ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia - Social and Behavioral Sciences 53 (2012) 1069 – 1079
Procedia
Social and Behavioral Sciences

SIIV - 5th International Congress - Sustainability of Road Infrastructures

Definition of Homogenous Sections in Road Pavement Measurements

Salvatore Cafiso[a*], Alessandro Di Graziano[a]

[a]*Department of Civil and Environmental Engineering, School of Engineering, University of Catania, V. Andrea Doria 6 -95125 Catania, Italy*

Abstract

Road measured data become useful information only if treated in order to match each section with statistically meaningful values. The problem can be led back to the detection of those change-points that divide measures in homogenous segments along the series of measured data. In this paper a methodology to detect a change point is proposed, searching those points that minimize the sum of the squared errors respect to the series of data. The MINSSE (MINimization Sum of Squared Error) has been, therefore, compared with the AASHTO Cumulative Difference Approach and a bayesian approach for the retrospective detection of change-points.

Keywords: Pavement, Survey, Measurements, Homogenous sections.

1. Introduction

In road pavement survey and evaluation, measures of skid resistance, unevenness and capacity bearing can be considered as random variables whose dispersion is to be attributed both to random errors of measurement and small heterogeneities of the surface. For this reason, it is appropriate to analyze data from surveys in order to identify homogeneous sections, for which it makes sense to define a mean value of the measure that also present significant differences compared to the averages of the sections before and after. The definition of homogeneous sections cannot be neglected, since only starting from this segmentation it is possible to use such information to determine the appropriate maintenance interventions [1,2,3]. Figure 1 shows, by way of example, as from a set of data (A) in function of the way these are aggregated (B or C) different evaluations about the need for maintenance interventions may be obtained. If you proceed with a single section analysis (solution B) you have to make the treatment even in the initial stretch where the pavement conditions do not need any intervention. This information is, however, highlighted by aggregating the data into two sections of analysis (solution C).

* Corresponding author. Tel.: + 39 095 738 2213; fax: + 39 095 738 7913.
E-mail address: dcafiso@dica.unict.it

1877-0428 © 2012 The Authors. Published by Elsevier Ltd. Selection and/or peer-review under responsibility of SIIV2012 Scientific Committee
Open access under CC BY-NC-ND license. doi:10.1016/j.sbspro.2012.09.956

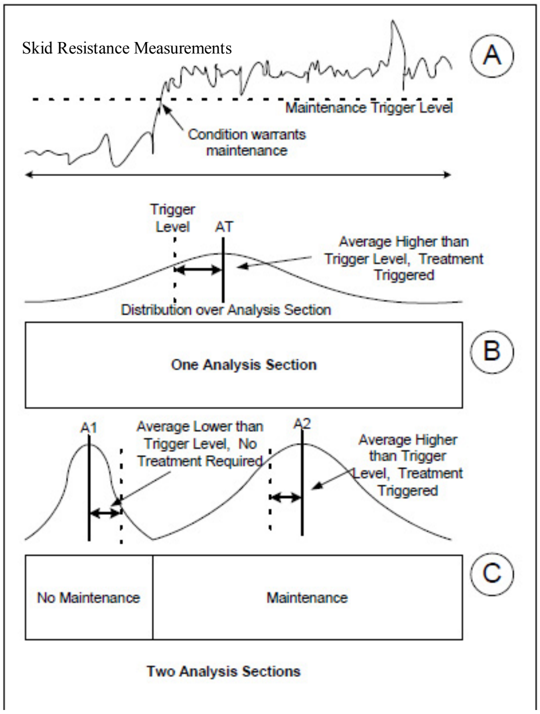


Fig. 1. Implications of sectioning [3]

Only since the early 80s approaches for the survey data treatment were defined based on rigorous statistical assumptions, implemented using simple mathematical models [4] or through more advanced logic [5,6]. Until then, the segmentation was performed according to the judgment of engineers by the way of simple rules of partition (fixed length of the segment) or with more articulate approaches but related to local pavement management systems and not always statistically reliable.

In modern literature on homogeneous segmentation [7,8,9], the most important systematic approaches are based on the "change-point" research. A change-point is a point which divides a set of observations, x_1, x_2, ..., x_n, into two groups, each of which can be described by a different statistical model. This difference can be associated with the change of a parameter, such as the average value, in a model otherwise unchanged. Before to discuss the proposed method to define homogeneous sections, it is useful to introduce other methods usually applied which will be used also as comparative results.

