# USING QUALITY CONTROL CHARTS TO SEGMENT ROAD SURFACE CONDITION DATA

Amin El Gendy
Department of Civil Engineering
University of Manitoba
Phone : (204) 474-8376
Fax : (204) 474-7513
umelgend@cc.umanitoba.ca

Ahmed Shalaby
Department of Civil Engineering
University of Manitoba
Phone : (204) 474-6818
Fax : (204) 474-7513
shalabya@cc.umanitoba.ca

**Reference**

**El Gendy, A., and Shalaby, A., Using quality control charts to segment road surface condition data, International conference on managing paving assets, ICMPA 2008, Calgary, AB, June 2008.**

**ABSTRACT**

Road surface condition data are collected for the purpose of building effective asset management systems to support the analysis of maintenance and rehabilitation strategies. With the growth of the collected data, there is a great need to identify homogeneous road segments to use these data effectively for planning maintenance or rehabilitation strategies. For this reason, a road section may be segmented into homogenous subsections which have statistically uniform properties. This paper discusses the current strategies for segmenting linearly referenced pavement condition data and the limitations of these segmentation methods are addressed. The classical cumulative difference approach and the absolute difference approach are reviewed. A third approach that uses quality control charts is introduced. The fundamental concepts of the quality control charts are reviewed and its suitability for segmentation is examined. Then, a target range is used to improve the selection of the c-chart control limits. Finally, an example to compare results of the c-chart segmentation with previous segmentation methods shows that the presented method is able to identify homogenous segment borders.

# 1. Introduction

Monitoring the pavement surface condition is an essential element of building an effective asset management system to support the analysis of maintenance and rehabilitation strategies. Surface condition information can be obtained by collecting either continuous or spot measurements of pavement response variables. Examples of these variables are roughness, deflection, serviceability index, friction number, pavement condition indices, or even individual distress severities such as percent cracking and rut depth.

With the growth of the collected data, there is a great need to identify homogeneous segments to use these data effectively for planning maintenance and rehabilitation strategies. The road section has to be segmented into homogenous subsections which have consistent statistical properties so that information stored for each segment can be summarized without losing any significant information within each segment. The ability to delineate the general boundary locations of these segments effectively is very important for the rehabilitation analysis. The proper consideration of segment variability is an effective way of rationalizing the condition of each segment and the investment needs.

This paper provides background information related to methodologies used to define the boundary limits of relatively uniform segments. Two of the classical approaches that address the problem of segmentation of linearly referenced data are discussed. These are the cumulative difference approach (CDA) recommended by AASHTO (*1*) and the absolute difference approach (ADA) introduced by El Gendy and Shalaby (2). The limitations of the pervious segmentation methods are examined

A new method based on statistical quality control charts (c-charts) is presented. The fundamentals of the quality control charts are reviewed then their use as a segmentation tool is evaluated. A numerical example is used to compare results of the c-chart segmentation with previous segmentation methods. Finally, the target range is used to enhance the determination of the c-chart control limits.

# 2. Current Profile Segmentation Methods and Limitations

The cumulative difference approach (*1*), CDA, is a simple and powerful analytical method for segmenting linearly referenced road condition information. The cumulative difference approach (CDA) creates segment borders at maxima and minima locations of the cumulative difference between a response indicator and the average response over an entire section. The approach can be applied to variety of measured pavement response variables such as IRI.

Several algorithms have been developed to enhance the performance of the CDA. Divinsky et al. (3) used the calculated standard deviation for a moving group of m measured (response) values as the input for further application of the cumulative difference approach procedure.

One of the significant limitations of the CDA is that it is affected by the overall average response. The CDA test the changes in mean with respect to the overall average or the mean of the entire section. Misra and Das (*5*) showed that the CDA fails to identify existence of homogeneous segments with various average response levels because all of these segments were either completely above or below the mean of an entire section.

