18,0993	2,69
D7	23,1456	0,9516	24,3238	6,31E-129	21,2804	25,0107	2,32
D8	27,2177	1,0396	26,1820	9,62E-149	25,1801	29,2553	2,25
D9	14,5703	1,0074	14,4629	3,48E-47	12,5956	16,5449	2,06
D10	0,5802	0,9893	0,5864	0,557588865	-1,3590	2,5194	1,9

As can be seen in Table 9, the value of *Significance F* is equal to zero which means that the null hypothesis can be rejected and the model as a whole can be considered as significant.

Also all regression coefficients in Table 10, except for the dummy variable *D10*, are significant with a p-value well below the limit at 0.05. The VIF-values are low for all coefficients and no high correlations were found in the correlation analysis (Table B3 in Appendix B).

5.5. Analysis and discussion of final regression model

An analysis of the regression coefficients and the regression model as a whole is presented below.

5.5.1. *Time*-variable

The low p-value of the C_{time} -coefficient tells that there is a significant time dependency in the model. The negative sign of the coefficient means that the road load decreases when the hour of the day increase.

The residual plot for the variable in Fig. 12 shows that the prerequisites from Section 3.6 are met. The vertical bar of outliers to the left in Fig.12 derives from one round of test with the same car, which indicates that something may be wrong with that test. No errors could however found in the test protocol and they were therefore not removed in order to not interfere with the natural variation.

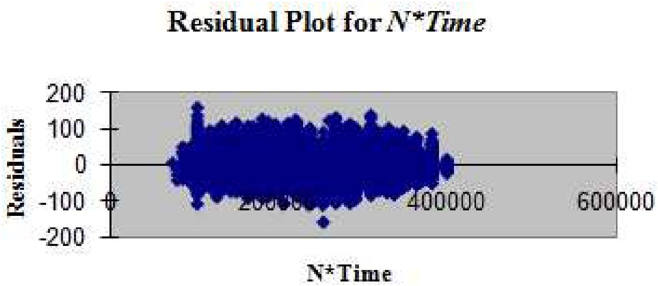

Fig. 12. *The residual plot for the rolling resistance coefficient Time.*

5.5.2. Ambient temperature

The coefficient related to the ambient temperature, C_{temp} , has been very stable for all different models tested. It says that the rolling resistance decrease when the temperature increase. That agrees well with the theory that the tire pressure, and thus the rolling resistance, is dependent of the ambient temperature.

The prerequisites from Section 3.6 are met, according to the residual plot in Fig.13. Also in this residual plot the outliers described in Section 5.5.1 can be seen.

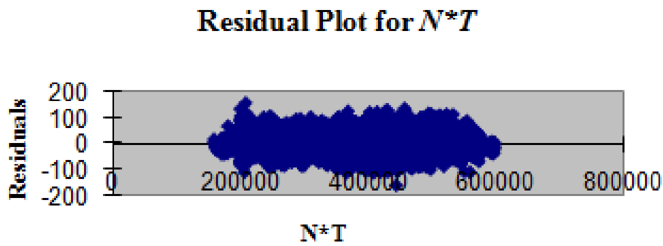

Fig. 13. *The residual plot for the ambient temperature times the normal force.*

5.5.3. Air resistance

C_L has appeared relatively stable between the different models, but there was a slightly difference when the dummy variables were introduced which tells that it probably is something more than just the $C_D \cdot A$-value that differentiates the cars in terms of air resistance. This effect is even more evident when it comes to the cross-wind coefficient, C_{Lbeta}, which differ greatly between the model with dummy variables and the one without. Based on that result, it can be assumed that another car model-specific variable is needed in the equation, for example a new $C_{D\_side} \cdot A_{side}$-variable where the area A and C_D are measured obliquely from the side.

In Fig.14 a slight departure of linearity can be seen for values of AIR exceeding 400N, which means vehicle speeds of 110 km/h and more. This effect may be due the lack of the vehicle speed term in the rolling resistance coefficient that was removed in Section 5.3 due to multicollinearity problems.

