Rescaled range analysis of service load data

K J Rathanraj[1]*, **A Srividya**[2], **A K Verma**[2], **R Raman**[3], **A V Mannikar**[4], and **V A Pankhawala**[4]

[1]Department of Industrial Engineering and Management, BMS College of Engineering, Bangalore, India

[2]IDP – Reliability Engineering Group, Department of Electrical Engineering, Indian Institute of Technology, Bombay, Mumbai, India

[3]Department of Material Science and Metallurgy, Indian Institute of Technology, Bombay, Mumbai, India

[4]Structural Dynamics Lab, The Automotive Research Association of India (ARAI), Pune, India

The manuscript was received on 24 April 2009 and was accepted after revision for publication on 11 August 2009.

DOI: 10.1243/09544070JAUTO1247

Abstract: Recent advances in data acquisition, storing, and recording technology are allowing increasingly detailed samplings of actual customer usage of product. Because of this, laboratories are coming forward to use the actual road data rather than simulated data or data collected from the proving ground. This research work is an attempt to find a representative length of data (short-term data) from the large-distance data (long-term data). For any particular class of vehicle, finding the short-term data, which are representative of long-term data, will avoid the collection of large-distance data in the future.

In this paper, using the Hurst methodology an attempt have been made to find the short-term data that have to be collected, so as to be representative of the long-term data. In this study, data have been collected both on the proving ground as well as on an actual road for a distance of 1700 km on a light commercial vehicle. It has been seen from the Hurst exponent that, collecting data for an actual road length of about 200 km would give a representation for an actual road length of 1000 km.

Keywords: service load, road load data acquisition, Hurst exponent

1 INTRODUCTION

The development cycle time of a vehicle model is continuously decreasing. This does not permit adequate time for proving and evaluating component, assemblies, and vehicles for durability by conventional measures. The technique calls for suitable concepts to be developed in order to compress the evaluation time. On the other hand, it is required to validate the service load data used for testing in order to meet the customer requirement for efficiency, reliability, and high performance of the vehicles. The time span available for bringing the new model to market is also becoming as short as possible.

Corresponding author: Department of Industrial Engineering and Management, BMS College of Engineering, Bull Temple Road, Basavanagudi, Bangalore 560 019, India.
email: rathanrajkj@yahoo.com

The concept of systematic measurement of service loads, analysis and application of fatigue evaluation to design optimization processes, and proving by simulation techniques are being applied increasingly by the automobile industry.

Durability assessment is carried out mainly to evaluate the structural adequacy and the fatigue life of the vehicle components. For product durability, a designer's objective is to specify the product that achieves some defined reliability level when placed in the hands of the customer. More details regarding evaluation for durability can be obtained from references [**1**] to [**4**].

It is observed that in the laboratory, in order to carry out durability testing, service data are collected from the proving ground [**5**]. Further, in recent years because of developments of new technologies in automated control of servohydraulic actuators, it has been possible to combine the above to simulate the dynamic effects of the test track on vehicles and their components. Laboratory simulation holds the potential for mainly accelerating the testing time, reducing testing and manufacturing start-up costs, and ultimately producing qualitatively better test results and higher quality for the vehicle. The ability to have measured vehicle field data and laboratory vehicle behaviour data readily available for study and comparison leads to a much greater understanding of the problems associated with vehicle design. Recently, owing to technological innovations, problems related to collecting and storing of large amounts of data seem to be solved. As a consequence of this, laboratories are coming forward to use the actual road data rather than simulated data or data collected from the proving ground. This research on finding the representative length of data (short-term data) from the large-distance data (long-term data) is an attempt in that direction.

In this paper, an effort has been made to find the short-term data from the long-term data using the Hurst exponent H, which measures the long-range statistical interdependence in the time series or histories. Here, an attempt has also been made to find the relationship or dependence between the different types of road and speed for a light commercial vehicle at fixed payload.