2. Method of Cumulative Differences

Cumulative Difference Approach (CDA) [4] can be illustrated by figure 2 using the initial assumptions of a continuous and constant response value (r_i) with three intervals along a project length L_s (Figure 2.a). The cumulative area A_x in a generic point x is simply the integral of the series along the distance:

$$A_x = \int_0^{x_1} r_1 dx + \int_{x_1}^{x} r_2 dx \quad (1)$$

Taking, instead, into consideration the average value of the measured characteristics:

$$\bar{r} = \frac{\int_0^{x_1} r_1 dx + \int_{x_1}^{x_2} r_2 dx + \int_{x_2}^{x_3} r_3 dx}{L_s} = \frac{A_T}{L_s} \tag{2}$$

the cumulative area media could be computed as:

$$\bar{A}_x = \int_0^x \bar{r} dx = \bar{r}x \tag{3}$$

In Figure 2.b the dashed line represents the cumulative area caused by the overall average project response.

The cumulative difference variable Z_x is simply the difference in cumulative area values ($Z_x = A_x - \bar{A}_x$) along the portion of measurement at a given x. Figure 2.c, plotting Z_x against distance x, allows to detect the position of the change-point (unit boundary) where the slope of Z_x function changes algebraic signs.

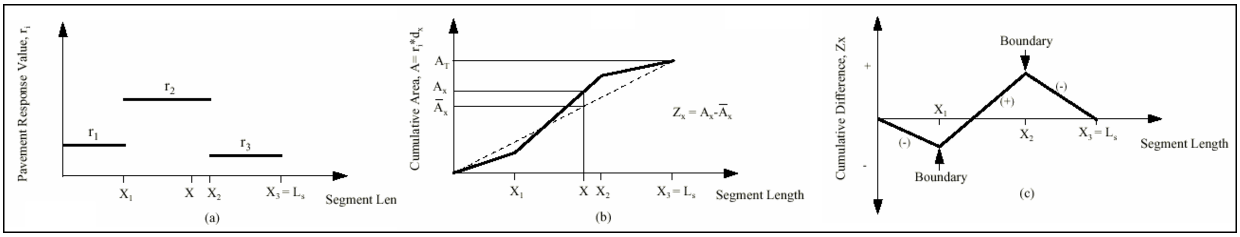


Fig 2. Cumulative Difference Approach [4]

The numerical approach proposed by AASHTO can be applied also in series of discontinuous measurements and never constant, which are actually the series of measurements of pavement characteristics, readily adapted to a computerized implementation. However, if the measurement series are not rather smooth with distinct changes in level, it was observed [10] that the method is highly sensitive to what exact part of the series is used for the analysis, for which even small changes of the points of beginning and end of the stretch in analysis determine changes in the definition of the homogeneous sections. To solve this problem and to introduce other needs in the maintenance process not considered in the classical CDA method, in the present work, the CDA was modified by adapting the methodology to an iterative process as reported in the flow chart of figure 3.

The new procedure introduces the concepts of minimum length L_{min} of a segment and of minimum difference between the mean values of adjacent segments Δmin, which represent two constraints necessary to make the segmentation produced by the CDA an effective support tool for the definition of homogeneous sections compatible with the maintenance interventions. While decreasing length, in fact, segmentation is better suited to the series of measures in the analysis but can reach values have little significance for the subsequent application of the maintenance. The minimum length of segments, generally, is set between 100 m and 300 m, based on considerations that take into account the costs of road agencies and users and the technical limitations imposed by the implementation of maintenance interventions. The definition of the minimum difference between the mean values of adjacent segments have to be referred to changes in the parameter under study which may affect the selection of the various maintenance operations. The procedure also provides for a statistical data check, through the application of the test t-student, so as to verify that differences in mean values of adjacent segments are statistically significant.