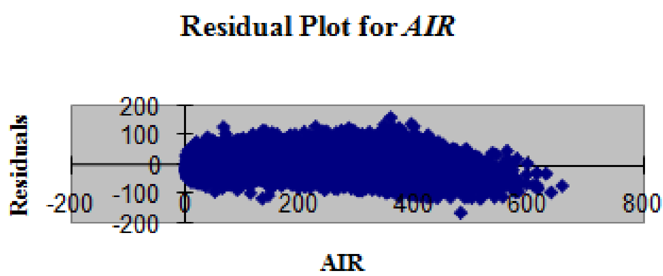

Fig. 14. *Residual plot for the air resistance.*

With the speed included in the rolling resistance coefficient, the residuals for *AIR* appeared much better for high speeds but shows instead a departure of linearity for low speeds, which is illustrated in Fig.15.

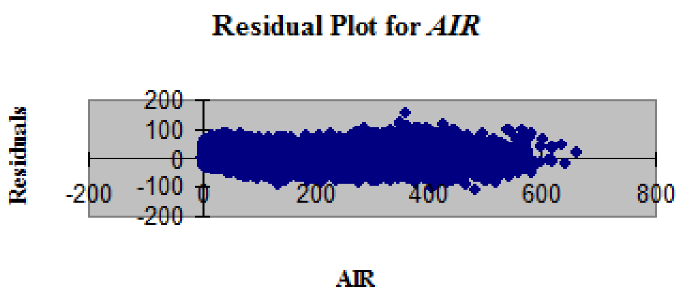

Fig. 15. *Residual plot for the cross-wind effect with the vehicle speed included in the rolling resistance coefficient.*

The speed must however be removed in order to solve the severe multicollinearity problem that occurs otherwise. Therefore, the model should be used with some caution for those high speeds. For lower speeds the model fulfils the requirements for linearity. The standard deviation seems to decrease for high values of $AIR \cdot \sin(\beta)$ in Fig.16, but the number of observations does also strongly decrease for increased values of $AIR \cdot \sin(\beta)$ and it is therefore not possible to determine if the standard deviation actually is equal for all levels or not.

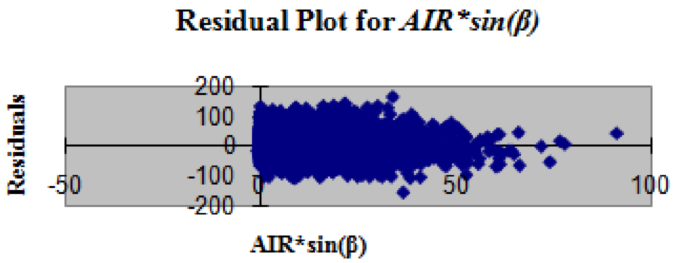

Fig. 16. *Residual plot for the cross-wind effect.*

5.5.4. Dummy variables

The value of the dummy variables for the different car variants shows the average difference in road load for the various variants relative to the reference vehicle, which in those calculations is the *D11*. For example, a coast-down test with a *D8* would in average yield a road load 27N higher than for the *D11* during the same boundary conditions. The values of the coefficients are in good agreement with earlier measurements performed by VCC. Interesting is that the coefficients for the two station wagons, *D8* and *D7* are quite similar and so also the coefficients for the two sedans, *D6* and *D9.* That strengthens the theory that a new kind of aerodynamic variable is needed in order to better calculate the cross-wind effects.

The dummy variable *Direc* indicates that it is more preferable to drive backwards on the test track than forwards. The road load is in average 12N higher when driving forward. The reason for this effect is far from obvious, especially since earlier measurements show that the test track is totally level. But the probable cause for this is the fact that the cars were not driven at the same part of the test track for the two directions and that the asphalt structure differs between those different parts. This is further discussed in Section 5.6.4. When introducing this dummy variable in the model, the other coefficients remained at the same values which indicate that this dummy variable measure something that is not cover earlier in the model, such as the road surface influence on the rolling resistance.

