Variable False Discovery Rate for Structural Feature Extraction from Noisy Measurements

Samer W. Katicha[1*], Amara Loulizi[2], John El Khoury[3], and Gerardo W. Flintsch[4]

Abstract

This paper presents a computational algorithm to identify features in a high dimensional vector of observations containing white Gaussian noise. The algorithm minimizes the classification error of features by adaptively selecting the level at which to control the false discovery rate (*FDR*). The motivation for developing the algorithm is feature extraction in the context of pavement structural evaluation although the proposed algorithm can be applied in a much wider setting. The proposed algorithm was used in a simulation study and the results compare favorably with those obtained using the state of the art empirical Bayes method of Johnstone and Silverman (2004 and 2005) in terms of total square error (L2 loss) and total error (L1 loss), but significantly outperforms the empirical Bayes method in terms of classification error. Furthermore, the proposed algorithm enjoys theoretical optimality results of *FDR* control as demonstrated by Abramovich et al. (2006) and the experimental simulation results show that the proposed algorithm achieves near optimal performance when compared to the best possible

[1] *Senior Research Associate, Center for Sustainable Transportation Infrastructure, Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg VA 24061, USA, email: skaticha@vtti.vt.edu*
[2] *Associate Professor, Ecole Nationale d'ingénieur de Tunis, ENIT, Département Génie Civil, B.P. 37 Le Belvédère 5 1002 Tunis, Tunisia, email: amlouliz@vt.edu*
[3] *Assistant Professor, Departmentrtation of Civil Engineering, Lebanese American University, Byblos, Lebanon, email: john.khoury@lau.edu.lb*
[4] *Professor, The Charles Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, Blacksburg VA 24061, USA, and Director, Center for Sustainable Transportation Infrastructure, Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg VA 24061, USA, email: flintsch@vt.edu*
** Corresponding Author*

classification. The application of the algorithm with wavelet transform denoising using data obtained from pavement structural evaluation is also illustrated.

1. Introduction

Structural health monitoring is a major research area in structural engineering. At the basic stage of structural health monitoring, noisy data is acquired for a response variable that relates to the structural condition. At the next step, feature extraction and identification allows the engineer to distinguish between the different health states of the structure. To be able to make good inference on the structure's condition, the engineer needs efficient and robust methods to extract the relevant features from noisy observations and determine their statistical significance. Suppose that $X = (X_1, \ldots, X_n)$ are noisy observations on a variable of interest such that

$$X_i = \mu_i + \varepsilon_i \tag{1}$$

where ε_i are $N(0, \sigma^2)$ random variables and $\mu = (\mu_1, \ldots, \mu_n)$ is the vector of mean "feature" values of the variable of interest (μ can be thought of the measured value in the ideal case where there is no noise in the measurements). These mean "features" is what we are interested in estimating. In many situations, X represents a series of measurements at different times (time series) or locations (spatial series). In general, without some knowledge of the structure of μ, the best estimate of μ is the measurements vector X itself. However, in many settings μ is sparse (here sparse means that significant portion of the μ_i are either zero or relatively small [compared to σ^2]). In that case, selecting a threshold that sets small "insignificant" measurements to zero while keeping large "significant" ones, can significantly improve estimating μ from X. Naturally, the choice of threshold, which is the main topic of this paper, has a great impact on the goodness

of the estimate. While the sparisty condition at first seems too restrictive, in nonparametric function estimation (time or spatial series estimation) using wavelets, the true wavelet coefficients of the function form a sparse sequence irrespective of the structure of the original μ_i. This feature of wavelets allows for efficient denoising and feature extraction from noisy signals with a properly chosen threshold.

2. Objective and Motivation

The main objective of this paper is to develop a denoising and feature extraction method, for structural health monitoring. The proposed method minimizes the classification error of features as either significant (different from zero) or not (zero) by varying the control of the false discovery rate (*FDR*). Our motivation for the proposed algorithm is to denoise measurements obtained by the Traffic Speed Deflectometer (TSD) for pavement structural evaluation purposes.

The remainder of the paper is organized as follows; in Section 3, we give a brief description of the TSD and the data it collects (more details on the TSD can be found in Flintsch et al. 2013). In Section 4, we present the denoising procedure which consists of 1) calculating the wavelet transform of the measurements, 2) determining a threshold below which the wavelet coefficients are set to zero, and 3) thresholding the wavelet coefficient and calculating the inverse wavelet transform to obtained the denoised signal. The main contribution of this paper is a variable *FDR* procedure to determine the threshold in step 2. In Section 5, we present wavelet transform denoising and the *FDR*. In Section 6, we formulate the proposed variable *FDR*. In Section 7, we compare the proposed method to determine the threshold with the state of the art empirical Bayes method of Johnstone and Silverman (2004, 2005). In Section 8 we present the results of the proposed denoising method applied to TSD deflection measurements and show how denoising

allows for more stable calculation of the asphalt layer modulus. We conclude the paper in section 9.

