INSTABILITY DETECTION FOR HIGH-SPEED WEIGH-IN-MOTION SYSTEMS TOWARDS DIRECT ENFORCEMENT

Loïc Warscotte[1], Jehan Boreux[2]
[1] *University of Namur, Belgium*
[2] *Public Walloon Service, Belgium*

Abstract

High Speed Weigh-In-Motion (HS-WIM) systems are used to automatically enforce weight restrictions on heavy vehicles operating within specified maximum weight limits. In Belgium, these systems have recently been implemented into the road surfaces of highways at several locations. They accurately estimate the weight of semi-trailers and vans moving at high speeds, meeting Class I standards according to the E1318-94 criterion. Despite the initial precision in weight measurement, it is crucial for the system to maintain calibration for reliable estimations over an extended period. Currently, periodic tests of estimation errors are conducted, and the parameters of the system are calibrated when necessary. The main objective of this research is to develop a complete automatic monitoring of this HS-WIM system, operating without the need for constant supervision. In the field of changepoint detection methods, papers have laid the theoretical background for developing such monitoring, assuming strong mathematical characteristics on the nature of the analyzed data and the changes. More recently, a breakthrough research on locally stationary time series extended the previous theory avoiding strong assumptions on the dataset. However, extensive calculations are required to perform these changepoint detections. Despite lower bounds of thresholds are suggested, no exhaustive solution is provided in the initial research. This paper then aims to present the complete practical implementation of unsupervised monitoring used for the HS-WIM. First, it especially addresses the issue of natural dependence between the HS-WIM data and external variables to the system by performing a pre-processing stage. Second, we propose parametrized approximations of the theoretical background thanks to characteristics of the HS-WIM data. These parameters are user-defined, considering the available time and memory capabilities. Third, we show that we can enhance the sensitivity of the monitoring by using our procedure compared to the usual parametric method.

Keywords: changepoint, weigh-in-motion, control-chart, non-parametric

1 Introduction

In Belgium, there has been a growing interest in High Speed Weight-In-Motion (HS-WIM) towards direct enforcement over the past several years. This system, made of piezo-quartz sensors, was installed in the road surface of several highways to enforce both the total weights and the single axle weights for the semi-trailers and vans. After a renewal of the pavement at the specific location of Louvain-La-Neuve, the class of the site was considered as excellent according to the OIML criterion, as detailed in [1]. The accuracy of the HS-WIM system according to the type of the vehicle is described in [7] at this specific location.

DOI: https://doi.org/10.5592/CO/CETRA.2024.1658

In addition to assessing the current accuracy of the WIM system, there is a need to ensure that both the electrical response of the sensors and the computational part of the weighing remain consistent of time. For example, the aging of the sensors, the deformation of the road surface or even a component breakdown could lead to disturb the estimation of the weight. In this situation, we then have to calibrate the parameters or repair the device, if necessary. Currently, checking the validity of weighing is performed manually during periodic testing. During these tests, the weighing errors of randomly selected vehicles are calculated, by comparing the weights estimation of the HS-WIM and a more accurate low speed WIM. This process of individually weighing vehicles twice is time consuming and not well suited.
To face this problem, we propose an unsupervised tracking of the HS-WIM device over time. We especially analyze the long-run stochastic process $X_{t,T}$ defined in Eq. 1.

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \epsilon\left(\frac{t}{T}\right) \tag{1}$$

The stochastic process $X_{t,T}$ is seen as a random variable at each time step t, from 1 to T. T is also known as the length of the process. To remain consistent with the literature of change-point detection, we use the rescaled time line $u = t/T$. We express $X_{t,T}$ as the mean μ of the process, in addition to ε which denotes the individual random weight deviation of each vehicle from the mean μ.
In this work, we focus on the problem of monitoring the consistency of the mean μ over time. To this end, we define an alarm $D_T(u)$ in Eq. 2, which must turn on if the mean of weight estimation $\mathbb{E}[X_{t,T}]$ starts to deviate, and stay turned off if no problem is detected. This alarm is typically easy to compute from the events of $X_{t,T}$ at each time step before T.

