Generating homogeneous road sections based on surface measurements: available methods

Reprint from the 2nd European Pavement and Asset Management Conference, 21-23 March 2004, Berlin, Germany.
(Paper no. 48)

Fridtjof Thomas
Generating homogeneous road sections based on surface measurements: available methods

(Paper no. 48)

Fridtjof Thomas
GENERATING HOMOGENEOUS ROAD SECTIONS
BASED ON SURFACE MEASUREMENTS:
AVAILABLE METHODS

Dr. Fridtjof Thomas
Swedish National Road and Transport Research Institute (VTI)
Box 760, SE-781 27 Borlänge, Sweden
Tel.: +46 243 73675 / Fax: +46 243 73671
E-mail: Fridtjof.Thomas@vti.se
http://www.vti.se/tek/fthomas

Abstract
Modern road profilers deliver long measurement series of road surface characteristics such as IRI or rut depth. These measurement series have to be “chopped up” into parts of similar values in order to summarize the information they contain and to generate road sections that are meaningfully handled as units in pavement management systems. Three algorithms for this purpose are described and compared with respect to methodological aspects as well as data requirements.

1 Introduction
Modern road profilers deliver long measurement series of road surface characteristics such as IRI-values (International Roughness Index values) or rut depth, where each single value represents a short part of the physical road. Most road administrations determine these parts to be between 10 m and 50 m long. This distance is the finest “resolution” of a road that can be handled in the associated pavement management system (PMS). Ideally, pavement management decisions should be based on the information about a road’s condition contained in these measurement series. Other information such as traffic volumes, the width of a road, speed limits and current type of pavement is important as well for creating road sections that are useful input units to a PMS.

Currently, road administrations experience the greatest need to identify so-called homogeneous sections when planning maintenance actions. Identifying candidate sections for maintenance in the near future is essentially a task of determining which parts of the measurement series exceed certain threshold values, and ensuring that these parts are not too short to be meaningful candidates for actions such as repaving.

However, the large amount of data collected by road profilers can be used for more than just identifying sections that fail some minimal requirements. Evaluating and comparing the information from different instances in these measurement series allows for a systematic monitoring of the road surfaces. This type of monitoring could, for example, provide insight as to when road deterioration accelerates, long before critical values are exceeded. Such an early warning would give a road engineer the opportunity to investigate a problem before severe damage with subsequent extensive maintenance requirements becomes a reality. Furthermore, a monitoring system could be used to evaluate whether past maintenance
actions were successful in solving the problem of concern. Also, cold-climate countries could identify road sections subjected to frost heave by measuring the affected road network repeatedly between the seasons.

Whether or not the measurement series are used for short- to medium-term maintenance planning or for a more ambitious monitoring of the road network in general, a prerequisite for most analyses is to identify the parts of the measurement series that are homogeneous with respect to a particular criterion. Typically, the sections so obtained will be compared to sections based on other criteria, and a final single partition of the road will be compiled for the particular problem under study. Which criteria to pay attention to and how best to combine different existing partitions of the road into a single one will depend on the particular problem at hand (see BENNETT (2004) for a description of typical steps).

Inventory data and manually obtained summary measures for the overall condition of relatively long road sections can exhibit identical values for a large number of consecutive values associated with the running length of a road. In contrast, the measurements from modern road profilers are rather volatile. Because adjacent measurements are essentially always different even when their overall level is unchanged, simple approaches like inserting a section border whenever a value is not exactly identical to the previous one do not provide a solution. Establishing upper and lower bounds of deviations that are tolerated without triggering a section border will mitigate the problem but not resolve it.

This paper describes three methods that explicitly address the problem of segmentation of measurement series obtained from modern road profilers. First, we will review the method recommended in AASHTO (1986) and discuss some extensions necessary to make it a fully automatic method suitable for the large amount of data that have to be processed by a road administration. Second, we will describe the method by RÜBENSAM and SCHULZE (1996) that first smooths a measurement series and then identifies section borders by evaluating absolute differences of neighboring values in this smoothed series. Third, we will outline the approach recently developed by THOMAS (2003, 2004) which views a part of a measurement series as homogeneous if that part can be described reasonably well by a single particular statistical process, and which inserts a section border whenever there is a change in the parameters governing that process. The three approaches are compared to each other. Some general remarks conclude the paper.

2 Available Methods for Road Segmentation

Throughout, we denote the measurements of a measurement series of length $n$ as $x_1, x_2, \ldots, x_n$ and assume that each $x_i, i = 1, \ldots, n$, represents an equally long stretch of the physical road.