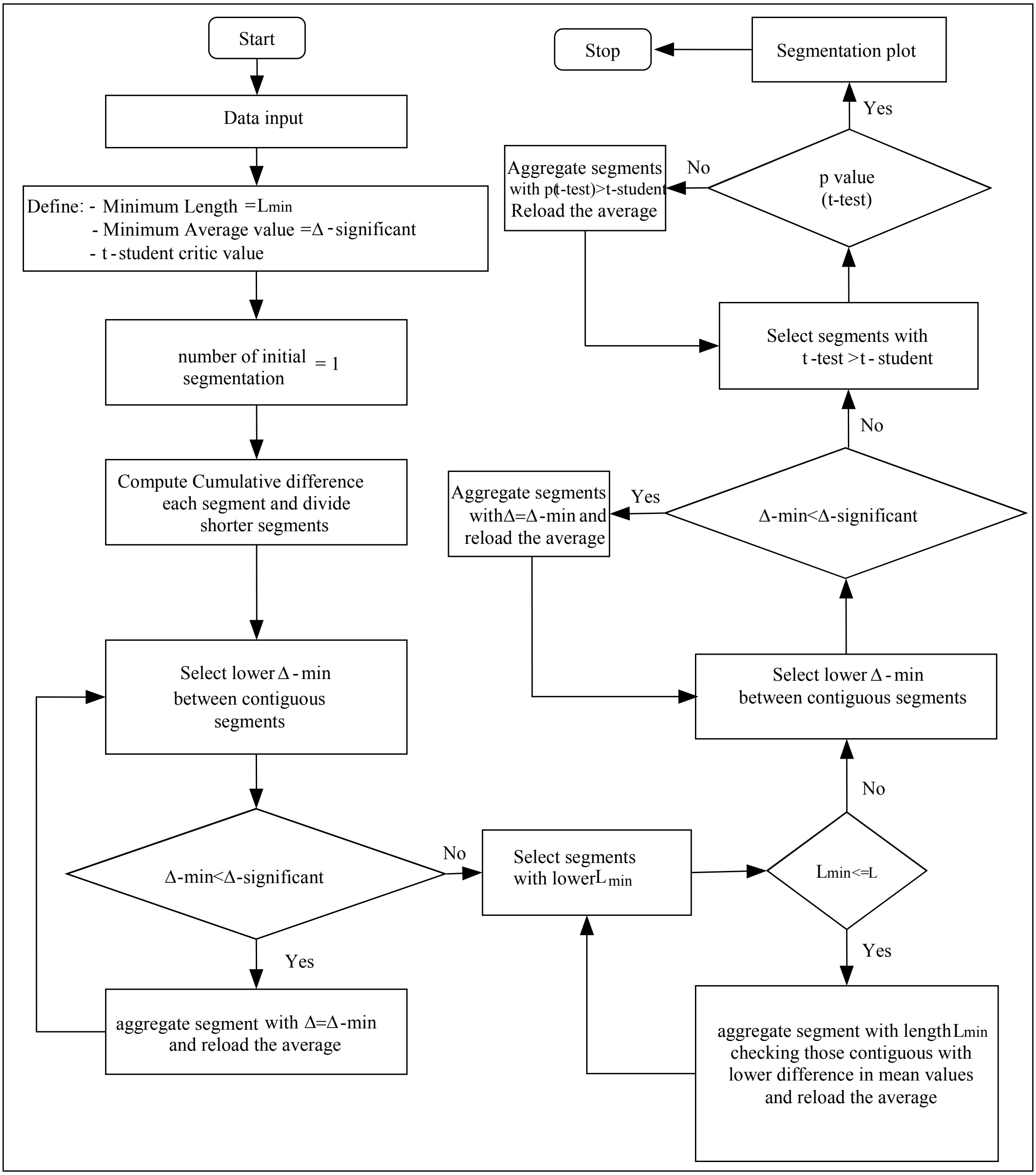


Fig. 3. Flow diagram for an application of the CDA

3. Bayesian method

With the bayesian approach [10], the measured parameter x_t is considered a random variable. In a formulation of the problem based on the retrospective research of a change-point, while $x_1, ..., x_n$ are assumed to be independent random variables distributed according to the following function:

$$x_t = \begin{cases} \theta_1(t) + \varepsilon_t & \text{per } t = 1,...,r, \\ \theta_2(t) + \varepsilon_t & \text{per } t = r+1,...,n, \end{cases} \tag{4}$$

where:
x_t: measured value;
$\theta(t)$: value assumed by the random variable θ in the measurement point t;
ε_t: measurement error in t (independent term);
r: position of the change-point.

If $p(\theta)$ is the prior distribution of the unknown quantities θ, $p(x|\theta)$ is the posterior distribution once x is observed and it can be obtained according to the following relationship, where $f(x|\theta)$ is the likelihood function:

$$p(x|\theta) \propto f(x|\theta) \times p(\theta) \tag{5}$$

The bayesian analysis of a change-point problem is based on the idea to compare the marginal posterior distribution $p(x|M_r)$ of different data sets θ_j with that obtained in the event of absence of change-point $p(x|M_0)$, in order to verify the one which best represents the real data.