For most of the road condition data, the change in cumulative differences from negative to positive or vice versa may exist within a short number of referenced measurements where there are no significant changes in the statistical properties of these measurements. Thomas (*4*) showed two methods for solving this problem either by smoothing the profile data before applying the CDA or by finding local peaks which their values are the most extreme within a window of seven neighbours to the left and seven neighbours to the right of each peak.

Thomas (*4*) also stated that since any series of cumulative differences starts and ends at zero, at least two segments will be identified in any dataset and has at least one peak in between. The same problem exists in the CART approach proposed by Misra and Das (*5*) where a section is divided into two segments in each step even if there is no change to its statistical properties.

Efforts were also made to develop different approaches to delineate a road profile. Thomas (*4*) implemented a segmentation algorithm based on Bayesian analysis. The algorithm searches a given set of measurements sequentially for transitions between neighboring homogeneous sections. A segment is defined to be homogeneous if the associated measurements can be described by a single first–order autoregressive process. Cuhadar et. al. (*6*) used the wavelet transform to automate the segmentation of pavement condition data. During the segmentation stage, singularities of the smoothed waveform are detected, and they are marked either as isolated singularities or border points.

Misra and Das (*5*) suggested a methodology for delineating homogeneous segments based on a combined approach of Classification and Regression Trees, CART. First, the search procedure divides a data set into two homogeneous segments by locating the position where the sum of the squared differences between the data in each segment and the corresponding mean of each segment is minimized. The procedure can be applied recursively to each segment until a maximum number of segments or a minimum segment length is reached. Then the potential for joining adjacent segments based on having similar statistical properties is examined. The joining is performed if the resulting segments are considered uniform.

El Gendy and Shalaby (2) proposed the absolute differences approach, ADA, that depends on limiting the absolute difference between response values within each segment by defining a sliding window controlling the maximum difference between individual responses within each segment. A graphical illustration of the ADA concept is shown in Figure 1. Segment border is determined when the absolute difference between the minimum and maximum responses reaches the target range, $r_{range}$, selected by the user and a new segment will start from this border. This concept is based on using a sliding window with a height equal to the target range. Each new segment will start from the point of intersection between the pavement response profile and the upper or lower limit of the pervious window. The target range is an arbitrary value and should be specified according to the required level of details.

# 3. Segmentation Using Quality Control Charts

## 3.1. General Model for Control Chart

In quality control process, The CDA approach is known as Cusum chart (*8*, *9*). In general, a change in slope of the Cusum indicates a change in the mean of the process. Thus the Cusum can indicate approximately when the mean of the process has shifted from the expected mean.

Another common approach used in quality control process is the Shewhart control charts or c-charts *(10)*. In the c-chart quality control, and to control the mean of a certain process, upper and lower control limits, UCL and LCL, are established from the standard deviation of a sample as shown in Figure 2. When referenced values start to fall outside the control limits, the system is considered in a state out of control and action should be taken to return the system to a state of control. The general model for control chart requires the selection of lower and upper control limits. As shown in Figure 2 the centreline CL, the upper control limit UCL, and the lower control limit LCL are:

$$\begin{aligned} UCL &= \mu + \mathrm{k}\sigma \\ CL &= \mu \\ LCL &= \mu - \mathrm{k}\sigma \end{aligned} \tag{1}$$

where k is the distance of the control limit from the centreline expressed in standard deviation unit.

Two sets of limits are widely used. The outer limits are called action limits, and are usually at 3σ while the inner limits, usually at 2σ, are called the warning limits. When responses start to fall outside the warning limits, the system is considered to be in a state out of control but no action will be taken to return the system to a state of control until responses are outside the action limits.

Although, the quality control approaches are on-line tools mainly used to control the mean of a process, the Cusum approach or CDA is able to identify homogeneous section borders. Similarly, the ability of c-chart to determine the homogenous section borders is examined.