2 BACKGROUND

The Hurst exponent H has broad applicability to all time series analysis because it is extremely robust. It can distinguish a random series from non-random series, even if the apparently random series is non-Gaussian (i.e. not normally distributed). Hurst invented the rescaled range analysis, also called the R/S or the Hurst method, for the evaluation of time-dependent hydrological data [**6**]. His original work is related to water reservoirs and the design of an ideal storage facility on the river Nile. Mandelbrot and Wallis [**7–11**] popularized the method to find wide applications in many fields of science. The Hurst exponent occurs in several areas of applied mathematics, including fractals and chaos theory, long-memory processes, and spectral analysis. Estimating the Hurst exponent for a data set provides a measure of whether the data are of a pure random walk nature or have underlying trends. Another way to state this is that a random process with an underlying trend has some degree of autocorrelation. When the autocorrelation has a very long (or mathematically infinite) decay, this kind of Gaussian process is sometimes referred to as a long-memory process. Processes that might naively be assumed to be purely random sometimes turn out to exhibit Hurst exponent statistics for long-memory processes. Hurst exponent estimation has wide applications ranging from biophysics to computer networking. There is sufficient literature on both theory and applications on testing for long-range dependence. Hurst measured how the reservoir level fluctuated around its average level with time. Depending on the length of time used for measurement, it could be expected that the range of this fluctuation would also change. If the series were random, the range would increase with the square root of time. To standardize the measure over time, Hurst decided to create a dimensionless ratio by dividing the range by the standard deviation of the observations. Hence, the analysis is called rescaled range analysis.

Detailed mathematical aspects of the method have been described elsewhere [**12**, **13**]. According to Weron [**14**], the Hurst exponent can be calculated in the following way: the analysis begins by dividing a time series of length L into d subseries of length n. Next, for each subseries $m = 1, \ldots, d$, first, find the mean E_m and standard deviation S_m, second, normalize the data $Z_{i,m}$ by subtracting the sample mean $X_{i,m} = Z_{i,m} - E_m$ for $i = 1, \ldots, n$, third, create a cumulative time series $Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ for $i = 1, \ldots, n$, fourth, find the range $R_m = \max\{Y_{1,m}, \ldots, Y_{n,m}\} - \min\{Y_{1,m}, \ldots, Y_{n,m}\}$, and, fifth, rescale the range R_m/S_m. Finally, calculate the mean value of the rescaled range for all subseries of length n, according to

$$\left(\frac{R}{S}\right)_n = \frac{1}{d}\sum_{m=1}^{d}\frac{R_m}{S_m} \tag{1}$$

It can be shown that the R/S statistics asymptotically follow the relation

$$\left(\frac{R}{S}\right)_n \approx cn^H \tag{2}$$

Thus, the value of H can be obtained by running a simple linear regression over a sample of increasing time horizons according to

$$\log\left(\frac{R}{S}\right)_n = \log c + H \log n \tag{3}$$

the Hurst exponent, which varies between 0 and 1, characterizes the long-memory effect of the time series data. There are three distinct classifications for the Hurst exponent:

(a) $H = 0.5$;
(b) $0 \leqslant H < 0.5$;
(c) $0.5 < H < 1$.

$H = 0.5$ denotes a truly random series. Events are random and uncorrelated. This also means that the present does not influence the future. R/S analysis can classify an independent series, irrespective of the shape of the underlying distribution. For $0 \leqslant H < 0.5$, the system is an anti-persistent series (chaotic). It is often referred to as 'mean reverting'; if the system has been up in the previous period, then it is more likely to be down in the next period. Conversely, if it was down previously, it is more likely to be up in the next period. When $0.5 < H < 1$, the system has a persistent or trend-reinforcing series, if the series has been up (down) in the last period, then the chances are that it will continue to be positive (negative) in the next period. Trends are apparent. The closer H is to 0.5, the noisier it will be, and the less defined its trends will be. Persistent series are fractional Brownian motion or biased random walks. The strength of the bias depends on how far H is above 0.5.