3. Description of the Traffic Speed Deflectometer

As shown in Figure 1, the TSD is an articulated truck with a rear axle load of 100 kN (22 kips), which in the model evaluated, uses four Doppler lasers mounted on a servo-hydraulic beam to record the deflection velocity of a loaded pavement. Three Doppler lasers are positioned such that they measure deflection velocity at a range of distances in front of the rear axle, at 100, 300, and 756 mm (3.9, 11.8, 29.8 inches). The fourth sensor is positioned 3600 mm (11.5 ft) in front of the rear axle largely outside the pavement deflection bowl, acting as a reference measurement. The beam on which the lasers are mounted moves up and down in opposition to the movement of the trailer in order to keep the lasers at constant height from the pavement surface. To prevent thermal distortion of the steel measurement beam, a climate control system maintains the trailer temperature constant at 20°C (68°F). Data is recorded at a survey speed of up to 80 km/h (50 mph) at a rate of 1000 Hz, which correspond to a 20 mm (0.8 in.) spacing between the raw measurements.

4. Proposed Denoising Procedure

The proposed denoising procedure consists of the following steps:

1. Calculate the discrete wavelet transform of the noisy signal which results in the noisy wavelet at each decomposition level
2. At each decomposition level determine a threshold below which the noisy wavelet coefficients are set to zero
3. Calculate the inverse wavelet transform to obtain the denoised measurements.

Steps 1 and 3 are part of any wavelet transform denoising procedure. Step 2 involves wavelet coefficient shrinking of which, thresholding (which we use in the proposed method), is a particular case. The proposed procedure to determine the threshold at each resolution level, assuming the noise standard deviation is equal to 1, is as follows (in case the noise standard deviation is not known, which is the case with the TSD measurements used in this paper, it can be estimated using the median absolute deviation of wavelet coefficients at the finest resolution level as suggested by Donoho and Johnstone (1995)):

Given n noisy wavelet coefficients $d_1, d_2, \ldots, d_n$, at a specific resolution level

1. Re-order the d_i ($i = 1, \ldots, n$) in decreasing absolute value $|d_1| \geq |d_2| \geq \ldots \geq |d_n|$
2. For each of the ordered d_i calculate the corresponding two-sided p-value p_i under the assumption of the null hypothesis (in this case for the standard normal cumulative distribution Φ)

   $$p_i = 2\left(1-\Phi\left(|d_i|\right)\right)$$

3. For each of the p_i's calculate $q_i = \frac{n}{i} p_i$ and $q_i^* = \frac{p_i\left(n+1-i\right)}{i\left(1-p_i\right)}$
4. Calculate a as the value that minimizes $\sum_{i=1}^{n}\left(aq_i - q_i^*\right)^2$
5. Determine k to be the index i for which $i/n - 2ap_i$ is maximized
6. The threshold is then $|d_k|$, and all d_i that satisfy $|d_i| < |d_k|$ are set to zero

The proposed procedure is inspired from control of the *FDR* (Benjamini and Hochberg 1995) in multiple hypothesis testing. Although Abramovich and Benjamini (1996) proposed using the *FDR* for wavelet transform denoising, their procedure pre-selects the level at which the *FDR* is

controlled. Johnstone and Silverman (2004, 2005) observed that the performance of *FDR* denoising depends on the selected level at which the *FDR* is controlled and that the optimal control level is data dependent. On the other hand, the proposed procedure allows the *FDR* to vary which significantly improves the denoising performance. In order to test the robustness of the proposed denoising method, simulation studies were performed and the results, as presented in section 7 of the paper, show that it performs nearly as good as knowing the optimal threshold.

5. Wavelet Transform Denoising, and *FDR* Control

A key feature of the discrete wavelet transform is its ability to compress the information of a (nonrandom) signal into relatively few large coefficients while most other coefficients can be neglected (essentially set to zero) without significant loss of information. Because white Gaussian noise is not compressible its transform is again white Gaussian noise with the same standard deviation and its energy is spread over all wavelet coefficients. Therefore, taking the wavelet transform of a noisy signal, the relatively few important wavelet coefficients will stick out from the noise. A decision process can then be implemented to differentiate the coefficients that are small (mostly noise) from those that are large (corresponding to the true signal). A popular choice of decision process is to select a threshold such that all coefficients lower than the threshold are classified as small while those higher than the threshold are classified as significant. The coefficients classified as small are then set to zero and the signal is reconstructed from the remaining significant wavelet coefficients. Donoho and Johnstone (1994) showed that this process optimally removes the noise from the signal in the sense that no other denoising procedure can give faster rates of convergence to the original signal. For more in-depth information on wavelet transform denoising, we refer the reader to the work of Donoho and his

collaborators (Donoho 1995; Donoho and Johnstone 1994, 1995; Donoho et al. 1995; Donoho and Coifman 1995).

Threshold Selection

Since the early work of Donoho and Johnstone (1994), numerous methods to modify the empirical wavelet detail coefficients have been proposed. Discussing all proposed methods to determine the threshold is beyond the scope of this paper and we only concentrate here on the empirical Bayes method of Johnstone and Silverman (2004 and 2005) for its good performance and the *FDR* method of Ambramovich and Benjamini (1996) as it forms the basis for the proposed method in this paper. For further details on other threshold selection methods we refer the reader to Antoniadis et al. (2001).