## 3.2. Estimating Mean and Standard Deviation from Segment Data

Different estimates of mean and standard deviation of the pavement response data result in different locations of segment border. In Figure 3, and when moving from left to right to test whether a response point is inside the action limits, the number of tested segment points increases; the sample size changes. Box and Lunceno (*9*) discussed several approaches for constructing and operating a control chart with a variable sample size. One of these approaches is to determine control limits for each individual segment based on its number of data points. At different locations in the segment, the true mean μ will not be known and will be replaced by an estimate based on averaging the responses which is:

$$\hat{\mu} = \bar{r} \tag{2}$$

where

$\hat{\mu}$ = estimate of mean for current segment

$\bar{r}$ = average of responses in current segment

Also the estimate of variance can be found as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} r_i^2 - n\bar{r}^2}{n-1} \tag{3}$$

where

$r_i$ = response value

$\hat{\sigma}^2$ = estimate of variance for current segment

$n$ = number of response points ($i$) in current segment

The estimated variance can be used to detect when the system is out of control or in the segmentation process where a new segment is generated. The estimated variances depend on the number of the data points, $n$. The smaller the number of the data points the larger the error in estimating the variance, therefore a minimum of $n$=5 is used in this paper.

At each point starting from the fifth point in each tested segment, new values of the mean and control limits are calculated according to equations 2 and 3 using the segment data up to this point. As shown in Figure 3, a new segment is started once the referenced response data falls outside the control limits.

### 3.3. Modifying C-Chart Control Limits Using Response Range

For improving the segmentation method and controlling the segmentation process, additional boundaries such as specifying a minimum segment length, or a maximum range of response within each segment can be added. These boundaries depend mainly on the data type (response values) and the purpose of the analysis. El Gendy and Shalaby (*4*) introduced a relationship between the minimum segment length and the maximum IRI range. Similar relationships can be constructed for other types of response.

Also the target range for ADA, which is predetermined by the user, can affect the segmentation. ADA uses the target range to test the information of the maximum and minimum responses only. Although the range is correlated to the standard deviation, it loses efficiency rapidly as it ignores statistical information in the segment other than the maximum and minimum responses. The c-chart approach can be considered to be an enhancement to the ADA approach where the range is determined autonomously by the control limits and the statistical information of the response data within the segment.

In the proposed delineation approach, using the 3-sigma limits combined with a specified target range is recommended and the modified control limits become:

$$\begin{aligned} UCL &= \hat{\mu} + c \\ LCL &= \hat{\mu} - c \end{aligned} \tag{4}$$

Where $c$ is the minimum of the $3\hat{\sigma}$ and 0.5 $r_{range}$.

## 3.4. C-Chart Delineation Algorithm

The algorithm of the proposed c-chart delineation approach can be summarized as follows:

- Proceeding from the fifth point of the segment to allow for reasonable estimate of the statistical parameters, the estimated mean $\hat{\mu}$ and variance $\hat{\sigma}$ of the segment are calculated according to equations 2 and 3 based on all the segment data points up to the tested point.
- The minimum of $3\hat{\sigma}$ and 0.5 $r_{range}$, is used to determine the control limits using equation 4.
- A new segment is started once the tested data point falls outside the control limits.
- The process continues until all the profile data is segmented.

Table 1 summarizes a comparison between the different segmentation approaches introduced. According to the segmentation criteria, CDA does not require predetermined inputs from the user while criteria information is required for both ADA and CART. Before segmentation, the user must provide the target range for ADA and the number of segments for CART. Although the target range is used in the proposed c-chart algorithm, a profile can be segmented successfully without it when there is no information about the referenced data.

# 4. The AASHTO Example

The AASHTO guide (*1*) example is segmented using the proposed algorithm with 2-sigma and 3-sigma control limits as shown in Figure 4. Segmentation using 2-sigma control limits produces more detailed segments. Some of these segments may be joined when tested for either minimum length or minimum range criteria. Conversely, segmentation using 3-sigma limit controls may produce segments that have a large variability of data. For example, data of the segment between km-posts 13.7 and 42.6 has two main parts; from km-posts 13.7 to 27.4 and km-post 27.4 to 42.6; with average friction numbers of 30.8 and 29.4 respectively.