The problem formulation is simplest to describe [11], when a sequence of random variables, $x_1, x_2, ..., x_n$, is partitioned by a change-point r ($1 \leq r < n$) into two subsequences $x_1, ..., x_r$ and $x_{r+1}, ..., x_n$, so that these subsequences are exchangeable, but the combined series is not. Denoting the model placing a change-point at r by M_r, one obtains the marginal posterior density of the observations, as:

$$p(x_1,...,x_n|M_r) = \iint \prod_{i=1}^{r} p_1(x_i|\theta_1) \prod_{i=r+1}^{n} p_2(x_i|\theta_2) \cdot p(\theta_1,\theta_2|M_r)\, d\theta_1\, d\theta_2 \tag{6}$$

where:
$p_1(x|\theta)$: posterior distribution with respect to the observed data in the first series x_i (i = 1, ... r);
$p_2(x|\theta)$: posterior distribution with respect to the observed data in the second series x_i (i = r +1, ... n);
$p(\theta 1,\theta 2|M_r)$: prior distribution for the model with change-point.

Denoting the model without change-point (r = n) as M_0, the marginal posterior distribution of the observations is:

$$p(x_1,...,x_n|M_0) = \int \prod_{i=1}^{n} p_1(x_i|\theta_1) \cdot p(\theta_1|M_0)\, d\theta_1 \tag{7}$$

where:
$p_1(x|\theta)$: posterior distribution with respect to the set of all observed data x_i (i = 1, ... n);
$p(\theta_1,|M_0)$: prior distribution for the model without change-point.

A comparison of these two models, with and without change point, may be based on the value assumed by the Bayes factor B_{r0}, referring to a judgment scale (Table 1):

$$B_{r0} = \frac{p(x_1,...,x_n|M_r)}{p(x_1,...,x_n|M_0)} \tag{8}$$

Table 1. Guidelines for interpreting Bayes factors [9]

B_{r0}	Evidence for M_r
<1	negative (data supports M_0)
1-3	not worth more than a bare mention
3-10	substantial
10-30	strong
30-99	very strong
>99	decisive

The model that includes a change-point r (M_r) and that one without change-point (M_0) are computed using an inference procedure, whose detailed discussion is beyond the scope of this presentation [9], based on the following statistical estimation for the measured values:

$$M_r \rightarrow \begin{cases} x_t = \alpha_1 + \beta_1 x_{t-1} + \varepsilon_t & \text{per} \quad t = 2,\ldots,r, \\ x_t = \alpha_2 + \beta_2 x_{t+1} + \varepsilon_t & \text{per} \quad t = r+1,\ldots,n-1. \end{cases} \tag{9}$$

or

$$M_0 \rightarrow x_t = \alpha_1 + \beta_1 x_{t-1} + \varepsilon_t, \quad t = 2,\ldots,n-1, \tag{10}$$

where α, β and ε_t are parameters characterizing the adopted measurement system (Figure 4).

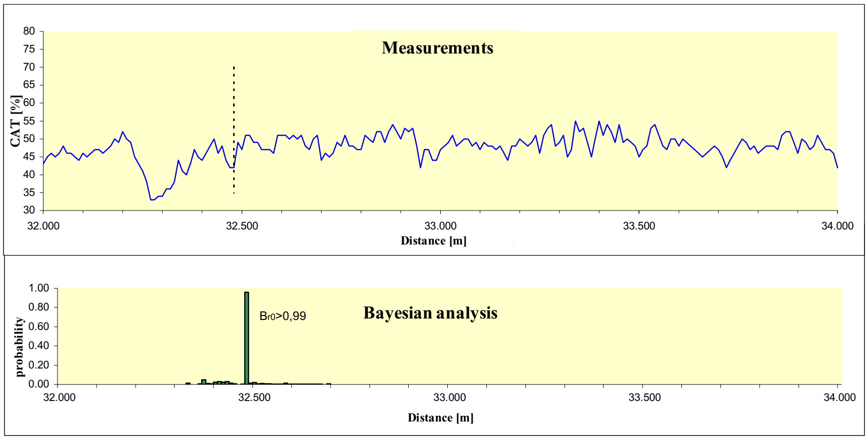


Fig. 4. Identification of a change-point using Bayesian analysis

The approach is analogous in the case of multiple change-points, so that, depending on the values of the Bayesian factors, the points where there is a greater probability of change of the measurements are determined. Although the procedure present a strong methodological approach, the complexity of the method makes it difficult to check L_{min} and Δ_{min}.