$$D_T(u) < \tau \text{ if } u \leq u_0 \\ \geq \tau \text{ otherwise} \tag{2}$$

In the equation above, the changepoint u_0 is then the critical time step from which μ is no longer consistent. The threshold τ turns the alarm on and has to be estimated by our method, explained hereafter.
Typical difficulties arise from the analysis of this stochastic process $X_{t,T}$. First, the reseach on electrical response of piezo materials sugggest that the weighing process is not consistent regarding temperature and speed fluctuations. This leads to unstability that does not have to be considered in the monitoring. Second, the Probability Density Function (PDF) of ε can be not related to a specific probability distribution, and even not consistent over time. These very general assumptions make the usual tools as the parametric Cumulative Sum (CuSum) very sensitive to false positives.
More recently, research on locally stationary processes gains attention in the literature in [3] and [5]. They provide strong theoretical properties with poor assumptions on the studied time series. However, the calculations become burdensome concerning long-run signals, which span over one hundred time steps. The difficult part is typically to compute a good value of τ without strong assumptions on ε. Computing a lower bound of τ by discarding some computation steps was initially proposed in [6]. They empirically show that this approximation gives a good estimation in practice for artificial stochastic processes containing a changepoint. However, this could be a crude choice for real data which does not contain any changepoint, or small acceptable variations regarding external conditions. This work then aims to provide a computational approximation of the theory concerning the changepoint detection of locally stationary processes. The following development is organized in three parts.

First, we present how we apply the transformation of the initial time series computed by the device to reduce the impact of external variables on the time series of interest. Second, we introduce the CuSum principle used in both parametric and non-parametric method. Third, we present the improved computational method that is based on the locally stationary theory, which computes an estimation of τ that is robust against false positives detection.

2 Pre-processing of the time series

The first stage involves selecting a time series for the changepoint detection. To justify our choice, we present several features of semi-trailers in Fig. 1, regarding the hour of the day. These features, as well as the following results, are computed thanks to a dataset collected from October 2022 to July 2023 containing hundreds of thousands trucks. We especially observe that the total weight is not consistent along the day. While the first axle weight remains relatively consistent, variations in total weight can be attributed to changes in semi-trailer loads. Additionally, we note a 30% mean difference in temperature between night and day. The small difference for the speed can be explained with the difference of loads.
From these observations, the time series that we decide to monitor is the first axle weight of the semi-trailers. Two clusters of semi-trailers are present in our database. The first one is a cluster of light vehicles, for which the loading increases the first axle weight, while the second cluster contains semi-trailers for which the loadings does not critically affect the first axles weight. We find that the first cluster, due to its variability in total weight depending on the hour of the day, is not suitable for monitoring purposes. Therefore, we opt to focus only on the heavier semi-trailers. To ensure consistency, we set the total weight lower limit at 35.1e+3 kg.

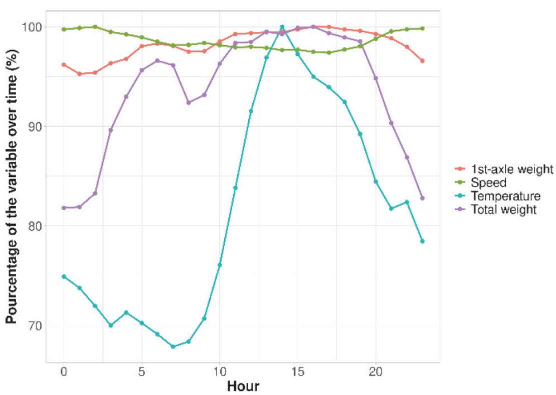

Figure 1 Mean value of features regarding the hour of day for the semi-trailers

Moreover, additional criteria based on the temperature and speed are operated, due to the influence of these external factors on the weight estimation [2]. The specific temperature range from 10°C to 15°C is considered for the monitoring. Despite the relative consistency of the speed, we include a speed range selection to take into account the possible modifications in the speed limits in the future.
The last modification is the mean operation of B consecutive selected first axle weights. This process has been shown to reduce the autocorrelation between time steps and provide a shrinking of the initial time series. This stage allows us to monitor smaller data sets which can be advantageous for detection methods with high complexity. In this work, we will use $B = 3$ to obtain a time series of length T around 2000.

The summary of the pre-processing stage and some descriptive statistics of the signal over time are available in Table 1. The time series values of consecutive time steps were gathered in 10 groups, in order to compute standard statistics in each group separately. This time grouping allows us to imitate the knowledge of the exact stochastic process over time, and verify the (un)consistency of these quantities. Each group contains approximately two hundred consecutive values of first axle weights, in which the mean, the median and the standard deviation were estimated.
First, we especially observe that the mean and the median are more consistent than the standard deviation, relatively compared to their respective statistics computed over all groups. The maximum relative difference is far below 1% for the mean and median, while it can reach 5% in the groups 2 and 7 for the standard deviation. Second, we notice a slight difference between the mean and the median for each group, even if relative consistency is observable across the groups, for each statistics independently. These observations encourage us to rather monitor the mean in order to reduce the number of false alarms.