Figure 4 also shows the CDA segmentation and the CART segmentation introduced by Misra and Das (*5*). Although segment borders at km-posts 8.0 and 91.7 do not exist in the CDA delineation, they are identified in both the CART and the c-chart approach. This would give the c-chart approach an advantage in detecting segment variations better than the CDA method. Segment borders at km-posts 23.3, 27.4, 112.7 and 119.0 appear in both CDA and the c-chart with 2-sigma control limits. These borders do not exist in both the CART and the c-chart with 3-sigma control limits. The c-chart using 3-sigma control limits is less sensitive to detect the segment borders therefore the number of borders is expected to be smaller than that when using the 2-sigma control limits. For the CART approach, these borders may not exist for one of the following conditions:

- CART algorithm takes into account the constraint of minimum segment length, therefore borders may not exist if they will introduce segments that are shorter than the minimum length.
- CART algorithm pre-selects the number of segments to be delineated. For example, if the number of segments is *n* then the algorithm will search for the best (*n-1*) positions for interior borders.

To validate the proposed method, the residual errors, the sum of squared errors (SSE), for the segmented profiles shown in Figure 4, are computed. The SSE is defined by:

$$SSE = \sum (r_i - \bar{r}_i)^2 \qquad (5)$$

Where $\bar{r}_i$ is the average of the response of the segment which the response $r_i$ belong to. It should be noted that SSE diminishes as the number of segment approaches the number of responses, *n*.

As listed in Table 2, c-chart segmentation using either 2-sigma or 3-sigma control limits produces SSE values smaller than SSE values of CDA which means that the proposed c-chart method is an acceptable segmentation tool. However CART criterion is to minimize the SSE, the segmentation results introduced by Misra and Das (*5*) have SSE higher than the two c-chart segmentation methods due to the early termination of the CART process when the pre-selected number of segments is reached.

Compared to other segmentation approaches, the number and position of segment borders, Figure 4, and the residual errors, Table 2, would support the conclusion that c-chart can be used to delineate homogeneous segments.

# 5. Conclusions

A new method for segmenting road profiles based on linear-referenced road data is presented. The proposed approach uses quality control charts to identify homogeneous road segments. The general theory of the quality control charts is reviewed then the use of c-chart as a segmentation tool is evaluated. The target range is used to enhance the determination of the c-chart control limits.

The AASHTO (*1*) example is used to compare results of the proposed approach with previous work. The number and position of segment borders, and the residual errors of the segmented profile would support the conclusion that c-chart can be used to delineate homogeneous segments.

A better algorithm would be able not only to segment a profile but also to determine whether the entire profile could be considered as one segment. CDA and CART approaches are unable to detect if the entire profile could be considered as one segment. For both methods there will be a minimum of two segments in each profile. ADA would consider the entire profile as one unit only if the user defined a target range larger than the range of the entire profile. C-chart would able to detect if the entire profile could be statistically considered as one profile.

Finally, the main advantage of the c-chart approach is that it is an autonomous method. Without any prior information about the statistical characteristics of the data, the proposed method can segment the linear-referenced road profile successfully. Moreover, if the characteristics of data are known, additional criteria such as target range can be incorporated.