3 EXPERIMENTS

The vehicle body is subjected to complex loads from each of the wheels of the vehicle especially during rough-road driving. The procedure for obtaining the road inputs involves acquiring the acceleration data from the accelerometers mounted at various predetermined locations of the prototype vehicle during the actual running of the vehicle, on different (representative) road surfaces, i.e. paved road, country road, highway, etc., which is decided on the basis of the severity and type of vehicle. Generally, road load data are acquired in terms of time series or history to quantify typical service conditions on different types of road and test track in order to obtain a realistic database. These in-service data can also be acquired the same way on an actual production vehicle as in the case of data acquired on a prototype. To cover typical in-service cycle data, the length of data acquisition corresponds to a few hundred kilometres to a few thousand kilometres of (intercity to interstate) travel. These data acquisition exercises are referred to as long-term data acquisition. Usually short-term service load data are used to predict the durability of the components or complete structure in the accelerated-testing laboratories. The service load data collected from the proving ground and short-term road data are used in the simulation test in laboratories. Therefore, it becomes necessary to find the approximate short-term road load data from the long-term or actual road load data for it to be more representative.

For collecting service load data, physical variables such as road surfaces, payload, and speed can be fixed. In this experiment, three types of road were selected, namely highway, country road, and paved road. The selected speeds of the vehicle were 20 km/h, 30 km/h, and 40 km/h on a paved road, 50 km/h and 80 km/h on a highway, and 25 km/h and 35 km/h on a country road. Once these variables are fixed, the location of the vehicle and the type of vehicle can be selected. A light commercial vehicle was used in the present study with four accelerometers fixed on the axle and four on the chassis for acceleration measurement.

The selection of accelerometers is made according to the requirement. See references [**15**] to [**17**] for details regarding the road load data acquisition. The data collected are required for validation in terms of their representativeness, repeatability, and reproducibility. To obtain better repeatability the whole exercise was repeated twice or more times depending on the accuracy and end usage requirements in the case of the proving ground. For data acquisition, 200 samples/s have been fixed on the basis of past experience that the maximum frequency may not exceed 60 Hz.

Figure 1 shows a typical time series acceleration signal of an axle. The measured random stress–time history is unique and contains the effects of both external loads and the dynamic response of the structure. This means that the histories are affected not only by the structural system but also by the location where the history has been observed. Figures 2 and 3 show the accelerometer locations on the front axle and chassis respectively.

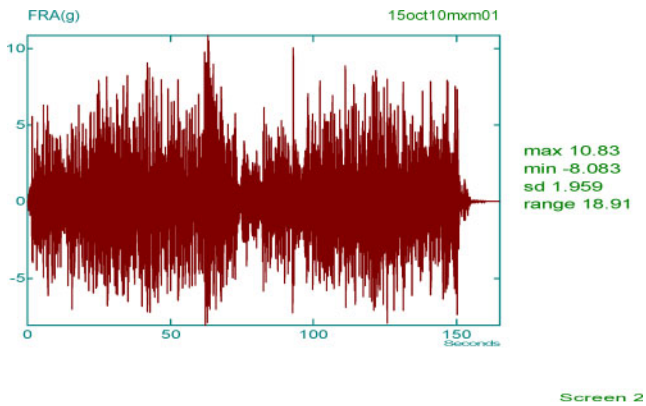

Fig. 1 Typical time series acceleration signal of the front right axle (FRA)

Fig. 2 Accelerometer location on the front axle

Fig. 3 Accelerometer locations on the front chassis

4 RESULTS AND ANALYSIS

Acceleration data on the light commercial vehicle are collected by running the vehicle on the proving ground at different speeds and also at a random speed on the actual road. Signal data collected are random and non-stationary and are analysed for estimation. Considering the data to be a Gaussian white-noise sequence (independent and $N(0, 1)$ distributed) of length $L = 2^N$ where $N = 8, 9, \ldots, 16$, i.e. $L = 256, 512, \ldots, 65\,536$. In order to determine the Hurst exponent, log (R/S) is plotted against log N and the slope gives H. However, not all the points of this plot have the same statistical weights. When N is very large, only a few R/S data are obtained, and, so the statistics are poor. When N is small, a large number of R/S data can be calculated but their scatter is large. For this reason, the first and the last few points of the double-logarithmic plot are usually discarded. The calculated values of Hurst exponent for various road surfaces and speeds selected in this study are listed in Table 1.