Table 1 Descriptive statistics of $X_{t,T}$ after pre-processing gathered in 10 groups over time, with T = 2109

Group	Number	**Mean** [kg]	**Median** [kg]	**Standard deviation** [kg]
1	210	7288	7289	261
2	210	7294	7266	279
3	210	7290	7295	267
4	210	7280	7275	252
5	210	7275	7295	277
6	210	7282	7273	260
7	210	7249	7235	279
8	210	7264	7269	266
9	210	7307	7318	260
10	219	7280	7281	240

3 Parametric cumulative sum

This section is dedicated to the application of the standard Cumulative Sum (CuSum) Principle [4] to our database. The assumptions that has to be met for this tool are the Independent and Identically Distributed (IID) and the Gaussianity of ε at each time step. Regarding the quantities in Table 1, we can already expect a lot of false positives, due to the difficulty of the time series to verify these assumptions. The practical tool is defined in the Eq. (3).

$$\begin{aligned} s_t^+ &= \max\left\{0, X_{t,T} - \mu_0 - \frac{\delta\sigma_0}{2} + s_{t-1}^+\right\} \\ s_t^- &= \min\left\{0, X_{t,T} - \mu_0 + \frac{\delta\sigma_0}{2} + s_{t-1}^-\right\} \end{aligned} \quad (3)$$

In these equations, s_t^+ (resp. s_t^-) denotes the positive (resp. negative) accumulation value over the time steps t, which is used to detect positive (resp. negative) changes in the mean. σ_0 denotes the standard deviation of these Gaussian variables ε. δ is a value fixed by the user, and stands for the number of σ_0 from which the mean can deviate, without being considered as a changepoint. It is conventionally accepted that δ is fixed at 1.

The procedure in Eq. (3) especially consists of testing the null hypothesis H_0 that $\mu(t/T) = \mu_0$ is consistent for all time steps, against the alternative hypothesis H_1 that there exists a changepoint t_0 after which the mean starts to deviate, such that $\mu(t/T) = \mu_0 + \delta\sigma_0$. Accumulations start at the artificial time step t = 0 with: $s^+_0 = s^-_0 = 0$. Since the value of μ_0 and σ_0 of the underlying process are obviously not known, they can be estimated with the testing data set, or during a training phase.
The accumulation values and the estimated changepoints are depicted in the Fig. 2. For this method, the condition that turns on the alarm for the changepoint is the overrun of the quantity $\tau = \pm 5\sigma_0$ by at least one of the two accumulation values. We observe that, even if we know that the device remains under control during this period, the CuSum detects several time steps for which the alternative hypothesis of deviation is not rejected. The false positives can be explained by the strong properties assumed by the method. In what follows, we introduced the extended version of this CuSum principle and how we can use it for our purpose.

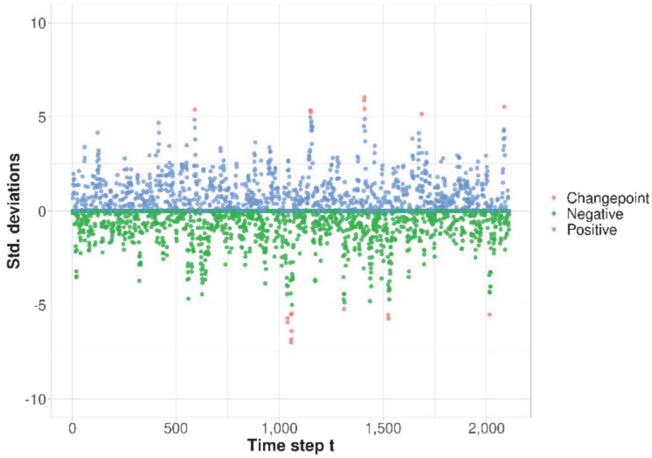

Figure 2 CuSum chart of the first axle weight time series with T = 2109

4 Non-parametric cumulative sum

4.1 Theory and assumptions

The extended theory to the non-parametric case presented in [6] assumes that the underlying process is locally stationary which means that, at any rescaled time step u, the locally stationary process of concern can be approximated by a stationary stochastic process. The $\varepsilon(t/T)$ can then have an unkown PDF and even not be stationary. We especially assume that the noise is a continuous and derivable mapping f from a uniform stationary process U_t. This a very general assumption and well-known PDFs such as Gaussian, exponential or Student-t actually verify this assumption.
In the following, we assume that the stochastic process of the first axle weight does not have to be neither IID nor Gaussian, but generated by the class of models: $X = \mu(t/T) + f(t/T, U_t)$. The breakthrough research in [6] shows that a good value of the threshold τ can be estimated from the quantiles of G(u), the maximum of absolute value of correlated zero-mean Gaussian variables, defined in Eq. (4).