# References

1. *AASHTO Guide for Design of Pavement Structures.* AASHTO, Washington, D.C. 20001, 1993.
2. El Gendy, A., and A. Shalaby. Detecting Localized Roughness Using Dynamic Segmentation, In *the First Annual Inter-University Symposium of Infrastructure Management (AISIM)*, University of Waterloo, Waterloo ON, August 6, 2005.
3. divinsky, M., S. Nesichi, and M. Livneh. Development of a Road Roughness Profile Delineation Procedure. *Journal of Testing and Evaluation*, JTEVA, Vol. 25, No. 4, July 1997, pp. 445-450.
4. Thomas, F. Generating Homogeneous Road Sections Based on Surface Measurements: Available Methods. In *the 2nd European Pavement and Asset Management Conference*, Berlin, Germany, paper No. 48, 21-23 March 2004.
5. Misra, R., and A. Das. Identification of Homogeneous Sections from Road Data, In *International Journal of Pavement Engineering Vol. 4,* Taylor & Francis Ltd., December, 2003, pp. 229 –233
6. Cuhadar, A., A. Shalaby, and S. Tasdoken. Automatic Segmentation of Pavement Condition Data Using Wavelet Transform. In *Proceeding of the 2002 IEEE Canadian Conference on Electrical and Computer Engineering*. ,2002, pp 1009-1014
7. Thomas, F. Statistical Approach to Road Segmentation. *Journal of Transportation Engineering,* Vol. 129, No. 3, May 1,2003, pp 300-308
8. Montgomery, D. Introduction to statistical quality control. John Wiley & Sons, Inc. 2nd edition, 1991
9. Box, G., and A. Lunceno. Statistical Control by Monitoring and Feedback Adjustment, Wiley Series in Probability and Statistics. John Wiley and Sons, Inc. 1997
10. Shewhart, W. Economic Control of Quality of Manufactured Product, New York, D. Van Nostrand company, inc., 1931.

# List of Tables

# List of Figures

TABLE 1 Comparison of segmentation methods

| | Characteristic | | | |
|---|---|---|---|---|
| Segmentation Method | Segmentation Criteria | Minimum number of segments | Final number of segments | Segment range |
| C-Chart | Standard deviation | One | Unlimited | Optional |
| CDA | Diversion from mean of entire profile | Two | Unlimited | Not specified |
| CART | Minimum sum of squared error | Two | Predetermined | Unlimited |
| ADA | Target range | One | Unlimited | Predetermined |

TABLE 2 Comparison of sum of squared errors (SSE) using three segmentation methods

| Segmentation Method | SSE [FN(40)] | Number of Segments |
|---|---|---|
| 2σ C-Chart | 264 | 19 |
| 3σ C-Chart | 331 | 11 |
| CDA | 521 | 11 |
| CART | 431 | 7 |