2.1 Cumulative Differences

The cumulative difference approach advocated by the AASHTO-Guide (AASHTO 1986, Appendix J) compares the sequence of actual cumulative sums in a measurement series with the sums that would have resulted from adding averages. A series $z_1, z_2, \ldots, z_n$ is constructed by calculating

$$z_k = \sum_{i=1}^{k} x_i - k \bar{x}, \text{ for all } k = 1, \ldots, n, \text{ where } \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i,$$
According to AASHTO (1986), a section border is indicated whenever the “trend” in the series of cumulative differences changes from positive to negative or vice versa. However, relatively volatile measurement series like the ones typically obtained with modern road profilers need additional criteria to restrict the number of identified sections, since “non-substantial” changes in trend are very frequent in these series and should be ignored.

The section borders indicated in Figure 1 were obtained by finding “substantial” local peaks in the series of cumulative differences, where a peak is identified as “substantial” if its value is the most extreme within a window of seven neighbors to the left and seven neighbors to the right of it. Requiring more or fewer neighbors to be less extreme than the tentative peak respectively decreases or increases the number of identified sections.
This is a simple example of a rather arbitrary criterion that might be used to control the number of generated sections. However, since any series of cumulative differences starts and ends at zero and has at least one peak in between (unless all measurement values are identical), it is virtually true that at least two sections will be identified in any series unless further requirements prevent this.

Candidates for additional restrictions are a threshold for the minimal length of resulting sections and minimal differences of arithmetic mean values of resulting adjacent sections. PING et al. (1999) elaborate along these lines and also suggest using two-sided t-tests to assess the statistical significance of differences in mean values.

At least two conceptual problems are associated with this t-testing. First, a standard t-test requires the measurements to be statistically independent (conditionally on the common mean), something which is not the case for the measurement series of IRI, rutting and friction that has come to the author’s attention. These measurements typically exhibit pronounced first-order autocorrelation, thus violating the assumptions in a standard t-test. Secondly, applying t-tests repeatedly a large number of times in different parts of a long measurement series (or the entire road network!) plague such a procedure with problems of mass-significance. Be that as it may, PING et al. effectively accept all differences exceeding an exogenously given threshold as statistically significant by recommending a critical value for the test-statistic of exactly zero.

The cumulative difference approach in AASHTO (1986) is a special case of CUSUM-procedure (acronym of cumulative sum), widely used in fields such as industrial production control. The simplicity of calculating CUSUMs makes these procedures attractive. However, the substantial part of the analysis is the interpretation of the “behavior” of CUSUMs.

CUSUM-procedures in the statistical (quality control) literature are often designed for on-line monitoring situations, where measurements are subsequently taken and a decision has to be made whether, for example, a production process has to be stopped in order to recalibrate machines. This problem is conceptually different from the road segmentation challenge as the latter is a problem where all data before and after a suspected change are accessible (often addressed as an off-line problem). Furthermore, road segmentation intents to identify the exact location of a change in a measurement series, whereas on-line control procedures focus on the question if a change has occurred up to the most recent observation regardless where. As a matter of fact, the cumulative difference approach in AASHTO (1986) is closely related to a particular on-line monitoring CUSUM-procedure, but with the “target value” in the process replaced by the data-dependent arithmetic mean of the measurement series under study.

A successful implementation of the cumulative difference approach rests upon development of sensible criteria to interpret the calculated series of cumulative differences. It is likely that any criteria that work well for a wide range of measurement series will be data-dependent, for example, by explicitly accounting for the variability in a particular measurement series under study. Unfortunately, such dependencies on the data under study inevitably destroy the most attractive feature of CUSUM-procedures, namely the already mentioned simplicity in calculation (because the exact criteria have to be calculated from the data).

2.2 Absolute Differences in Sliding Mean Values

RÜBENSAM and SCHULZE (1996) suggest first smoothing a measurement series and then analyzing the smoothed series. From the measurement series $x_1, x_2, \ldots, x_n$, they construct a smoothed series by calculating the arithmetic mean of a measurement value and its $q$
Fig. 2: Approach using absolute differences of smoothed series

neighbors to the left and the right. Hence, such a window contains \( 2q + 1 \) measurements and “slides” along the measurement series. To be specific, they calculate

\[
y_i = \frac{1}{2q+1} \sum_{j=-q}^{q} x_j \text{ for all } i = q+1, \ldots, n-q,
\]

where \( q \) governs the degree of smoothing achieved by these sliding mean values. Thereafter, the absolute difference between neighbors \( 2d + 1 \) units apart is calculated as

\[
z_i = |y_{i-d} - y_{i+d}| \text{ for all } i = d + q + 1, \ldots, n - q - d.
\]