From Table 1, it can be observed that the H values that are less than 0.5 are anti-persistent and have an increasing trend in the past, which implies a decreasing trend in the future, and vice versa. This is true, as the proving ground gives a unique set of road inputs to the vehicle, which means that there is much less chance of having a long memory and the stretches on which the data are collected are of smaller length; i.e. a 1 km stretch for a paved road and 5 km on a country road. From the table, it can also be observed that the H values obtained from front right and left axles on the paved road and the country road decrease as the speed increases. In contrast, on the highway the persistence value increases slightly with increase in the speed.

The estimation of the Hurst exponent alone is not enough. A measure of the significance of the results is also needed. Traditionally, the statistical approach is to test the null hypothesis of no or weak dependence versus the alternative of strong dependence or long memory at some given significance level. Based on reference [**14**], Table 2 shows the empirical confidence interval for R/S statistics and sample length $L = 2^N$. According to the empirical confidence interval, the values calculated for sample

Table 1 Hurst exponent values for different types of road and different speeds on the proving ground at a fixed payload of 75 per cent on a light commercial vehicle

			H value for the acceleration data of the following							
Type of vehicle	Type of road	Speed (km/h)	Front right axle	Front left axle	Rear right axle	Rear left axle	Front right chassis	Front left chassis	Rear right chassis	Rear left chassis
Light commercial vehicle	Paved	20	0.5244	0.5363	0.5346	0.5186	0.4505	0.4385	0.4133	0.4303
		30	0.5158	0.5186	0.4953	0.5091	*	0.4025	0.42	0.4327
		40	0.4933	0.4885	0.5138	0.5053	*	0.3852	0.4253	0.4683
	Highway	50	0.4965	0.5053	0.4992	0.4977	0.4638	0.5246	*	*
		80	0.5342	0.5129	0.4868	0.4809	0.5425	0.5677	*	*
	Country	25–30	0.5332	0.5053	*	*	*	0.6286	0.589	0.5298
		30–35	0.5055	0.5097	0.5114	0.5266	*	0.6725	0.5508	0.4705

*No proper signal has been collected to calculate the *H* value

Table 2 Empirical confidence intervals

Level (%)	Confidence interval for $n_1 > 50$ Lower bound	Upper bound
90	$0.5 - \exp[-7.35\ \log(\log N) + 4.06]$	$\exp[-7.07\ \log(\log N) + 3.75] + 0.5$
95	$0.5 - \exp[-7.33\ \log(\log N) + 4.21]$	$\exp[-7.20\ \log\ (\log N) + 4.04] + 0.5$
99	$0.5 - \exp[-7.19\ \log(\log N) + 4.34]$	$\exp[-7.51\ \log\ (\log N) + 4.58] + 0.5$

length $L = 65\,536$ ($N = \log_2 65\,536 = 16.0$) are shown in Table 3.

Looking at the interval values in Table 3 and comparing them with the proving ground values, it can be stated that all the values fall within the interval. This signifies that it obeys the null hypothesis, meaning that there is no long memory in these data. It can also be seen that the data presented in Table 3 constitute a random series.

Further, road load data on the actual road with random speed were collected on the light commercial vehicle. Analysing the road load data for H value provides a persistent value for both the front right axle (FRA) and the front left axle (FLA) for a distance of 1000 km, which is shown in Table 4 and plotted in Fig. 4. This means that there is long memory in the data, indicating some consistency. The main objective of this experiment is to predict the short-term distance from the long-term distance. With this objective the short-term distance was found by splitting the data into a 100 km patch and a 200 km patch and their H values are calculated. The obtained H values are listed in Table 4 and the same values are plotted in Fig. 5. It can be observed that the values of 100 km are not consistent and have values less than the 99 per cent confidence interval.