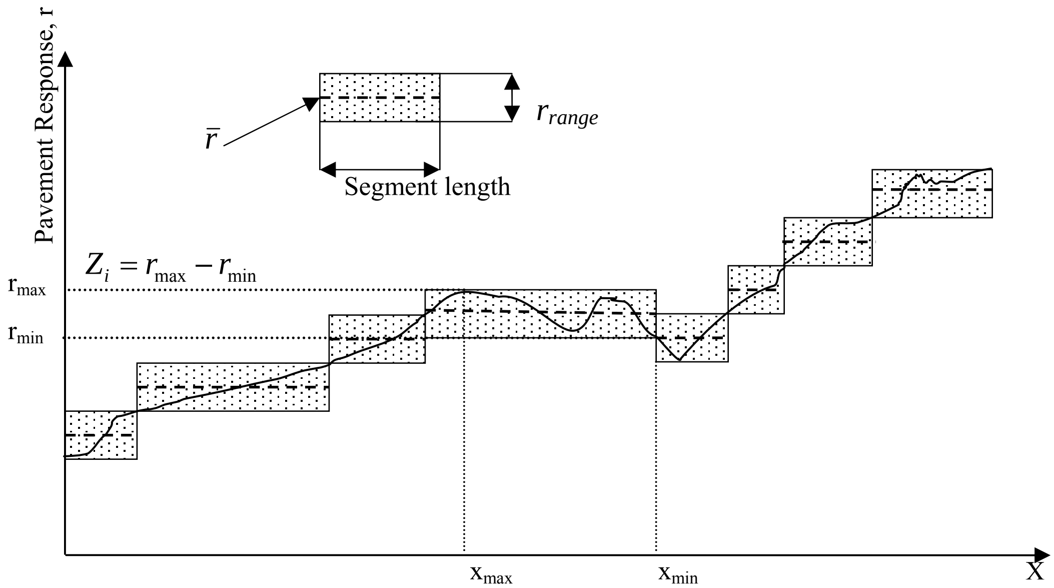


**FIGURE 1 The absolute difference approach based on user-selected response range**

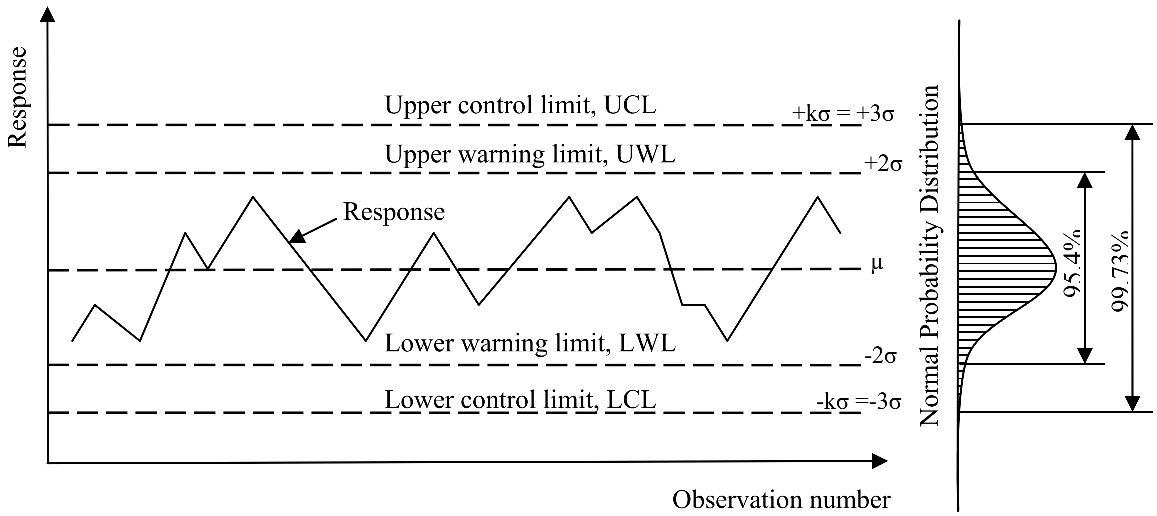


**FIGURE 2 Typical control chart showing warning limits (±2σ) and control