Note that the resulting series \( z_{d+q+1}, \ldots, z_{n-q-d} \) contains only \( n - 2(q + d) \) elements, and that section borders in the beginning or end of the measurement series cannot be detected by the method as it is. This is not a serious shortcoming when long measurement series are analyzed. Anyhow, in case of relatively short measurement series or extensive smoothing (large values for \( q \)), it might be worth defining the sliding mean values differently for the “end portions” of the measurement series in an attempt not to lose these parts for the analysis.

Section borders are then identified by the following technique: A threshold value \( Z_{\text{crit}} \) is selected that plays the role of a watchdog for the series of absolute differences. All parts of \( z_{d+q+1}, \ldots, z_{n-q-d} \) below \( Z_{\text{crit}} \) are ignored, and the positions of the maxima in the connected parts of \( z_{d+q+1}, \ldots, z_{n-q-d} \) exceeding \( Z_{\text{crit}} \) are taken to coincide with a section border in the original measurement series. RÜBENSAM and SCHULZE (1996) take these borders only as tentative and check for various requirements that have to be satisfied in addition. Also, this procedure is not capable of detecting gradual changes in the form of slight trends over a
longer stretch of the road; RÜBENSAM and SCHULZE (1996) check resulting sections explicitly for this situation in order to determine whether additional section borders have to be established.

Figure 2 displays the method when applied to the IRI-series from Figure 1 with the particular setting of \( q \approx d \approx 15 \) and \( Z_{\text{crit}} = 0.242 \). (Note that \( q \) and \( d \) do not have to be identical.) The positions of the five indicated section borders coincide with the positions of the maxima in the “hills” above the threshold value indicated by the vertical dashed line in the lower diagram in Figure 2. It is straightforward to derive the solution for other than the here shown threshold by shifting the horizontal line in the lower diagram accordingly and observing the consequences of such a shift. If the threshold \( Z_{\text{crit}} \) had been chosen slightly lower, a sixth section border between 161 and 204 would have been indicated.

Observe that even a slightly higher threshold would lead to an additional section border to the left of 101. This is so, because a slightly higher threshold would separate the multimodal “hill” (currently having one maximum at 101) into two “hills” with one maximum each. This is an important observation because it demonstrates that the number of section borders is not a decreasing function of the threshold, unless all “hills”, however small or large, are strictly unimodal. This unimodality can only be achieved by excessive smoothing.

Smoothing, however, disguises the information about sudden changes in a measurement series for the following reason. Since a suspected change in the measurement series is abrupt, the information about the location of that change is “clearest” by comparing its immediate neighbors. Averaging of measurements corrupts this “pure” information in the neighborhood of that location by mixing values from both sides of the suspected change. Therefore, smoothing a measurement series by a sliding mean might be expected to do more harm than good when the task is to identify the location of a sudden change.

2.3 Bayesian Segmentation Algorithm

A close study of the measurement series from modern road profilers for IRI and rutting reveals the following findings. The variability of individual measurements around a thought mean value is often close to the same size as the “jumps” between mean values of adjacent homogenous road sections. On the other hand, a fairly large number of measurements to both sides of a suspected section border are often available to facilitate the detection of relatively small shifts in mean value.

Furthermore, the variability itself is subject to often abrupt changes. Sometimes these changes coincide with changes in a series’ level, but this is not generally the case. Moreover, adjacent measurements are typically not statistically independent, but correlated. A closer look often reveals positive first-order autocorrelation of varying magnitude. Typically, the autocorrelation coefficient is somewhere between 0.2 and 0.7, but more extreme values are possible.

All these features reflect differences in the associated physical road surface and the measurement situation, which is itself to some degree a function of the road surface. Therefore, one might attempt to identify the parts of a measurement series that have been taken on similar parts of the road and under similar circumstances, where similarity is defined as similarity in level, degree of variability, and degree of autocorrelation in the measurement series. Whenever one or more of these criteria change, the measurement series is “chopped up”. The identified parts are then homogeneous in the sense that they can be described by a single first-order autoregressive process with constant parameters.
Fig. 3: Identification of a section border using a Bayesian approach

This is the underlying idea of the statistical method described in detail in THOMAS (2003). This method follows Bayesian principles of information processing, but is restricted to situations where it is known beforehand that at most one section border is reflected in the measurement series. THOMAS (2004) makes this method operational for arbitrary long measurement series with an arbitrary number of section borders. The necessary computations are more involved than those associated with methods described above, but by no means prohibitive. The reader is referred to THOMAS (2003, 2004) for the technical details.