4. MINimization Sum of Squared Error (MINSSE) method

The aim of the minimum sum of squared errors method [12] is the research of those change-points that minimize the sum of squared errors defining homogeneous sections within parameters L_{min} and Δ_{min}. The series of n measurements may contain k (k<n) change-points $r_1, r_2, \ldots, r_k$. Considering all possible positions of k change-points the following set is obtained:

$$\kappa_k = \{(r_1, r_2, \ldots, r_k) : 0 < r_1 < r_2 < \ldots < r_k < n\} \tag{11}$$

Each element of the set determines a partition of the n measures into k +1 segments:

$$S_1 = \{x_1,\ldots,x_{r_1}\},\ S_2 = \{x_{r_1+1},\ldots,x_{r_2}\},\ \ldots,S_{k+1} = \{x_{r_k+1},\ldots,x_n\} \qquad (12)$$

with a mean value $\bar{x}_{S_1}, \bar{x}_{S_2}, \ldots, \bar{x}_{S_k+1}$. The sum of squared error of this partition is given by the following:

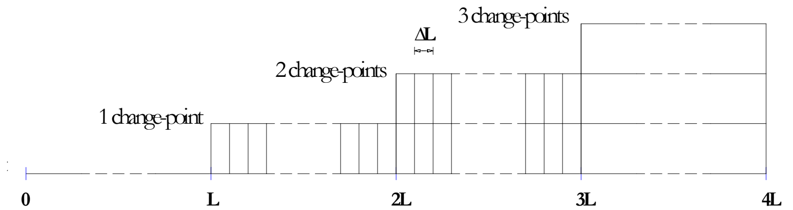
$$SSE_k = \sum_{j}^{k+1} \sum_{i \in S_j} \left(x_i - \bar{x}_{S_j} \right)^2 \qquad (13)$$

The method identifies, among all the segmentations obtained for k = 1, ... p-1, the one that minimizes SSE_k; where p represents the maximum number of homogeneous sections in the segment of length L:

$$p = \text{int}\,\frac{L}{L_{min}} \qquad (14)$$

Change points research is based on a continuous overlapping of the segments as represented in figure 5.

Once change points minimizing the SSE are identified, contiguous sections whose averages do not meet test t-student and adjacent segments with the difference of average values less than Δ_{min} (Figure 6) are aggregated.



Change-point	n° homogenous sections
0	1
1	21
2	66
3	1

Fig. 5. Combination of segmentations in a stretch 4Lmin, n = 40

As the method is strongly conditioned by the definition of the minimum difference between the mean values of adjacent segments, a procedure to support the engineer judgment in the choice of Δ_{min}, for each type of measured values has been implemented. Specifically, the ratio among the different values of SSE_i obtained with the variation of Δ and the SSE* obtained with the all set of data (one segment) permits to check the incidence of the choice of Δ_{min} in the segmentation process. In the example of figure 7, referring to Rut Depth, can be observed that only for values of Δ higher than 1.5 mm the ratio $\Sigma SSE_i/SSE^*$ increase in a significant way. Therefore, in the example a choice of about Δ_{min} =1.5 could be a good compromise between accuracy and precision in the segmentation results.

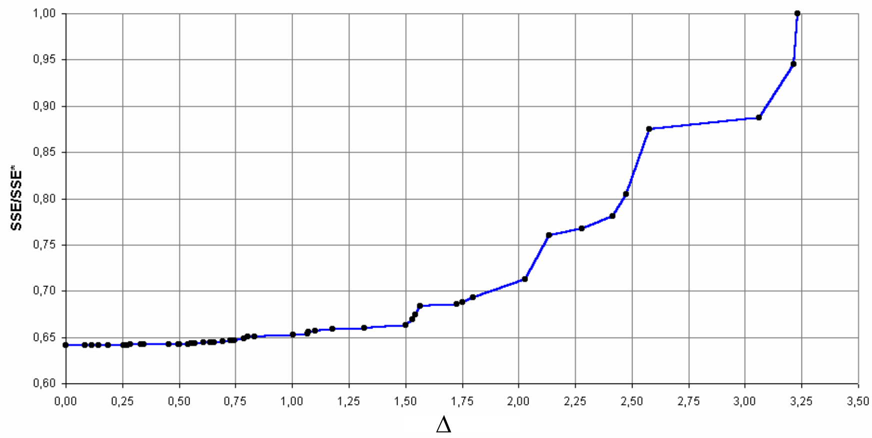


Fig. 7. Variation of the ratio ΣSSEi / SSE * depending on Δ.