5.5.5 Total model

The final regression model provides an *adj-*R^2-value at over 96% which means that it is less than 4% of all road load variation that cannot be explained. Those 4% consists of both measure errors and parameters that were not included in the model. Since the meteorological measuring equipment is located at the side of the test track, it could be as much as 1km between the coasted car and the weather station and it is hence possible that the measured wind is not the one that actually affects the car. In Section 5.3 it was shown that introducing the vehicle speed v in the rolling resistant coefficient increased the *adj-*R^2-value to 97.6%. Thus, there is potential of further improve the model by solving the multicollinearity in another way than removing the correlated variable.

5.6. Stability of parameters with respect to subsets of data

The final regression model was used on three different subsets of data, one for each coast-down expedition. The result is illustrated in Appendix D. The comparison is not entirely fair since the various coast-down expeditions consists of different numbers of test, but it gives an indication of which parameters that appears stable.

5.6.1 C_{time}

The coefficient for the *Time* variable, C_{time}, are negative for all expeditions, but varies between $-2\cdot10^{-5}$ and $-7\cdot10^{-5}$.

5.6.2 C_{temp}

The ambient temperature coefficient, C_{temp}, appears relatively stable for the different expeditions.

5.6.3. C_L

The C_L coefficient is very stable and varies just between 0.997 and 1.014. This is probably due to that the AIR-variable mainly depends on the vehicle speed, which is accurately measured without much measuring faults.

5.6.3. C_{Lbeta}

This cross-wind parameter is rather stable for the two coast-down expeditions performed in 2010; it varies between 0.278 and 0.243. But for the expedition in 2011 the C_{Lbeta}-value is only 0.124, about the half of the other two. It is significant, but with a relatively small margin compared to the two older expeditions; the p-value is only 0.04. That indicates a problem with the wind direction data. It is known that the wind gauge was turned 180 degrees at an unknown date during the expedition in 2011 and the dataset was compensated for that by adding 180 degrees to the wind direction for those dates when the wind gauge most likely was turned and that provided the highest value of *adj*-R^2.
It is also known that minor modification was done to the weather station in the middle of the expedition in 2011. By only using weather data before that date, the C_{Lbeta} instead increased to 0.49 and the p-value decreased significantly. The C_L coefficient was also slightly affected and decreased to 1.006. Thus, this modification of the weather station probably affected the accuracy of the wind data.

5.6.4. Dummy variables

The direction dummy for the two expeditions in 2010 both show that it is significantly preferable to drive backwards on the test track, even if the value is a bit unstable and vary from 15,6N to 22,6N. The value for 2011 is instead slightly negative, -1.4N, and with a higher p-value than for the other expeditions. Those results agree well with the fact that at the expeditions performed in 2010 the cars were not driven at the same part of the test track for the different directions. When the cars were driven at the direction called forward, they pass a larger section of the track with inferior asphalt quality than when they are driven at the other direction. The expedition in 2011 was performed on two different test tracks at APG and the tests were performed at the same part of the track for the different directions. It is hence very likely that the dummy variable *Direc* can be used as a measure for the road surface influence on the rolling resistance. In the report by Karlsson, et al. (2011) it stated that differences in road surface quality can influence the road load at the same order as found by this dummy variable.

The stability of the dummy variables for the car-variant is naturally hard to validate since they are not generic for all expeditions.

6. Conclusions

With a multiple regression analysis performed with the model

$$F = \underbrace{N \cdot (C_{temp} \cdot T + C_{time} \cdot Time)}_{F_{RR}} + \underbrace{(C_L + C_{Lbeta} \cdot \sin(\beta)) \cdot AIR}_{F_{AIR}} + \underbrace{Direc + D1 + ... + D11}_{Dummies}$$

on a dataset containing coast-down results from three independent coast-down expeditions with in total eleven different car-variants, over 96% of the total variation of the road load was explained. The values of the regression coefficients are presented in the table below.