To determine the threshold in the empirical Bayes approach, a prior designed to capture the sparseness of the wavelet expansion is selected to model the wavelet detail coefficients at each resolution level. In most applications, the prior is a mixture of two distributions, γ_1 and γ_2, one distribution designed to capture negligible coefficients, and the other designed to capture significant coefficients as follows

$$d_{jk} \sim (1-\pi_j)\gamma_1(\mu_{1,j}) + \pi_j \gamma_2(\mu_{2,j}) \quad (2)$$

The level-dependent parameters of the mixture prior, $\mu_{1,j}$, $\mu_{2,j}$, and π_j ($0 \le \pi_j \le 1$), are obtained from the empirical wavelet coefficients using empirical Bayes methods based on the marginal maximum likelihood after which Bayes' theorem is used to calculate the posterior distribution of the wavelet coefficients. The coefficients are then estimated at each resolution level using Bayes'

rule to minimize the loss. For the mixture prior, Johnstone and Silverman (2004 and 2005) proposed using an atom of probability at zero and a heavy-tailed distribution such as the Laplace distribution and showed that this choice results in near optimal rates of convergence.

The *FDR* method to determine the threshold has been proposed by Abramovich and Benjamini (1996). The method consists of applying the *FDR* control procedure of Benjamini and Hochberg (1995) (at each resolution level) at a predetermined level $0 < q < 0.5$ and determine the threshold as $|X_k|$, where k is determined through the *FDR* controlling procedure. Abramovich et al. (2006) showed *FDR* thresholding with $0 < q < 0.5$ achieves near optimal rate of convergence and their theoretical results are somewhat stronger than those obtained by Johnstone and Silverman (2004 and 2005) for the empirical Bayes method.

In a simulation study, Johnstone and Silverman (2004) showed that their empirical Bayes method outperformed the *FDR* method of Abramovich and Benjamini (1996) (and other methods such as Stein's Unbiased Risk Estimate [SURE]); the main drawback of the *FDR* method being that the choice of the q value was critical to the performance of *FDR* thresholding. For sparse signals, a low *FDR* value performed better than a high *FDR* value while for dense signals a high *FDR* value performed better than low *FDR* values. In the proposed method in this paper, the value of q is selected based on the data to reduce the classification error.

Shrinking and Thresholding Functions

In wavelet transform denoising, modification of empirical wavelet coefficients is performed using a shrinking or a thresholding function. Following the definition of Johnstone and Silverman (2004), a shrinking function $\delta(x, t)$ is an antisymmetric increasing function on $(-\infty,\infty)$ such that $0 \le \delta(x, t) \le x$. A thresholding function is a shrinking function with $\delta(x, t) = 0$ if and

only if $|x| \le t$. In their original work on wavelet thresholding, Donoho and Johnstone (1995) used the hard, δ_H, and soft, δ_S, thresholding functions defined as follows

$$\delta_H(x,t) = \begin{cases} 0 & \text{if } |x| \le t \\ x & \text{if } |x| > t \end{cases} \tag{3}$$

and

$$\delta_s(x,t) = \begin{cases} 0 & \text{if } |x| \le t \\ \text{sgn}(x)(|x| - t) & \text{if } |x| > t \end{cases} \tag{4}$$

where sgn(x) is defined as follows

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \ge 0 \end{cases} \tag{5}$$

The graphical representation of the hard and soft thresholding functions is shown in Figure 2 (along with the identity function for comparison). A more robust thresholding function, δ_r, can be defined for $1 \le r < \infty$ as follows

$$\delta_r(x,t) = \begin{cases} 0 & \text{if } |x| \le t \\ \text{sign}(x)|x|\left(1 - \dfrac{t}{|x|}\right)^r & \text{if } |x| > t \end{cases} \tag{6}$$

For $r = 1$ the soft thresholding function is recovered while as $r \to \infty$, the thresholding function converges to the hard thresholding function. In between these two extremes, the thresholding function continuously transitions from soft to hard thresholding (see Figure 2). We use this

thresholding function, δ_r, with the proposed denoising method and a data dependent choice of r which we explain later.

The False Discovery Rate

The *FDR* was proposed by Benjamini and Hochberg (1995) in the setting of multiple hypothesis testing. Most of the wavelet transform coefficients, d_i, of a given signal are small (essentially zero) while a few coefficients are large (significantly different form zero). This property allows solving the problem of wavelet transform denoising in terms of hypothesis testing with the following null and alternative hypothesis:

$$H_0: d_i = 0, \text{ vs. } H_a: d_i \neq 0 \quad (7)$$

Hypothesis testing in the case of a single observation is usually performed at a significance level α, for example $\alpha = 0.05$. While this may be adequate in the case of a single observation, for multiple observations, the probability of observing instances where observations from the null hypothesis are rejected becomes substantial and the approach becomes inadequate. For example, given 1,000 observations from the distribution of the null hypothesis (i.e. $d_i = 0$ for all i), hypothesis testing at $\alpha = 0.05$ will on average result in wrongly rejecting the null hypothesis for 50 observations.

To address the drawbacks of multiplicity, Benjamini and Hochberg (1995) proposed controlling the *FDR*. The *FDR* controls the rate of false rejections among all rejections of the null hypothesis. As an example, suppose that controlling the *FDR* of n at 0.1 (10 percent) results in m rejections of the null hypothesis. Then, it is expected that $0.1 \times m$ of the rejections are actually false rejections (meaning the null hypothesis is true) and $0.9 \times m$ of the rejections are actually true

rejections (meaning the alternative to the null hypothesis is true). To formalize the concept of *FDR*, Table 1 summarizes the possible outcomes in multiple hypothesis testing (Benjamini and Hochberg 1995, Storey 2002).