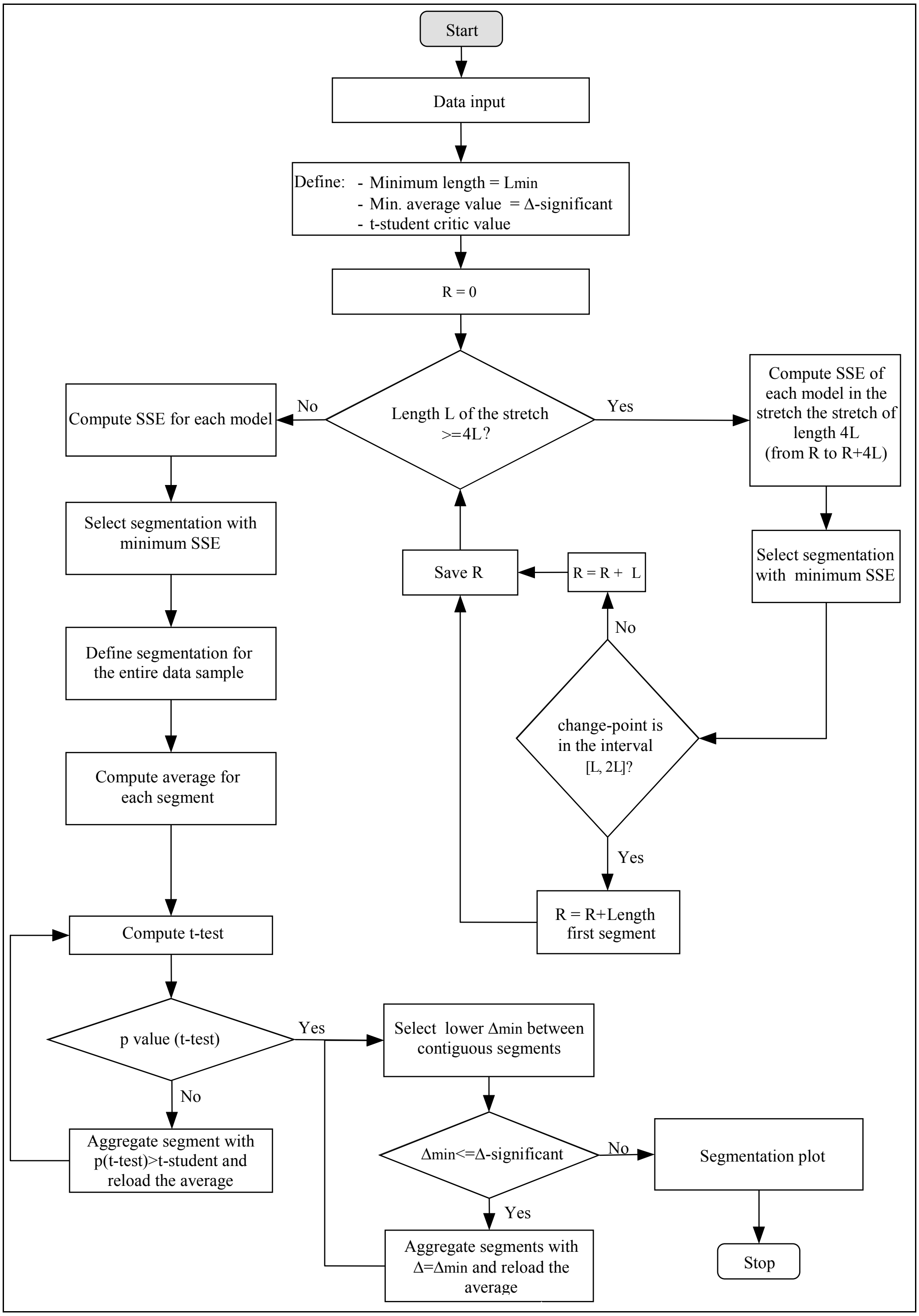


Fig. 6. Flowchart of MINSSE

5. Application

Methodologies above presented have been applied for the data analysis of a 20 km survey carried out on the highway SS417 in Italy. During the survey skid resistance (SCRIM), unevenness and rutting (ARAN) with a sampling interval of 10 m were measured. A comparison of different segmentations referred to data processing of skid resistance and rutting for only few kilometres of survey are reported in figure 8 and figure 9 respectively.