	Coef	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	159,0079	0	156,5652	161,4507
C_time	-0,0001	2,02E-61	-0,0001	0,0000
C_temp	XXX	6,70E-299	XXX	XXX
CL	1,0130	0	1,0098	1,0161
CL_beta	0,2463	5,34E-18	0,1906	0,3021
Direc	12,5329	1,49E-207	11,7427	13,3231
D1	59,9321	0	56,9788	62,8854
D2	40,7072	0	38,6450	42,7694
D3	-10,7814	4,55E-27	-12,7401	-8,8228
D4	-4,5659	7,43E-06	-6,5626	-2,5692
D5	31,1567	7,69E-223	29,2631	33,0503
D6	16,3041	2,24E-70	14,5088	18,0993
D7	23,1456	6,31E-129	21,2804	25,0107
D8	27,2177	9,62E-149	25,1801	29,2553
D9	14,5703	3,48E-47	12,5956	16,5449
D10	0,5802	0,557588865	-1,3590	2,5194

From the above table it can be seen that all regression variables, except the dummy variable *D10*, are significant for a confidence level at 95%. The analysis of the coefficients showed the value of C_{time} fluctuates between the different coast-down expeditions but that it is always significantly negative. That means that the road load is lower when performing the coast-down test later in the day, but exactly how much lower differ from various expeditions and the actual cause for this cannot be established with the present data. The coefficient C_{temp} has appeared stable for the different models and datasets and the value can be seen as secured for this model. The air resistance parameter C_L has also been stable around 1.01, but a small departure of linearity for high vehicle speeds, 110 km/h and more, can be seen in the residual plot. This problem is probably due to the lack of the vehicle speed term in the rolling resistance coefficient that was removed due to severe multicollinearity problems. Therefore, the model should be used with some caution for those high speeds. The other air resistance parameter for the cross-wind effects, C_{Lbeta}, appears rather stable for two of the expeditions with a value between 0.243 and 0.278, but for the last expedition the value is about the half and barely significant. The reason for this behaviour is probably those experiments that were done with the wind gauge during the expedition that may have perturbed the wind direction data. It is hence not possible to determine a totally sure value of C_{Lbeta}, even if it most likely is close to the value in the above table.

The value of the dummy variable coefficient *Direc* tells that it is in average 12N heavier to drive in the direction called "Forward" on the test track, compared to "Backward". For the expeditions performed in 2010 that value is even higher, between 15,6N to 22,6N, but for the last expedition in 2011 the coefficient is slightly negative and also barley significant. In 2010, when the cars were driven at the direction called forward, they pass a larger section of the track with inferior asphalt quality than when they are driven at the other direction. The expedition in 2011 was performed on two different test tracks at APG and the tests were performed at the same part of the track for the different directions. Thus, it can be considered that the differences in road surface has a relatively large impact on the rolling resistance and that it is important to run the coast-down test where the asphalt quality is best.

The car specific dummies cover those differences between the car variants that the model does not handle, different gearboxes for example. Those are no absolute values but relative values to the reference dummy, *D11*. The differences between the variants are considered as reasonable according to the experience of previous tests at VCC. When introducing those dummy variables in the model, the air resistance coefficients changed and C_{Lbeta} in particular. That indicates that another car specific aerodynamic variable is needed in the model, probably some kind of a $C_{D\_side} \cdot A_{side}$-variable where the area A and C_D are measured obliquely from the side.

To sum up, the regression model meets the goal of this master thesis by explaining just over 96% of the total variation in the coast-down results and thus describing the true road load within less than +/- 4%. The model can be used to normalize the boundary conditions at a coast-down expedition in order to investigate if whether the obtained results are representative in relation to the circumstances or not. In that way the number of test runs can be decreased and the coast-down expedition may be made quicker, cheaper and with more precise and reliable result, which also was the purpose of this thesis.

6.1. Future work

To make this model even better and more accurate, the multicollinearity problem must be solved in another way. A vehicle speed variable in the rolling resistance parameter would increase the *adj*-R^2-value and probably also solve the problem in the air resistance variable for high speeds.