Figure 3 shows the result of applying THOMAS (2003) to the IRI measurement series already analyzed in Figures 1 and 2. (The series is displayed on a logarithmic scale in the upper diagram.) The lower diagram shows in its right corner a probability of 0.95 that a single transition is present in the measurement series. Equivalently, the odds in favor of one transition as opposed to no transition are 19:1. This is strong evidence for a transition somewhere, but the location of that transition is not indicated by that probability (or odds ratio).

Information about the location of a transition is obtained from the probability plot in Figure 3 (lower diagram). In this particular case, only one location (161) is a candidate for such a transition. It is not unusual that several locations are plausible candidates for a single transition. The probabilities depicted as bars in Figure 3 are measures of the relative support of competing locations. The mode of this distribution is therefore an obvious candidate for selecting the location of a transition, since—by definition—the mode receives the highest support of all possible locations.
This distribution is conditional on the existence of one transition; the height of all bars will therefore always sum to unity by a basic property of probability measures. One has first to decide whether a transition is present or not, which is easily done with help of the probability Pr(one transition) depicted in the right corner of the lower diagram. A threshold for this probability is all that is needed to automate the procedure. This procedure is used iteratively in THOMAS (2004) in order to detect an arbitrary number of transitions in long measurement series.

Identified homogeneous sections may or may not differ in level from their neighboring sections, but they will differ in some respect. Figure 4 demonstrates this. The upper diagram shows IRI measurements from Austria; each observation represents 25 m of the pavement. The lower diagram displays the result rendered by the Bayesian segmentation algorithm. The section between location 230 and 254 (representing 600 m of the pavement) exhibits considerably less variability than the adjacent sections and is therefore identified as a separate section. Also, location 937 is identified as a section border, even though the difference of the associated mean values is small. The measurements to both sides of this location tend to be of different magnitude, and they are sufficient in number to establish this section border.

Some of these section borders will not be of interest when scheduling maintenance actions. It is a straightforward exercise to investigate differences, for example, in the levels once the section borders are determined. As already mentioned in the Introduction to this paper, sometimes only the level is of interest to the road engineer, while monitoring generally requires more features to be picked up and evaluated.

The method by THOMAS (2003, 2004) builds explicitly on a particular statistical model and evaluates the information in a given measurement series by assuming this model to hold. Even when the model assumptions do not hold exactly, the statistical process used is often a good approximation of more involved processes and will consequently render satisfying results. However, the algorithm should be expected to fail in case of seriously violated model assumptions. Measurement series containing excessive outliers might not be handled well. Missing values are not allowed, and parts of measurement series with successive identical values can cause computational problems (identical values imply zero variance and, as a consequence, division by zero whenever the variance is used to standardize the variability in a measurement series).

3 Comparison of Methods

3.1 Definition of Homogeneity

The Bayesian segmentation algorithm is built around statistical properties of measurement series that include, but are not limited to, the mean value. A part of a measurement series is defined to be homogeneous if it can be described by a single first-order autoregressive process.

The cumulative difference approach and the approach utilizing absolute differences based on sliding mean values are not specific in what criteria are used to define “homogeneity” in a measurement section. However, it appears to be the case that changes in level are the feature of interest, and that changes in variance and the like are not intended to be handled.
Fig. 4: Comparison of the three methods described
This is partly a consequence of different expectations for the algorithms. The latter two are intended to generate road sections as candidates for maintenance actions. Therefore, even changes in a measurement series’ level below any critical field value are not really of concern.

In contrast, the Bayesian segmentation algorithm is not intended to solve the problem of identifying road sections for maintenance planning per se, but is intended to “chop up” measurement series in parts that are for some reason different from their adjacent parts. If these differences are of practical concern for a specific task will need further investigation. This post-processing is a step of utilizing a given segmentation for a well defined practical purpose, and is not a necessity to overcome certain shortcomings in the basic approach.

3.2 Adaptation of the Algorithms

The measurement series from different countries may or may not “look differently”. Exact details depend not only on the type of measurement (IRI, maximum rut depth, average rut depth, friction, etc.) but also on the overall state of the road network, the measurement equipment used, the training of operators, the processing of the measurement series before they are stored in the database, and so forth. Despite that, one should not overemphasize these differences as many measurement series across countries, pavement types, and technical solutions in road profilers do exhibit similar structure.