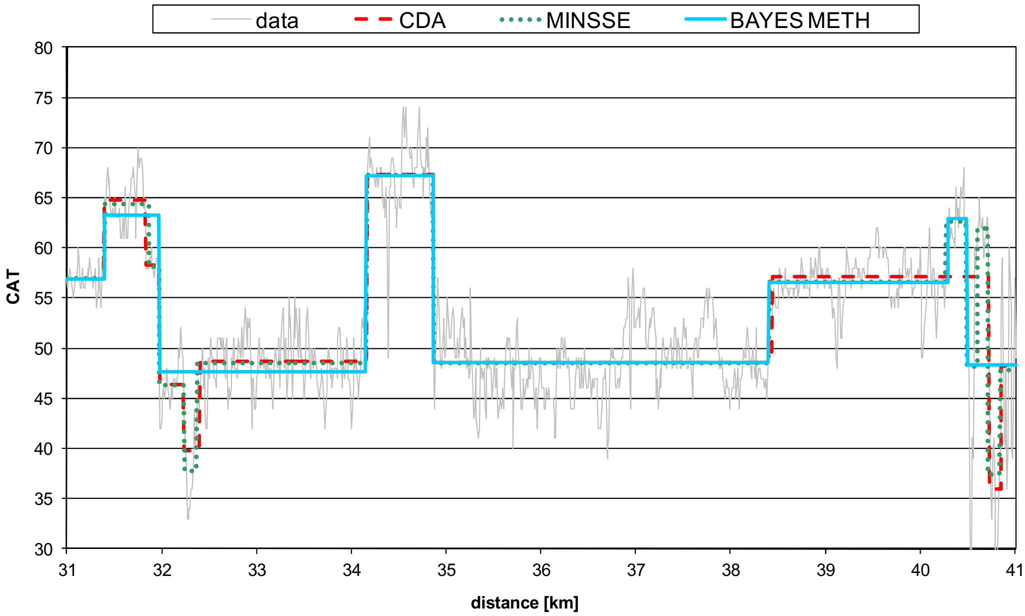


Fig. 8. Definition of homogeneous sections for the skid resistance data

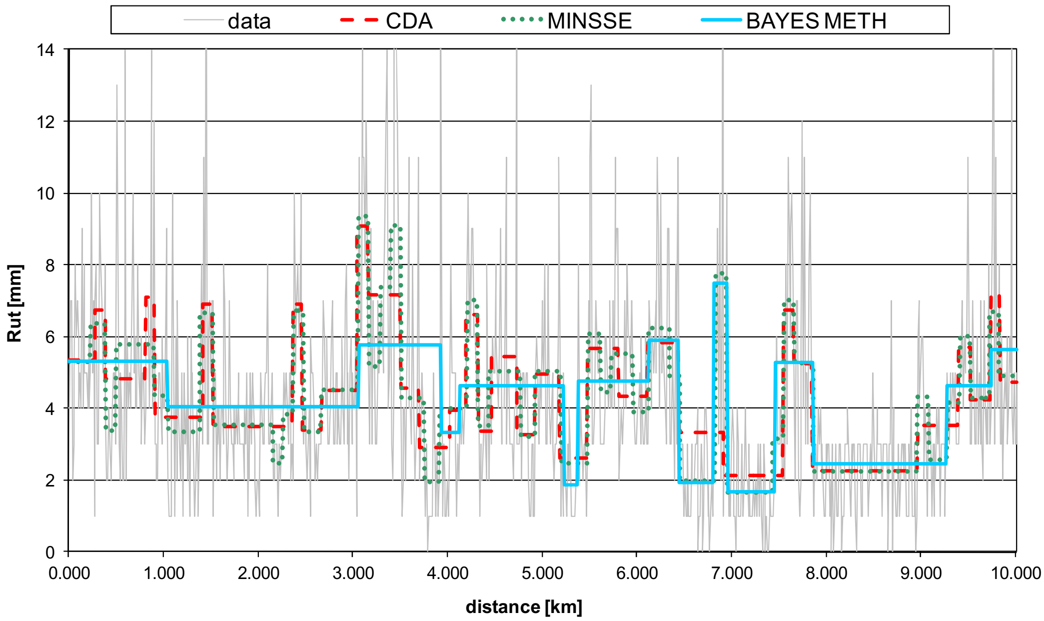


Fig. 9. Definition of homogeneous sections for the rutting data

The methods were applied as above described, by setting a minimum segmentation length of 100 meters and the following significant changes Δ:

- CAT [%] ± 5
- Ruth [mm] ± 1
- IRI [mm/ m] ± 0.5

CAT from Km 0+000 to Km 10+020

6	CDA	MINSSE	BAYES METH
CDA	19	14	8
MINSSE	-	22	10
BAYES METH	-	-	15

CAT from Km 31+000 to Km 41+020

4	CDA	MINSSE	BAYES METH
CDA	10	9	4
MINSSE	-	13	5
BAYES METH	-	-	7

IRI from Km 0+000 to Km 10+020

6	CDA	MINSSE	BAYES METH
CDA	28	19	7
MINSSE	-	40	8
BAYES METH	-	-	9

IRI from Km 31+000 to Km 41+020

3	CDA	MINSSE	BAYES METH
CDA	23	15	3
MINSSE	-	34	4
BAYES METH	-	-	4

Rut from Km 0+000 to Km 10+020

5	CDA	MINSSE	BAYES METH
CDA	33	28	5
MINSSE	-	44	10
BAYES METH	-	-	14

Rut from Km 31+000 to Km 41+020

7	CDA	MINSSE	BAYES METH
CDA	17	10	10
MINSSE	-	32	13
BAYES METH	-	-	23

Fig. 10. Comparison of the number of change-points for each methodology

Then, the three different methods above illustrated were compared quantitatively, as shown in figure 10: in each table the number of change-points returned by each methodology (cells along the diagonal), the number of change-points selected by all the three methods (cell in the upper left), the number of change-points selected by only two methods (other cells) are reported. The three methods provide similar results in the treatment of series which have less variations in the measured data (Figure 8), while significantly differ with series having a wider variation from the average value (Figure 9). Basing on the experimental results, the bayesian methodology returns a more macroscopic segmentation and the result is largely independent from the series of measures under study. The CDA returns a segmentation that in some places is unrepresentative of the detected data although characterized by a great number of change-point. The MINSSE returns a segmentation which follows in a microscopic way the series of data, so that it is more highly sensitive to the type of series. Many of the change-points found by the bayesian approach are almost always returned by this method, too.

6. Conclusions

The problem of searching in the series of measured data those change-points able to divide a long stretch of road into sections that can be considered homogeneous with respect to relevant aspects was analyzed by comparing three different methods based on different criteria of data processing.

In the bayesian method uncertainties about the existence and location of a possible change-point are reported in terms of probability, so that the analysis requires an accurate parameters definition for the identification of a change-point. The complexity of the method and the definition of the probability functions stiffen the application. However, the method is characterized by a rigorous statistical basis, which makes it an excellent method to check the reliability of the other methodologies.

From this point of view it was observed that the segmentation performed by the CDA fails to identify some significant variations, departing from the results provided by the bayesian approach.

With the new proposed method (MINSSE) the segmentation is often referred to the minimum length of the section, postponing further aggregation in function of the maintenance treatments. Moreover, the segmentation returns the majority of the change-points identified by the Bayesian method. From this point of view the MINSSE is extremely valuable as permits a good interpretation of the survey data and it is well suited to the maintenance programming requirements by the way of managing both the Δ_{min} and L_{min} .

References

[1] Cafiso S., Di Graziano A., Kerali H.R. and Odoki J.B. (2002) Multicriteria Analysis Method for Pavement Maintenance Management. *Transportation Research Record 1816*, Journal of the Transportation Research Board. Washington, DC, pp. 73 – 84

[2] Bird R., Cafiso S., Di Graziano A., Laganà G. (2000). Definizione della banca dati per la manutenzione delle pavimentazioni della rete stradale italiana. *Proceedings of 10th SIIV Conference* Catania

[3] Bennett C. (2004). Sectioning of road data for pavement management. *Proceedings of 6th International Conference on managing pavements* Brisbarne

[4] AASHTO (1993). AASHTO Guide for design of pavement structure. Appendix J: Analysis Unit Delineation By Cumulative Differences. *American Association of State Highway and Transportation Official*, Washington D.C.

[5] Andonuadis M., Altschaeffl A.G. and Chameau J.L. (1985). The Use of Fuzzy sets mathematics to assist pavement evaluation and management. *Final Report n° FHWA/IN/JHRP* n°14

[6] Camomilla G., Marchionna A. (1992). Procedura per il controllo delle caratteristiche superficiali e la programmazione della manutenzione. *Journal of Autostrade* vol. 3

[7] Misra R., Das A. (2003). Identification of homogeneous sections from road data. *International Journal Pavement Engineering* vol. 4, pp 229–233

[8] Ping W.V., Yang Z., Gan L., Dietrich B. (1999). Development of procedure for the automated segmentation of pavement rut data. *Transportation Research Record* 1655 Transportation Research Board, Washington, D.C.

[9] Thomas F. (2001). A Bayesian Approach to Retrospective Detection of a Change-point in Road Surface Measurements. *Doctoral Dissertation* Department of Statistics, Stockholm

[10] Thomas F. (2005). Automated Road Segmentation Using a Bayesian Algorithm. *Journal of Transportation Engineering* vol. 131 issue 8

[11] Smith A.F.M., Roberts G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society*, Series B n° 55

[12] Cafiso S., A. Di Graziano, F. Condorelli (2003). Definizione delle sezioni omogenee rei rilievi su pavimentazioni stradali. *Journal of Strade e Autostrade* vol. 5