To further understand the road surface impact on the total road load, better and larger numbers of measurements of the road surface is needed in order to analyze it in a regression model, but most important is to make it possible to connect those results to the coast-down results so the data can be used.

Bibliography

Ahern, J. L. (2004) *International Gravity Formula(e).* http://geophysics.ou.edu/solid_earth/notes/potential/igf.htm (12 March 2012).

Bauer, H. (1996) *Automotive handbook Bosch, 4:th edition.* Stuttgart: Robert Bosch GmbH.

Ender, P. (1998) *Statistical Tables.* http://www.philender.com/courses/tables/dist3.html (3 April 2012).

Grandin, U. (2012)*Dataanalys och hypotesprövning för statistikanvändare.* Uppsala: Swedish Environmental Protection Agency.

Heisler, H. (2002) *Advanced vehicle technology.* Oxford: Butterworth-Heinemann.

Hilmersson, M. (2010) *Coast down - Easy description.* [Internal Power Point at VCC]

Karlsson, R., Hammarström, U., Sörensen, H., Eriksson, O.(2011) *Road surface influence on rolling resistance* . Linköping : VTI.

Körner, S.,Wahlgren, L. (2006) *Statistisk Dataanalys.* Lund: Studentlitteratur.

Nationalencyklopedin. *Multikollinearitet* http://www.ne.se/multikollinearitet (24 April 2012).

Nielsen, L., Sandberg,T. (2002) *A new model for rolling resistance of pneumatic tires* . Linköping : Vehicular Systems, ISY.

Petterson, E. (2003) *Matematisk statistik.* Gothenburg: Matematiklitteratur i Göteborg.

Prerau, D. (2006) *Seize the Daylight.* New York: Basic Books.

ProfTDub.(2010) *Calculating Variance Inflation Factors in Excel 2007.[Youtube]* http://www.youtube.com. (6 April 2012).

Regulations No, 83-05. (2009) *Emissions - Light Duty Vehicles.* InterRegs Ltd.

Shelquist, R. (2011) *An Introduction to Air Density and Density Altitude Calculations* . http://wahiduddin.net/calc/density_altitude.htm (17 February 2012).

Simon, L,J. (2004) *Detecting multicollinearity using variance inflation factors.* http://online.stat.psu.edu/online/development/stat501/12multicollinearity/05multico_vif.html (14 April 2012).

Sundell, A. (2010) *Guide: Regressionsdiagnostik – multikollinearitet.* http://spssakuten.wordpress.com/2010/10/16/guide-regressionsdiagnostik-%E2%80%93-multikollinearitet/ (20 April 2012).

UCLA. (2007) *UCLA: Academic Technology Services, Statistical Consulting Group.* . http://www.ats.ucla.edu/stat (17 April 2012).

UKY. (2010) *Multicollinearity in Logistic Regression.* http://www.uky.edu/ComputingCenter/SSTARS/MulticollinearityinLogisticRegression.htm (15 April 2012).

Umeå University (2004) *Numeriska Metoder*. http://www8.cs.umu.se/kurser/TDBA68/HT04/forelasant/F3.pdf (24 April 2012).

UNSW.(2011) Goodness Of Fit http://web.maths.unsw.edu.au/~adelle/Garvan/Assays/GoodnessOfFit.html (12 April 2012).

Zejda, M.(2008) *Astronomy & Astrophysics.* http://www.aanda.org/index.php?option=com_article&access=standard&Itemid=129&url=/articles/aa/full/2008/37/aa8632-07/aa8632-07.right.html (5 April 2012).