In Table 1, the rejections, R, of the null hypothesis are the observed discoveries. V represents those discoveries that are false while S represents those discoveries that are true ($R = V+S$). V and S are unobservable random variables. The false discovery proportion (*FDP*) is defined as

$$FDP = \begin{cases} \dfrac{V}{V+S} & if\ V+S \neq 0 \\ 0 & if\ V+S = 0 \end{cases} \tag{8}$$

Benjamini and Hochberg (1995) defined the *FDR* as the expected value of the *FDP*

$$FDR = E\left(FDP\right) = E\left(V/R \middle| R > 0\right) P\left(R > 0\right) \tag{9}$$

and showed that it can be controlled at a user specified level $0 < q < 1$ for all values of $0 \leq n_0 \leq n$ (see Table 1 for the definition of n_0 and n). For more information on the *FDR*, the reader is referred to the article of Benjamini and Hochberg (1995) as well as the work of Storey (Storey 2001 and 2002) which defined a similar but slightly different property he called the positive false discovery rate (*pFDR*).

Control of the FDR

Benjamini and Hochberg (1995) proved that the *FDR* can be controlled at a desired level q using Simes' step-up procedure (Simes 1986) as follows (We refer to this procedure as BH)

Given n noisy measurements $X_1, X_2, \ldots, X_n$, of a feature vector $\mu_1, \mu_2, \ldots, \mu_n$, of which $n_0 \leq n$ satisfy the null hypothesis H_0: $\mu_i = 0$, and n-n_0 satisfy the alternative hypothesis H_a: $\mu_i \neq 0$

1. Re-order the X_i in decreasing absolute value $|X_1| \geq |X_2| \geq ... \geq |X_n|$
2. For each X_i calculate the corresponding two-sided p-value p_i under the assumption of the null hypothesis [for the standard normal cumulative distribution Φ:

$$p_i = 2\left(1-\Phi\left(|X_i|\right)\right)]$$

3. Let k be the maximum i such that $p_i \leq \frac{i}{n} q$
4. Reject H_0 for all $i \leq k$.

BH is conservative and the *FDR* is actually controlled at a lower level depending on n_0

$$FDR \leq \frac{n_0}{n} q \leq q \qquad (10)$$

This shows that control of the *FDR* can be improved (increase test power) with knowledge of n_0. Methods to estimate n_0 have been proposed by Benjamini and Hochberg (2000), Storey (2002) and Benjamini et al. (2006). Gavrilov et al. (2009) suggested an adaptive method to estimate n_0 by replacing iq/n in step 3 of BH with $iq/\left[n+1-i\left(1-q\right)\right]$ and showed that this modified procedure controls the *FDR*. This adaptive method gives the estimate $n_{0_i} = n+1-i\left(1-q\right)$.

6. Variable *FDR* procedure to minimize classification error

Control of the *FDR* does not take into account the number of missed discoveries, *T*. In denoising, to improve the estimation of the function, both *V* and *T* need to be minimized. We therefore propose a denoising method that aims to minimize the total number of classification error, *T*+*V*. Before we present the details of the proposed method, we state the following proposition which we will use to minimize the classification error.

240 **Proposition** Given n noisy measurements $X_1, X_2, \ldots, X_n$, of a feature vector $\mu_1, \mu_2, \ldots, \mu_n$, of

241 which $n_0 \leq n$ fall under the null hypothesis H_0: $\mu_i = 0$, and n-n_0 fall under the alternative

242 hypothesis H_a: $\mu_i \neq 0$. Re-order the X_i in decreasing absolute value $|X_1| \geq |X_2| \geq \ldots \geq |X_n|$ and

243 denote by S_i the number of correct discoveries when i null hypotheses are rejected and V_i the

244 corresponding number of false discoveries, then the quantity, $\max_i \{S_i - V_i\}$, minimizes the

245 classification error (i.e. $\max_i \{S_i - V_i\} \Leftrightarrow \min_i \{T_i + V_i\}$).

246 **Proof.** The proof follows simple algebraic manipulation. From Table 1 we have that

247 $T_i = (n - n_0) - S_i$. Therefore

248 $$\min_i \{T_i + V_i\} = \min_i \{(n - n_0) - S_i + V_i\} \Leftrightarrow \min_i \{-S_i + V_i\} \Leftrightarrow \max_i \{S_i - V_i\}$$

249 Where the first equivalence relationship is obtained by noting that $(n - n_0)$ is a constant and

250 therefore can be discarded.

251 To develop the proposed method, recall the definitions of $FDP = V/(V+S) = V/R$ and

252 $FDR = E(FDP)$. Rearranging terms, the following equation relating V, R and FDP can be

253 obtained.