For the 99 per cent confidence level

$$P\left[\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n_1}}\right)\right] < \mu < P\left[\left(\bar{x} + 2.58\frac{\sigma}{\sqrt{n_1}}\right)\right]$$

For the distance of 100 km

$$0.7689 < \mu < 0.9002(\text{FRA}),$$
$$0.7535 < \mu < 0.8709(\text{FLA})$$

For the distance of 200 km

$$0.8044 < \mu < 0.9032(\text{FRA}),$$
$$0.8256 < \mu < 0.9397(\text{FLA})$$

However, the 200 km patch shows as more consistent; and most of the values lie within the confidence interval and the one H value that is higher may be considered as an outlier. Thus, the data have to be collected for a distance of around 200 km in order to represent the actual road for this type of vehicle. It is known that, in the Hurst analysis, the value of persistence after reaching some point starts to slope

Table 3 Empirical values calculated at 90 per cent, 95 per cent, and 99 per cent confidence levels

Level (%)	N	Confidence interval for $n_1 > 50$ Lower bound	Upper bound
90	16	0.467 79	0.531 43
95	16	0.4618	0.536 79
99	16	0.449 82	0.546 03

Table 4 H values and the correlation coefficients R^2 for distances of 100 km, 200 km, and 1000 km in a light commercial vehicle at random speed

	100 km distance				200 km distance				1000 km distance			
	FRA		FLA		FRA		FLA		FRA		FLA	
Patch	H	R^2	H	R^2	H	R^2	H	R^2	H	R^2	H	R^2
1	0.7264	0.9233	0.7665	0.9339	0.8738	0.9156	0.8653	0.9276	0.8055	0.9223	0.8977	0.9298
2	0.8235	0.9556	0.9096	0.9449								
3	0.8650	0.9233	0.8595	0.9278	0.8044	0.9260	0.8334	0.945				
4	0.8382	0.8981	0.8927	0.9373								
5	0.7120	0.856	0.7449	0.9158	0.9156	0.9325	0.9479	0.9021				
6	0.7708	0.9416	0.8372	0.9147								
7	0.9500	0.9292	0.9479	0.8681	0.8461	0.9362	0.8461	0.9362				
8	0.7384	0.9109	0.7699	0.9237								
9	0.9141	0.9323	0.8613	0.9297	0.8292	0.9416	0.9208	0.9292				
10	0.7839	0.9209	0.7559	0.8995								