Appendix A - Calculation of air density

The air density can be written as:

$$\rho_{air} = \frac{P}{R_d \cdot T} \cdot \left(1 - \frac{0.378 \cdot p_v}{P}\right)$$

Where:

P = total air pressure [Pa]

R_d = specific gas constant for dry air = 287,05 [J/(kg*°K)]

T = temperature [°K]

p_v = pressure of water vapour (partial pressure) [Pa]

The vapour pressure, p_v, can expressed as:

$$p_v = \phi \cdot p_{sat}$$

Where:

ϕ =relative humidity [%]

p_{sat} = saturation vapour pressure [Pa]

To determine p_{sat}, a simplification of Herman Wobus polynomial[1] can be used with good accuracy, especially at higher air temperatures where the saturation pressure becomes significant for the density calculations. The saturation vapour pressure is then expressed as: (Shelquist, 2011).

$$p_{sat} = c_0 \cdot 10^{\frac{c_1 \times T_C}{c_2 + T_C}} \ \textit{[mbar]}$$

Where:

C0 = 6,1078
C1 = 7,5
C2 = 237,3

1.P_{sat} = 6.1078/(c0+T*(c1+T*(c2+T*(c3+T*(c4+T*(c5+T*(c6+T*(c7+T*(c8+T*(c9))))))))))8

Appendix B – Correlation analyses

Table B1. *Correlation analysis for the first regression model including dummy variables*

	N	*N*T*	*AIR*	*AIR* sin(β)*	*Direc*	*D1*	*D2*	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*
N	1,00															
*N*T*	0,41	1,00														
AIR	0,07	-0,08	1,00													
AIR sin(β)*	0,06	0,08	0,40	1,00												
Direc	0,00	0,00	-0,02	-0,03	1,00											
D1	0,00	0,03	0,02	0,00	0,00	1,00										
D2	-0,06	-0,08	0,02	-0,04	0,00	-0,04	1,00									
D3	-0,42	-0,10	-0,03	-0,07	0,00	-0,05	-0,09	1,00								
D4	-0,33	-0,14	-0,02	-0,09	0,00	-0,05	-0,08	-0,09	1,00							
D5	-0,08	0,02	0,04	-0,07	0,00	-0,06	-0,10	-0,11	-0,11	1,00						
D6	0,23	0,17	0,01	0,06	0,00	-0,07	-0,12	-0,13	-0,13	-0,15	1,00					
D7	0,28	0,14	-0,02	0,03	0,00	-0,06	-0,10	-0,11	-0,11	-0,13	-0,16	1,00				
D8	0,58	0,24	0,07	0,11	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,14	-0,12	1,00			
D9	0,31	0,04	0,00	-0,10	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,13	-0,12	-0,10	1,00		
D10	-0,38	-0,23	-0,05	0,07	0,00	-0,05	-0,09	-0,10	-0,09	-0,11	-0,13	-0,11	-0,10	-0,10	1,00	
D11	-0,27	-0,17	-0,04	0,08	0,00	-0,05	-0,08	-0,09	-0,09	-0,10	-0,13	-0,11	-0,09	-0,09	-0,09	1,00

Table B2. *Correlation analysis for the regression model with the vehicle speed, v, included in the rolling resistance coefficient.*

	N Time*	*N*v*	*N*T*	*AIR*	*AIR* sin(β)*	*Direc*	*D1*	*D2*	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*
N Time*	1,00																
*N*v*	-0,12	1,00															
*N*T*	0,27	-0,04	1,00														
AIR	-0,13	0,96	-0,08	1,00													
AIR sin(β)*	-0,08	0,41	0,08	0,40	1,00												
Direc	0,00	0,00	0,00	-0,02	-0,03	1,00											
D1	0,08	0,02	0,03	0,02	0,00	0,00	1,00										
D2	-0,01	-0,01	-0,08	0,02	-0,04	0,00	-0,04	1,00									
D3	0,02	-0,07	-0,10	-0,03	-0,07	0,00	-0,05	-0,09	1,00								
D4	0,06	-0,06	-0,14	-0,02	-0,09	0,00	-0,05	-0,08	-0,09	1,00							
D5	0,05	0,03	0,02	0,04	-0,07	0,00	-0,06	-0,10	-0,11	-0,11	1,00						
D6	0,07	0,03	0,17	0,01	0,06	0,00	-0,07	-0,12	-0,13	-0,13	-0,15	1,00					
D7	-0,13	0,00	0,14	-0,02	0,03	0,00	-0,06	-0,10	-0,11	-0,11	-0,13	-0,16	1,00				
D8	0,19	0,11	0,24	0,07	0,11	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,14	-0,12	1,00			
D9	-0,02	0,03	0,04	0,00	-0,10	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,13	-0,12	-0,10	1,00		
D10	-0,12	-0,04	-0,23	-0,05	0,07	0,00	-0,05	-0,09	-0,10	-0,09	-0,11	-0,13	-0,11	-0,10	-0,10	1,00	
D11	-0,16	-0,05	-0,17	-0,04	0,08	0,00	-0,05	-0,08	-0,09	-0,09	-0,10	-0,13	-0,11	-0,09	-0,09	-0,09	1,00