Regardless, each of the algorithms described here needs some adaptation to the measurement series that are to be analyzed. AASHTO (1986) does not contain any details on how to evaluate the series of cumulative differences. Many criteria can be thought of and software is available that has a number of these criteria implemented. The absence of a well specified model for the measurement series makes it impossible to derive good criteria on theoretical grounds. One will either have to develop such a model for the measurement series at hand or will have to rely upon ad hoc criteria that appear to work.

The approach using absolute differences based on sliding mean values needs three decisions: the size of the sliding window for computing the means, the spacing of smoothed values used to calculate the absolute differences, and a threshold value that triggers the search for a maximum in the series of absolute differences. The first two aspects are conceptually simple. Somewhat more difficult to understand is the role of the threshold value, because there is no simple and intuitively sound relationship between the value of the threshold and the number of identified sections (see Section 2.2).

The Bayesian segmentation algorithm needs a decision of the threshold at which to insert a section border. This threshold is a regular probability, and the higher this value the stronger must be the support for a section border for it to be placed in the measurement series. In addition, the measurement series have to be checked for compliance with the model assumptions. It is likely that a transformation of all measurement series prior to the analysis is recommendable. THOMAS (2004) gives details about such transformations.

3.3 Information about Section Borders

All algorithms return essentially a list of section borders without further information. Only the approach described in THOMAS (2003) provides additional information about the relative support of competing locations of a section border.
3.4 Data Requirements

Modern road profilers most typically deliver measurements representing equidistant parts of a road. None of the methods should be very sensitive to values representing parts of slightly unequal length. AASHTO (1986) provides formulas for calculating cumulative differences in cases where measurement values represent sections of unequal length.

The Bayesian segmentation algorithm should be the one most sensitive to “doubtful” data quality. Measurement series containing missing data cannot be handled by the algorithm. Furthermore, extreme outliers and repeated measurements with identical values are potentially not handled well. This algorithm necessitates therefore a thoughtful approach to outlier detection and imputation of missing values before a measurement series is submitted to the algorithm. In any case, all algorithms benefit from a careful approach to these important aspects of data processing.

4 Conclusion

Many road administrations possess steadily growing databases containing a huge amount of information in form of surface measurements obtained by modern road profilers. Much effort is rightfully devoted to the quality assurance of the process of data collection and storage. These good data should form the basis for assessing the current state of the entire road network as well as the condition of individual roads.

Using such data for monitoring road surfaces and forming meaningful input units for road deterioration models will require a sensible approach to the method of segmenting these long measurement sections into parts that describe road sections that are homogeneous with respect to the criteria of interest. Three algorithms for this purpose have been described.

The environment in which road profilers operate is only partially controlled. Regardless which algorithm is chosen by a road administration, there will be situations where the implemented algorithm does not perform well. The most “corrupted” measurement series that can be found in the database should therefore not form the basis for an evaluation of the algorithms. Likewise, it is not a good idea to “calibrate” an algorithm using very simple (or even stylized) situations and to hope for that the algorithm will perform still acceptable when less well behaved measurement series are analyzed.

A good approach for evaluating the algorithms is to retrieve a reasonably large number of measurement series from the database and to split these randomly into three groups. Use the first group to calibrate the algorithms in any preferred way. Then apply the calibrated algorithms to the measurement series in the second group and observe how well they perform. Improve the algorithms and check them finally against the measurement series in the third group. Keep in mind that the algorithm is supposed to perform reasonably well for measurement series you have never seen before and will possibly never ever look at yourself!

In a broader perspective, it is not self-evident that road segmentation always has to be based on the same measurement series that describe the criterion of interest. Being precise with the meaning of words, IRI and rut depth are not measured, but calculated from various signals typically obtained from laser units, accelerometers, gyrocompasses, and so forth. ROUILLARD et al. (2000) analyze, for example, signals from a road profiler that are much closer to the “raw signals” of the laser units than the IRI (they work essentially with narrowly spaced measures of elevation changes of a pavement). It might be easier to identify the locations of particularly pronounced elevation changes in these “more direct” measurements,
and to segment a road based on the nature of a sequence of these elevation changes. Once sections with similar structure in these elevation changes are identified, the IRI might be calculated for these identified sections. Nevertheless, a suitable segmentation algorithm for the analysis of the “raw data” is still needed.

REFERENCES