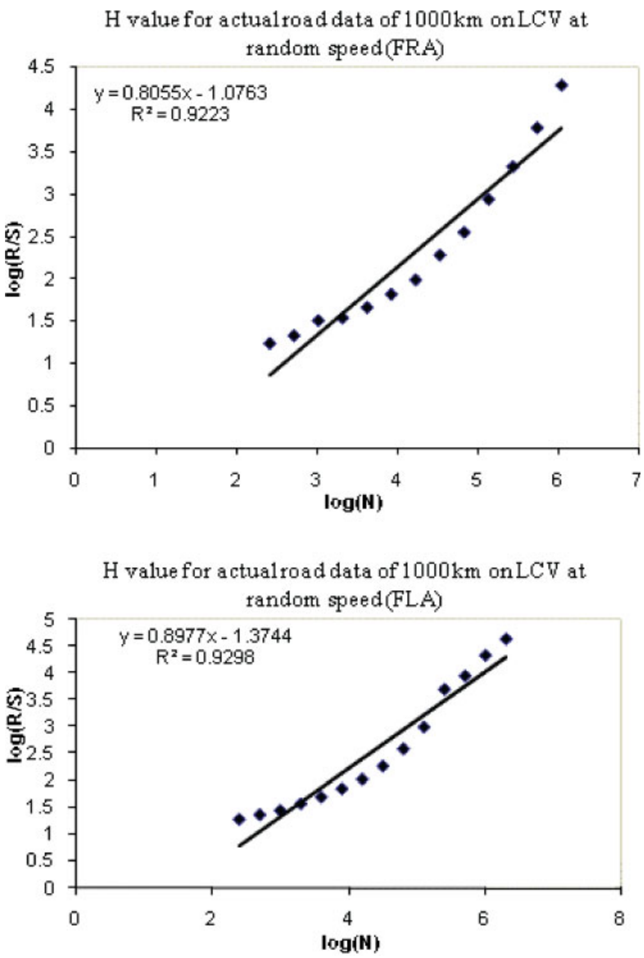

Fig. 4 Typical graph showing the Hurst exponent value for a light commercial vehicle (LCV) on a highway at a random speed for a distance of 1000 km: (a) FRA; (b) FLA

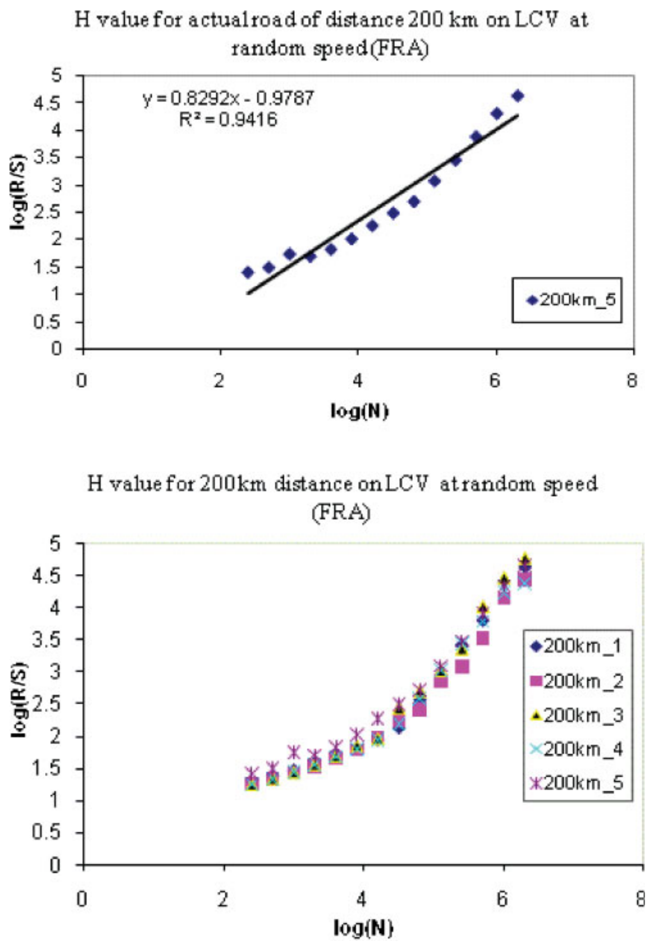

Fig. 5 *H* value for a typical 200 km patch for a light commercial vehicle (LCV) and also for a different patch of the FRA

towards 0.5. If it is possible to achieve this interaction value for the long term, then with some confidence level it can be stated that the maximum distance that has been collected is the long-term value. This could not been achieved in the present study because of the huge number of data, which require more than 2 GB random-access memory for computation.

5 CONCLUSION

By the Hurst exponent, classification applied to the collected data, prima facie, the following conclusions can be drawn.

1. From the proving ground data, it could be concluded that the majority of the values are weakly anti-persistent but they lie within the confidence interval of the random series.
2. On an actual road at a random speed for a light commercial vehicle, the *H* exponent measured for 1000 km (long term) shows a strongly persistent value and the data achieved at a distance of around 200 km (short term) can be considered as representative. This will drastically reduce effort, time, and cost.

ACKNOWLEDGEMENTS

This research work is being jointly carried out at the Indian Institute of Technology, Bombay, Mumbai, India, and the Structural Dynamics Laboratory, The Automotive Research Association of India (ARAI), Pune, India. The authors are grateful to the management of ARAI for giving them the opportunity to carry out the research work. They would also like to thank Mr M. R. Saraf, Mr Prashanth Pawar, Mr Y. V. Dhage and all other staff members of the Structural Dynamics Laboratory, ARAI, for their help with the experimental work.