Table B3. *Correlation analysis for the final regression model*

	*N*Time*	*N*T*	*AIR*	*AIR* sin(β)*	*Direc*	*D1*	*D2*	*D3*	*D4*	*D5*	*D6*	*D7*	*D8*	*D9*	*D10*	*D11*
*N*Time*	1,00															
*N*T*	0,27	1,00														
AIR	-0,13	-0,08	1,00													
AIR sin(β)*	-0,08	0,08	0,40	1,00												
Direc	0,00	0,00	-0,02	-0,03	1,00											
D1	0,08	0,03	0,02	0,00	0,00	1,00										
D2	-0,01	-0,08	0,02	-0,04	0,00	-0,04	1,00									
D3	0,02	-0,10	-0,03	-0,07	0,00	-0,05	-0,09	1,00								
D4	0,06	-0,14	-0,02	-0,09	0,00	-0,05	-0,08	-0,09	1,00							
D5	0,05	0,02	0,04	-0,07	0,00	-0,06	-0,10	-0,11	-0,11	1,00						
D6	0,07	0,17	0,01	0,06	0,00	-0,07	-0,12	-0,13	-0,13	-0,15	1,00					
D7	-0,13	0,14	-0,02	0,03	0,00	-0,06	-0,10	-0,11	-0,11	-0,13	-0,16	1,00				
D8	0,19	0,24	0,07	0,11	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,14	-0,12	1,00			
D9	-0,02	0,04	0,00	-0,10	0,00	-0,05	-0,09	-0,10	-0,10	-0,11	-0,13	-0,12	-0,10	1,00		
D10	-0,12	-0,23	-0,05	0,07	0,00	-0,05	-0,09	-0,10	-0,09	-0,11	-0,13	-0,11	-0,10	-0,10	1,00	
D11	-0,16	-0,17	-0,04	0,08	0,00	-0,05	-0,08	-0,09	-0,09	-0,10	-0,13	-0,11	-0,09	-0,09	-0,09	1,00

Appendix C – Statistics for regression analyses

Table C1. *Statistics for regression model without the C_{00}-coefficient*

	Coef	SE Coef	t Stat	P-value	Lower 95%	Upper 95%	VIF
Intercept	153,2163	1,2040	127,2532	0	150,8563	155,5763	
C_temp	XXX	0,0000	-41,7044	0	XXX	XXX	1,30
CL	1,0159	0,0016	630,3632	0	1,0127	1,0190	1,26
CL_beta	0,2651	0,0286	9,2618	2,20E-20	0,2090	0,3213	1,29
Direc	12,5886	0,4057	31,0283	8,91E-207	11,7933	13,3838	1,00
D1	56,7089	1,5038	37,7115	2,51E-301	53,7614	59,6564	1,30
D2	39,1506	1,0546	37,1223	2,46E-292	37,0834	41,2178	1,79
D3	-12,6739	0,9991	-12,6851	9,71E-37	-14,6322	-10,7155	1,95
D4	-7,0641	1,0139	-6,9669	3,33E-12	-9,0515	-5,0767	1,91
D5	29,1996	0,9650	30,2590	5,84E-197	27,3082	31,0911	2,24
D6	14,5534	0,9156	15,8942	1,47E-56	12,7587	16,3481	2,65
D7	23,3121	0,9576	24,3439	3,91E-129	21,4351	25,1891	2,32
D8	24,0887	1,0289	23,4128	1,03E-119	22,0720	26,1053	2,18
D9	13,5216	1,0119	13,3626	1,46E-40	11,5382	15,5050	2,05
D10	-0,0241	0,9950	-0,0242	0,980659	-1,9744	1,9262	1,90