$$G(u) = \max_{w \le v \le u}\left\{\left|H(v,w)\right|\right\}$$
$$\text{with cov. matrix entries } C_{(v,w),(v',w')} = Cov\left[H(v,w), H(v',w')\right] \tag{4}$$

where each w, v and u are rescaled time steps and H(v,w) is the CuSum value between time steps v and w. Therefore, the more is the proximity between v and v' or w and w', the more is the covariance between H(v, w) and H(v', w'). The covariance function Cov is known and detailed in [6]. C is then the matrix of covariance, for which the element C_{ij} in the i^{th} row and j^{th} column identifies i to (v, w) and j to (v', w').
As detailed in the aformentioned paper, it is computationally hard for long-run time series to obtain the PDF of G, due to the requirement of decomposing the non-sparse matrix $[C_{ij}]$ with $T^4/4$ entries. As it was proposed, a lower bound could be found by generating only few Gaussian variables. It provides a lower bound for the true value of τ, which is easily computed because of the reduced covariance matrix C'. Authors of [6] show good estimations of the changepoint u_0, for some types of time series which contain a changepoint. However, we will show that removing variables leads to increase the number of false positives for more complex stochastic processes, such as the HS-WIM time series.

4.2 Contributions

To increase the robustness of the approximation, we present how we modify the previous approximation of τ to obtain a less crude estimation of the threshold. Our computational method is based on a more general procedure to decrease the number of entries in C, without decreasing too much the number of Gaussian variables in G(u).
The first step of our method is refining the selection of H(v,w)'s that are considered in Eq. (4). Since we want to compute the maximum of Gaussian variables, removing those one with too big variance leads to a very poor approximation of the last quantiles. By inspecting the definition of the covariance function, we are not uniformly selecting the values of v and w, but focusing the selection of H(v, w)'s when v is sufficiently high, and w is far enough from 0 and v. Afterwards, we uniformly select equispaced values of v and w in the remaining subset of Gaussian variables, due to the high covariance between H(v, w)'s for which the values of w's and v's are close. The second step applied to reduce the computational demand is increasing the sparsity of the matrix C, without removing the corresponding generated random variables. While removing variables from G(u) does not affect the characteristics of a covariance matrix, discarding non-diagonal elements modifies its eigenvalues and does not ensure that the modified matrix remains a covariance matrix. The conservative reduction that we applied involves a block-diagonal selection and zeroing off-diagonal elements within blocks, as described in the eigenvalues decomposition in Eq. (5), with a simple example of 2 blocks C_1 and C_2.

$$C' \equiv \begin{pmatrix} C_1 & 0 \\ 0 & C_2 \end{pmatrix} = \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix} \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} V_1 & 0 \\ 0 & V_2 \end{pmatrix}^T \tag{5}$$

For this specific approximation of C by C', if C_1 and C_2 are covariance matrices, then C' also inherits the property of positiveness. V_k and Λ_k are respectively the matrices of eigenvectors stacked by columns and the diagonal of eigenvalues of the corresponding covariance matrix C_k. Due to the structure of C', the eigenvalues of each C_k can be independently computed, that fastens the calculations and saves Random Access Memory.
Additionally, these blocks have to be carefully selected in order to not disrupt too much the initial matrix C. To this end, we especially remove covariances for which values of w's and v's are too far. In practice, the number of block diagonals must be much bigger than 2 in order to reduce properly the memory complexity. All these subsets are computed according to defined-user parameters, which had to be fixed regarding the available resources.

Consequently, the key-point of our method is getting a balance between the number of generated random variables and the number of kept off-diagonal entries. The reduction of off-diagonal entries provides a sparse covariance matrix, that can be easily stored as a Compressed Sparse Row (CSR) matrix. We then test our method Sparse-cov compare to the basic Uniform method with the approximated τ with 5% risk, for the approximation after one step (τ_0) or after two steps (τ). The corresponding two steps are presented in [6] and the slight difference between the two estimations is not discussed in this work.
The Fig. 3 depicts the extended CuSum in green, easily computed from the first axle weights given by the HS-WIM system. We notice that the first detected false positive appears after a long period, in comparison with the parametric CuSum chart in the Fig. 2. We additionally observe the different approximation of τ, and the difficulty of the most basic method to generate a robust threshold for a long-run time series. We emphasize the fact that the parameters of the methods were fixed to require the same memory space.