limits (±3σ)**

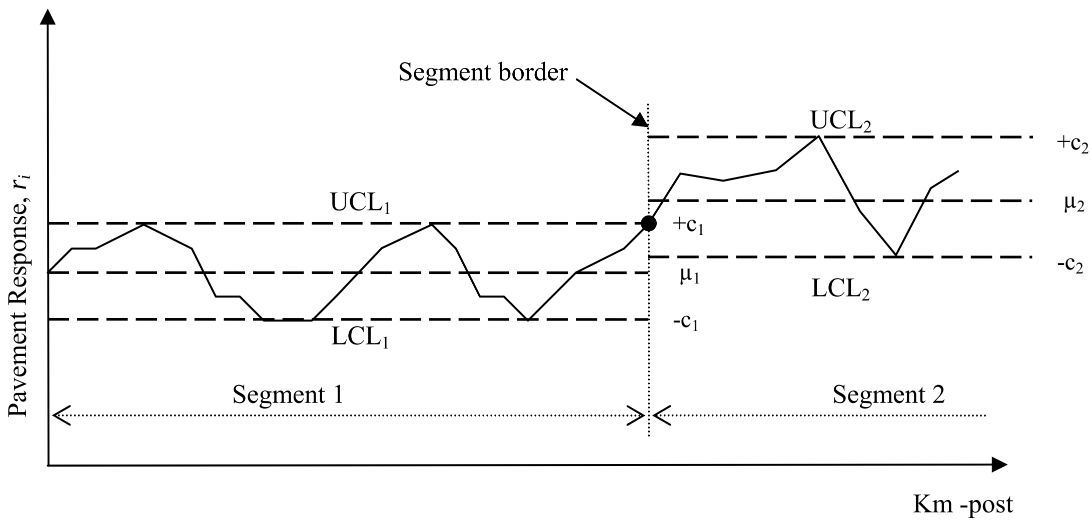


**FIGURE 3 Identification of homogeneous segments using c-chart approach**

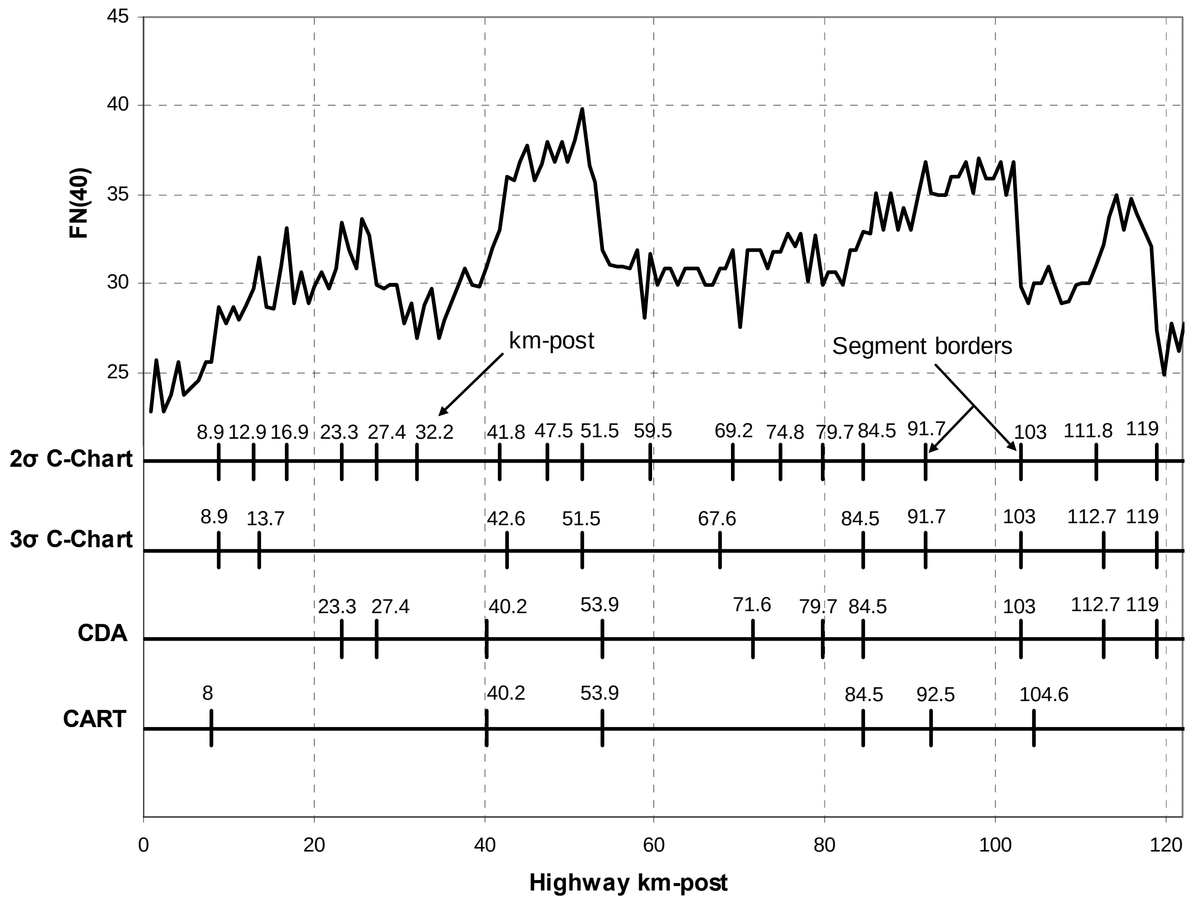


**FIGURE 4 Delineating a Friction Number profile using various methods**