254 $$V = R \times FDP \quad \text{if } R > 0 \tag{11}$$

255 and $V = 0$ if $R = 0$ (12)

256 R is observed while V and the FDP are not known. Using Equation 11, we propose to

257 approximate the FDP with FDR to obtain an approximation to V

$$V \approx R \times FDR \tag{13}$$

In the BH procedure, the p-values are compared to $\frac{i}{n_0} q = \frac{i}{a \times n} q$ and the null hypothesis is rejected if $p_i \leq \frac{i}{a \times n} q$ with q being the level at which the *FDR* is controlled. This approach fixes the *FDR* and then determines the rejections of the null hypothesis. Alternatively, for each p_i, we can calculate the corresponding minimum q value that will ensure that $p_i \leq \frac{i}{a \times n} q$ as follows

$$q_i = a \frac{n}{i} p_i \tag{14}$$

This definition of q_i is similar to the on proposed by Storey (2002). Since this q_i is essentially the lowest *FDR* control limit it can be used to approximate the *FDR* in Equation 13. This allows estimating the number of false discoveries, V_i and the number of true discoveries, S_i for each q_i as follows

$$V_i \approx R_i q_i \text{ and } S_i \approx R_i (1 - q_i) \tag{15}$$

Therefore

$$S_i - V_i \approx R_i (1 - q_i) - R_i q_{ii} \tag{16}$$

However, $R_i = i$ so that

$$S_i - V_i \approx i(1 - q_i) - i q_i = i - 2 i q_i \tag{17}$$

Using the definition of q_i given in Equation 14, we get that

$$S_i - V_i \approx i - 2nap_i \quad (18)$$

Maximizing Equation 18 minimizes the classification error and is equivalent to maximizing

$$i/n - 2ap_i \quad (19)$$

To specify the thresholding function δ_r, we propose to calculate $r = 1/(2q_i)$, $0 < q_i < 0.5$. We can justify this choice heuristically. q_i gives an estimate of the proportion of false discoveries. If this proportion is low then most of the discoveries are true, the calculated r is high and the discoveries are not significantly modified. However, if q_i is close to 0.5, this suggests that a significant portion of the discoveries are actually false (i.e. just noise). The calculated r in this case is close to 1 and the discoveries are shrunk which causes the noise to be attenuated. The simulation studies show that such a choice of r results in good mean square error of the estimated measurements.

Related Work (Higher Criticism)

A concept similar to what we propose is that of the higher criticism (*HC*) proposed by Donoho and Jin (2004, 2008, 2009). Their procedure is essentially the same as the proposed method in this paper except for the replacement of $i/n - 2p_i$ by

$$HC = \frac{i/n - p_i}{\sqrt{i/n\left(1 - i/n\right)}} \quad (20)$$

Though the *HC* has good theoretical properties, we show using a simulation study that, at best, it performs similarly to our proposed criterion and in some cases noticeably performs worse. Note

that the main objective of HC is not to classify features but to detect whether there is evidence of the presence of features.

7. Simulation Studies

First Simulation Study: Johnstone and Silverman (2004)

In the first simulation study, we compare the performances of the proposed method, the empirical Bayes method of Johnstone and Silverman (2004), and the HC method of Donoho and Jin (2008). We reproduced the experiment performed by Johnstone and Silverman (2004) in which they compared their empirical Bayes method with a number of other methods including the *FDR* with different q values. The setup consists of estimating a sequence μ_i of 1,000 observations with $\mu_i = 0$ for all observations except at K randomly chosen positions, where it takes the value μ_0. For each i, a data value $X_i \sim N(\mu_i,1)$ is generated, and μ is estimated using the empirical Bayes method, the *HC*, and the proposed method in this paper.

The parameter K controls the sparsity of the signal, and the values for which results are reported are 5, 50, and 500 (meaning 5, 50 and 500 observations of the original 1,000 observations are different from zero). The other parameter μ_0 gives the strength of the nonzero signal and the values for which results are reported are 3, 4, 5, and 7 (as an example, for $K = 50$ and $\mu_0 = 4$, we have 950 observations from the standard normal distribution centered at 0 and 50 observations from the standard normal distribution centered at 4). One hundred replications are carried out for each value of K and μ_0. For the empirical Bayes method, the posterior median is used with both w and a estimated by marginal maximum likelihood (see Johnstone and Silverman 2004, 2005). For each replication, the total squared error of the estimation $\sum(\hat{\mu}_i - \mu_i)^2$, the total error of the

313 estimation $\sum|\hat{\mu}_i - \mu_i|$ and the total classification error $T+V$ are calculated. The average over 100
314 replications are reported in Table 2 which also shows the classification performance of the oracle
315 threshold. The oracle threshold is the (unknown) threshold that results in the minimum
316 classification error at each replication.

317 Results of the *HC* are similar to the results of the proposed method for $K = 500$ and $K = 50$
318 except for $\mu_0 = 3$. However, for $K = 5$ (all μ_0 values) and $K = 50$ and $\mu_0 = 3$, the *HC* gives very
319 poor results.

320 Results of the empirical Bayes method and our proposed method in terms of square error (L2
321 norm) and absolute error (L1 norm) are comparable. However, in terms of classification error,
322 the empirical Bayes method can have very poor performance in case of dense (high K values)
323 signals and low intensities of μ_0. Finally, the proposed method's performance is comparable to
324 the oracle performance of classification.