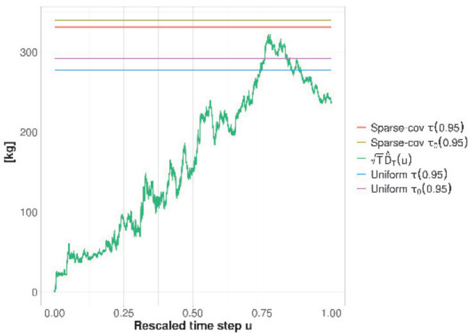

Figure 3 Non-parametric CuSum for the HS-WIM time series, with T = 2000

5 Conclusions

Numerous variables can have an impact on the weight estimation. We showed that reducing the effective range of temperatures, speeds and weights leads to a more stable time series. Despite this selection, parametric assumptions are difficult to meet in practice for the HS-WIM time series. This could be explained by the non specific PDF of the first axle weight and the non-stationarity. For example, we especially noticed unconsistency of the standard deviation over time. In this work, we assume that the first axle weight belongs to the family of locally stationary processes, and we apply a non-parametric detection test to detect deviations in the mean of first axle weights. This test typically requires to generate a lot of correlated variables which quickly leads to a covariance matrix C impossible to handle.
Therefore, selecting a subset of equispaced time steps is commonly used to overcome the computational demand. However, we especially show that, for a long-run and general time series such as the HS-WIM signal, performing such a selection drastically decreases the quality of the threshold estimation, implying a higher false positives detection rate. To overtake this issue, we opt for another procedure to make the covariance matrix lighter while not reducing too much the total amount of variance. This procedure is based on increasing the sparsity of C, without removing its diagonal entries. This leads to a C' which is block diagonal, and which keeps the properties of a covariance matrix.
By using the same amount of memory space, we conclude that our computational method is more suited for the monitoring of our stochastic process, due to its ability to properly compute more conservative bounds for τ.

The HS-WIM system can then be monitored using the first axle weight time series, after applying the pre-processing. Regarding the properties of the HS-WIM time series of length $T = 2000$, the critical threshold that has to be used in order to detect unstability is 350 kg.
In this work, the developed procedure was validated in the scope of HS-WIM data. However, this type of changepoint detection is valuable for every electronic system that has to be monitored, and which does not rely on a specific probability law. Moreover, detecting changepoint does not only involve breakdown of electronic components, but also detecting anomalies or changes in driver's behaviour concerning weight, but also speed for enforcement or for the forecasting of traffic jams.
Restrictions enforcement is one of the typical fields which are involved with one-sided long tail probability densities, due to the psychological effect of the restrictions on the driver's behaviour. Therefore, attempts of practical improvements for non-parametric detection certainly deserves a gain of attention in traffic management.

References

[1] Antofie, A., Boreux, J.: Approach of the walloon legal metrology (Belgium) for weigh-in-motion (wim) free flow direct enforcement, 8th International Conference on Weigh-in-Motion, 2019.

[2] Gadja, P., et al.: Accuracy maps of weigh-in-motion systems for direct enforcement. MDPI Electronics, 12 (2023) 1621, DOI: 10.3390/electronics12071621

[3] Koo, B., Linton, O.: Estimation of semi-parametric locally stationary diffusion models, Journals of Econometrics, 170 (2012), pp. 210–233, DOI: 10.1016/j.jeconom.2012.05.003

[4] Runger, G., Testik, M.: Multivariate extensions to cumulative sum control charts, Quality and reliability engineering international, 20 (2004), pp. 587–606

[5] Vogt, M.: Non parametric regression for locally stationary times series. The Annals of Statistics, 40 (2012), pp. 2601–2633, DOI: 10.1214/12-AOS1043

[6] Vogt, M., Dette, H.: Detecting gradual changes in locally stationary processes. The Annals of Statistics, 43 (2015), pp. 713–740, DOI: 10.1214/14-AOS1297

[7] Warscotte, L., Boreux, J.: Direct Enforcement in Belgium with High Speed Weigh-in-Motion (HS-WIM), MDPI electronics, 12 (2023) 55, DOI: 10.3390/electronics12030555