REFERENCES

1 **Rathanraj, K. J., Srividya, A., Verma, A. K., Mannikar, A. V.,** and **Pankhawala, V. A.** Road load data analysis (RLDA) using Hurst: a new approach. In Proceedings of the Third International Conference on *Automotive and fuel technology*, New Delhi, India, 16–18 January 2004, SAE paper 2004-28-0096 (SAE International, Warrendale, Pennsylvania).

2 **Grote, P.** and **Grender, G.** Taking the test track to the lab. *Automot. Engng*, 1987, **95**(6), 61–64.

3 **Lund, R. A.** and **Donaldson Jr, K. H.** Approaches to vehicle dynamics and durability testing. SAE paper 820092, 1982.

4 **Haq, S., Temkin, M., Black, L.,** and **Bammel, P.** Vehicle road simulation testing, correlation and variability. SAE paper 2005-01-0856, 2005.

5 **Murphy, R. W.** Endurance testing of heavy duty vehicles. SAE paper 820001, 1982.

6 **Hurst, H. E.** Long-term storage capacity of reservoirs. *ASCE Trans.*, 1951, **116**(2447), 770–808.

7 **Mandelbrot, B.** and **Wallis, J. R.** Noah, Joseph and operational hydrology. *Water Resources Res.*, 1968, **4**(5), 909–918.

8 **Mandelbrot, B.** and **Wallis, J. R.** Computer experiments with fractional Gaussian noises. Part 1, averages and variance. *Water Resources Res.*, 1969, **5**(1), 228–241.

9 **Mandelbrot, B.** and **Wallis, J. R.** Computer experiments with fractional Gaussian noises. Part 2, rescaled ranges and spectra. *Water Resources Res.*, 1969, **5**(1), 242–259.

10 **Mandelbrot, B.** and **Wallis, J. R.** Computer experiments with fractional Gaussian noises. Part 3, mathematical appendix. *Water Resources Res.*, 1969, **5**(1), 260–267.

11 **Mandelbrot, B.** and **Wallis, J. R.** Some long-run properties of geophysical records. *Water Resources Res.*, 1969, **5**(2), 321–340.

12 **Mandelbrot, B.** and **Wallis, J. R.** Robustness of the rescaled range R/S in the measurement of non cyclic long run statistical dependence. *Water Resources Res.*, 1969, **5**(5), 967–988.

13 **Feder, J.** *Fractals*, 1988 (Plenum, New York).

14 **Weron, R.** Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A*, 2002, **312**, 285–299.

15 **Saraf, M. R.** and **Jambhale, M. S.** Effect of payload, road surface and speed of vehicle on durability. SAE paper 2003-26-0033, 2003.

16 **Reddy, C. V. R., Satheesh, K. M., Yang, X., Sudhakar, M.,** and **Withrow, R. J.** Analysis of road load data to extract statistical trends in spindle loads for vehicles with variants. SAE paper 2005-01-0859, 2005.

17 **Pawar Prashant, R.** and **Saraf, M. R.** Measurement of road profile and study of its effect on vehicle durability and ride. SAE paper 2009-26-070, 2009.

APPENDIX

Notation

d	total number of divisions of length L
E_m	mean of subseries m
H	Hurst exponent
L	length of the time series
n	length of the subseries m
n_1	number of samples
N	sample length
R_m	range of the subseries m
$(R/S)_n$	mean of the rescaled range for the length L
S_m	standard deviation of the subseries m
$\bar{x}$	mean of the given sample of size n_1
$X_{i,m}$	normalized data of the ith element of the subseries m
$Y_{i,m}$	cumulative data of the ith element of the subseries m
$Z_{i,m}$	data of the ith element of the subseries m
μ	mean of the population
σ	standard deviation of the given sample of size n_1