325 *Second Simulation Study: Function Estimation*

326 In the second simulation study, we compare the performance of the proposed method with the
327 performance of the empirical Bayes method in signal denoising. The signals we use are the
328 Bumps, Blocks and Doppler signals sampled at 2048 points. These signals, shown in Figure 3,
329 were proposed by Donoho and Johnstone (1994) to evaluate the performance of denoising
330 methods and have become standard test functions for denoising procedures. The signals have
331 high spatial variability and therefore are difficult to denoise. All three signals have a standard
332 deviation of 7. The noisy realizations shown in Figure 3 consist of the signals with added white
333 Gaussian noisy with standard deviation of 5. Figure 4 shows the denoised signal using the
334 proposed method and for comparison purposes, the best performing moving average for each

function is also shown. In the simulation study, we did not assume knowledge of the noise standard deviation. We estimated it from the empirical wavelet coefficients at the finest resolution level with the robust estimator based on the median absolute deviation as suggested by Donoho and Johnstone (1995)

$$\hat{\sigma} = \frac{\text{median}\left(\left\{ \left| \hat{d}_{J\text{-}1,k} \right| : k = 0, 1, \ldots, 2^{J-1} - 1 \right\} \right)}{0.6745} \qquad (21)$$

For the wavelet transform, we used the translation invariant wavelet transform, for both the empirical Bayes method and the proposed method, because of its superior performance (Coifman and Donoho 1995, Johnstone and Silverman 2005). The mother wavelet used is the symlet with two vanishing moments. The performance of the proposed method and of the empirical Bayes method in terms of mean square error is practically the same (Table 3). The advantages of the proposed method are its superior classification performance (see previous simulation study), its simplicity compared to the empirical Bayes method, and the possibility of interpreting the significance of the features through the statistic q_i which is the minimum *FDR* value that results in i rejections of the null hypothesis.

8. Analysis of TSD Measurements

In this section we present the result of denoising deflection slope measurements using wavelet transform with the threshold determined using the proposed method. We show how denoising allows for more accurate estimation of the deflection slope and how this is essential to estimating the pavement asphalt layer elastic modulus.

Denoising and Feature Extraction

Figure 5 shows a set of noisy pavement deflection slope measurements collected on four different sites with the sensor located 100 mm from the rear axle. Sites 1 and 2 are flexible pavements. Three replicate measurements were collected for Site 1 while five replicate measurements were collected for Site 2. Site 3 is a composite pavement with five replicate measurements. Site 4 is a rigid pavement with five replicate measurements. Figure 6 shows the denoised measurements using the proposed method. In this figure, the different replicate measurements were shifted upwards by different amounts so that it is easy to visualize them and therefore differentiate between them. For Site 1 replicates are shifted by 0, 0.1, and 0.2 mm/m while for Site 2, Site 3, and Site 4, replicates are shifted by, 0, 0.5, 1.0, 1.5, and 2.0 mm/m. In general, measurements were repeatable for all sites. Comparing Figures 5 and 6, shows how the proposed denoising procedure can extract the structural features. For instance, the denoised data from site 4, which was analyzed in more details in Katicha et al. (2013, 2014), show large localized weak spots (high deflection slope values shown as spikes in Figure 6) from the beginning of the tested section until a distance of 780 m, and beyond 1380 m. Site 4 consists of an old jointed concrete pavement with a hot-mix asphalt wearing surface layer. Regularly spaced reflective cracks, which are due to the relatively weak joints, are visible on top of the wearing surface. The section between 780 and 1380 m has recently been rehabilitated. Half of the section completely rebuild by removing the jointed concrete. The other half was treated by milling the asphalt layer and overlaying the jointed concrete with a 35 to 40 mm axoshield layer, which is used to mitigate reflective cracking, and then placing a 25 to 30 mm asphalt layer. The results obtained in Figure 6 show that the denoising correctly reflects the construction history of the pavement section, specifically, the spikes show up mainly in the portion of the section that was not rehabilitated.

One way to evaluate the performance of a denoising procedure in simulation studies is to compare the sum of squares of the error between the original signal and the true signal (what would be measured if we did not have noise in the measurements) and the sum of squares of the error between the denoised signal and the true signal. Unfortunately, in practical applications, the true signal is not known. However, because multiple measurements were collected, the average of these multiple measurements is closer to the true signal than any of the individual measurement. Therefore, the average of all runs can be used as a substitute to the true signal to show that the denoised measurements from a single run are closer to the average of all runs than the individual raw measurements. To evaluate how close the denoised or raw measurements are to the average of all runs, we calculate the sum of squares of the difference between each set of data (denoised and raw measurements) and the average measurements of all runs. The results are summarized in Table 4. In all cases, the individual denoised run had a lower sum of squares of differences than the any of individual raw measurements.

Calculation of the Asphalt Layer Modulus

Data denoising is also critical when the deflection slope measurements are used to calculate specific pavement properties through nonlinear regression equations. For instance Katicha et al. (2014) showed how the pavement effective structural number, SN_{eff}, could be significantly affected by the noise when calculated using the raw deflection slope measurements. Here, we show a similar noise behavior in the case of estimating the asphalt layer elastic modulus using Equation 22 which was developed by Xu et al. (2002):

$$\log\left(E_{ac}\right) = -1.0831 \times \log\left(SCI\right) - 2.6210 \times \log\left(H_{ac}\right) + 0.0019 \times H_{ac} + 8.0889 \quad (22)$$

Where E_{ac} is the asphalt layer modulus expressed in MPa, H_{ac} is the asphalt layer thickness in mm, and *SCI* is the surface curvature index in mm defined as the pavement deflection at the point of load application and the pavement deflection 300 mm away from the point of load application. The *SCI* can be calculated from numerical integration of the deflection slope as follows (Flintsch et al. 2013):

$$SCI = 0.1\frac{S_{100}}{2} + 0.2\left(\frac{S_{100} + S_{300}}{2}\right) \quad (23)$$

Where S_{100} and S_{300} are the deflection slopes 100 mm and 300 from the point of load application expressed in mm/m.