Table C2. *Statistics for regression model with the vehicle speed included in the rolling resistance coefficient*

	Coef	SE Coef	t Stat	P-value	Lower 95%	Upper 95%	VIF
Intercept	104,5316	1,1055	94,5575	0	102,3647	106,6984	
C_time	0,00005	0,0000	-20,6117	1,76293E-93	0,0000	0,0000	1,19
C_v	0,0005	0,0000	111,2299	0	0,0004	0,0005	13,10
C_temp	XXX	0,0000	-50,7290	0	XXX	XXX	1,35
CL	0,5784	0,0041	140,6794	0	0,5703	0,5865	13,01
CL_beta	0,0794	0,0227	3,5009	0,000464658	0,0350	0,1239	1,29
Direc	9,6131	0,3217	29,8853	2,839E-192	8,9826	10,2436	1,00
D1	60,3946	1,1982	50,4030	0	58,0460	62,7433	1,31
D2	44,5178	0,8374	53,1632	0	42,8765	46,1592	1,80
D3	-5,7242	0,7960	-7,1915	6,62048E-13	-7,2843	-4,1640	1,97
D4	3,0542	0,8130	3,7567	0,000172602	1,4607	4,6477	1,95
D5	30,7072	0,7683	39,9693	0	29,2013	32,2131	2,26
D6	11,1794	0,7298	15,3180	1,10851E-52	9,7488	12,6099	2,69
D7	18,3642	0,7579	24,2290	5,9445E-128	16,8786	19,8498	2,31
D8	18,0877	0,8308	21,7722	5,7995E-104	16,4593	19,7161	2,26
D9	6,8122	0,8042	8,4710	2,5886E-17	5,2360	8,3885	2,06
D10	-1,6993	0,7870	-2,1591	0,030852785	-3,2419	-0,1566	1,89

Appendix D – Regression analyses for each expedition

Autumn 2011

	Coef	P-value	Lower 95%	Upper 95%
Intercept	198,6129	0	194,1488	203,0770
C_time	-0,00007	9,73E-54	-0,00007	-0,00006
C_temp	XXX	9,42E-82	XXX	XXX
CL	1,0146	0	1,00930	1,0199
CL_beta	0,1239	0,04149376	0,0048	0,2430
Direc	-1,4492	0,03231125	-2,7762	-0,1223
D1	29,4268	1,67E-85	26,5183	32,3352
D2	9,6115	2,66E-21	7,6287	11,5944
D3	-41,7326	0	-43,6071	-39,8580
D4	-35,2903	2,37E-264	-37,2077	-33,3730

Autumn 2010

	Coef	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	169,65588	0	166,38818	172,92358
C_time	-0,00004	4,31E-17	-0,00004	-0,00003
C_temp	XXX	1,15E-171	XXX	XXX
CL	0,9971	0	0,9930	1,0012
CL_beta	0,2426	4,65E-12	0,1739	0,3113
Direc	22,6576	0	21,5904	23,7248
D6	1,6435	0,03533243	0,1129	3,1740
D7	8,7591	3,44E-26	7,1420	10,3762
D8	12,7037	1,23E-44	10,9361	14,4712

Spring 2010

	Coef	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	146,2845	0	141,2321	151,3368
C_time	-0,00002	2,32E-02	-0,00003	0,00000
C_temp	XXX	2,15E-72	XXX	XXX
CL	1,0911	0	1,0821	1,1001
CL_beta	0,2781	1,78E-05	0,1512	0,4049
Direc	15,5010	2,16E-55	13,6040	17,4160
D10	0,0703	0,94291	-1,8552	1,9959