Figure 7 shows the calculated E_{ac} for Site 1 and Site 2 using the deflection slope measurements without denoising. Since Equation 22 includes the logarithm of *SCI*, *SCI* should be strictly positive. In the calculations, we have limited *SCI* values to a minimum of 0.01 mm (i.e. *SCI* values lower than 0.01 were set to 0.01). Results presented in Figure 7 show that the noise in deflection slope measurements is significantly amplified and results in large spikes in the calculated E_{ac}. The large spikes appear randomly and have no relation to the actual E_{ac} of the hot-mix asphalt layer. Figure 8 shows the results of E_{ac} calculation using denoised deflection slope measurements. In this case, the calculated E_{ac} is free of large spikes that result from the noise. Note that for Site 1, two large spikes are observed between 1600 and 1800 m. These are repeated in each of the three sets of measurements which suggest that they are not due to random noise but rather to a deterministic and physical feature in the pavement.

9. Conclusion

In this paper, we developed a method to identify features from noisy measurements. The method minimizes the classification error of whether an observed measurement comes from the null hypothesis (i.e. it is due to noise) or from the alternative (i.e. it is a true feature). We applied the proposed method to measurements obtained from the TSD for the purpose of pavement structural evaluation and show how denoising allows 1) identification of interesting features from noisy measurements and 2) calculation of the pavement asphalt layer modulus. We also compared the method with the state of the art empirical Bayes method of Johnstone and Silverman (2004 and 2005) as well as the similar concept of *HC* (Donoho and Jin 2008) and showed that the proposed method is very competitive. In fact, the proposed method either significantly outperforms the other two methods or when it is not the best performing method, it is not significantly outperformed.

References

Abramovich F., and Benjamini, Y. (1996) "Adaptive thresholding of wavelet coefficients" *Computational Statistics and Data Analysis*, Vol. 22(4), pp. 351–361.

Abramovich F., Benjamini, Y., Donoho, D.L., and Johnstone, I.M. (2006) "Special invited lecture: Adapting to unknown sparsity by controlling the false discovery rate" *The Annals of Statistics*, Vol. 34(2), pp. 584–653.

Antoniadis, A. Bigot J. and Sapatinas T. (2001) "Wavelet estimators in nonparametric regression: a comparative study" *Journal of Statistical Software*, Vol. 6, no. 6, pp. 1–83.

Benjamini, Y., and Hochberg, Y. (1995) "Controlling the false discovery rate: a practicel and powerful approach to multiple testing", *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 57(1), pp. 289–300.

443 Benjamini, Y., and Hochberg, Y. (2000) "On the adaptive control of the false discovery rate in
444 multiple testing with independent statistics", *Bioinfromatics*, Vol. 19(3), pp. 368–375.
445 Benjamini, Y., Krieger, A.M., and Yekutieli, D. (2006) "Adaptive linear step-up procedures that
446 control the false discovery rate", *Biometrika*, Vol. 93(3), pp. 491–507.
447 Coifman, R.R., and donoho, D.L. (1995), "Translation-invariant denoising" *Wavelets and*
448 *Statistics, Lecture Notes in Statistics*, Vol. 103, pp. 125–150.
449 Donoho D. (1995) "De-noising by soft-thresholding" *IEEE Transactions on Information Theory*,
450 Vol. 41, no. 3, pp. 613–627.
451 Donoho D. and Johnstone I. (1994) "Ideal spatial adaptation by wavelet shrinkage" *Biometrika*,
452 Vol. 81, no. 3, pp. 425–455.
453 Donoho D. and Johnstone I. (1995) "Adapting to unknown smoothness via wavelet shrinkage"
454 *Journal of the American Statistical Association*, Vol. 90(432), pp. 1200–1224, DOI:
455 10.1080/01621459.1995.10476626
456 Donoho D. and Jin J. (2008) "Higher Criticism for detecting sparse heterogeneous mixtures" *The*
457 *Annals of Statistics*, Vol. 32(3), pp. 962–994.
458 Donoho D. and Jin J. (2008) "Higher criticism thresholding: Optimal feature selection when
459 useful features are rare and weak" *Proceedings of the National Academy of Science*, Vol. 105,
460 no. 39, pp. 14790–14795, DOI: 10.1073/pnas.0807471105
461 Donoho D. and Jin J. (2009) "Feature selection by higher criticism thresholding achieves the
462 optimal phase diagram" *Phylosophical Transactions of the Royal Society A*, Vol. 367, pp.
463 4449–4470, DOI: 10.1098/rsta.2009.0129
464 Flintsch, G.W., Katicha, S.W., Bryce, J., Ferne, B., Nell, S., and Diefenderfer, B. (2013),
465 "*Assessment of Continuous Pavement Deflection Measuring Technologies*," Second Strategic

Highway Research Program (SHRP2) Report S2-R06F-RW-1, The National Academies, Washington, D.C.

Gavrilov, Y., Benjamini, Y., and Sarkar, S.K. (2009) "An adaptive step-down procedure with proven FDR control under independence", *The Annals of Statistics*, Vol. 37(2), pp. 619–629.

Johnstone, I. and Silverman, B. (2004) "Needles and straws in haystacks: empirical Bayes estimates of possibly sparse sequences" *Annals of Statistics*, Vol. 32(4), pp. 1594–1649, DOI 10.1214/009053604000000030

Johnstone I. and Silverman B. (2005) "Empirical Bayes selection of wavelet thresholds" *The Annals of Statistics*, Vol. 33(4), pp. 1700–1752, DOI: 10.1214/009053605000000345

Katicha, S.W., Flintsch, G.W., and Ferne, B. (2013), "Optimal averaging and localized weak spot identification of Traffic Speed Deflectometer measurements", *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2367, pp. 43–52.

Katicha, S.W., Flintsch, G.W., Ferne, B. and Bryce, J. (2014) "Wavelet denoising of TSD deflection slope measurements for improved pavement structural evaluation," *Computer-Aided Civil and Infrastructure Engineering*, Vol. 29(6), pp. 399–415.

Simes, R.J. (1986) "An improved Bonferroni procedure for multiple tests of significance" Biometrika, Vol. 73(3), pp. 751–754.

Storey, J.D. (2002) "A direct approach to false discovery rates" *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 64(3), pp. 479–498.

Storey, J.D. (2002) "The positive false discovery rate: A Bayesian interpretation and the q-value" *Annals of Statistics*, Vol. 31(6), pp. 2013–2035.

Xu, B., Ranjithan, S.R., and Kim, Y.R. (2002) "New relationships between falling weight deflectometer deflections and asphalt pavement layer condition indicators" *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1806, pp. 48–56.

Tables

Table 1 Outcome when testing n hypotheses

Hypothesis	Accept	Reject	Total
Null true	U	V	n_0
Alternative true	T	S	n-n_0
	W	R	N

Table 2 Simulation Results

		k = 5				k = 50				k = 500			
		3	**4**	**5**	**7**	**3**	**4**	**5**	**7**	**3**	**4**	**5**	**7**
norm2	**EB**	37.75	35.47	17.41	7.69	211.40	154.92	102.37	73.97	855.42	873.32	774.99	663.56
	Proposed	36.70	38.15	19.77	6.96	254.75	207.34	113.79	57.04	1108.54	826.50	621.66	516.68
	HC	296.04	207.45	192.62	147.69	240.96	195.36	125.98	60.61	1198.29	988.34	739.55	523.25
	Oracle	*37.46*	*42.83*	*23.21*	*7.77*	*262.36*	*221.44*	*144.77*	*55.15*	*1108.16*	*840.15*	*627.21*	*514.64*
norm1	**EB**	13.81	12.10	7.35	4.88	95.07	73.35	59.56	50.03	709.15	720.35	616.58	503.92
	Proposed	13.53	12.65	7.57	4.74	103.45	79.85	54.29	42.31	600.12	475.71	422.18	403.71
	HC	229.81	158.38	151.65	118.34	110.19	77.91	59.98	45.85	633.54	523.71	447.52	404.01
	Oracle	*13.33*	*12.84*	*7.89*	*4.66*	*104.67*	*81.77*	*58.54*	*41.20*	*599.33*	*477.28*	*420.68*	*401.69*
Classification Error	**EB**	6.82	4.48	1.05	0.41	29.25	14.50	8.52	4.19	500.00	499.94	306.26	98.84
	Proposed	4.93	3.69	1.55	1.15	26.62	11.63	4.59	1.61	94.17	32.67	9.05	1.43
	HC	484.71	342.10	329.79	261.72	46.62	12.73	14.40	11.10	94.73	32.58	9.10	1.40
	Oracle	*3.51*	*1.78*	*0.56*	*0.18*	*23.58*	*9.16*	*2.74*	*0.25*	*88.61*	*29.80*	*7.38*	*0.31*

Table 3 Mean square error of function estimation

Noise Standard Deviation	Bumps		Blocks		Doppler	
	FDR	EB	FDR	EB	FDR	EB
1	0.0999	0.1010	0.0760	0.0758	0.0859	0.0851
3	0.6987	0.7177	0.8037	0.7789	0.5353	0.5335
5	1.7905	1.8245	1.7776	1.7678	1.1741	1.1763
7	3.2898	3.3143	2.8128	2.8633	1.9086	1.9652

497 Table 4 Sum of squares of the differences ("error") with averaged measurements from all runs

Test	Site 1		Site 2		Site 3		Site 4	
	Original	Denoised	Original	Denoised	Original	Denoised	Original	Denoised
1	37.76	20.92	34.65	14.41	64.81	30.63	31.16	13.09
2	37.38	21.26	36.98	14.21	62.99	30.97	27.9	12.5
3	33.07	21.01	33.65	14.60	68.59	33.27	21.81	11.07
4	NA	NA	34.64	15.15	76.49	44.73	22.57	10.2
5	NA	NA	35.54	14.56	67.79	31.06	27.14	13.48

498

499

Figure 1
Click here to download high resolution image

Figure 2
Click here to download high resolution image

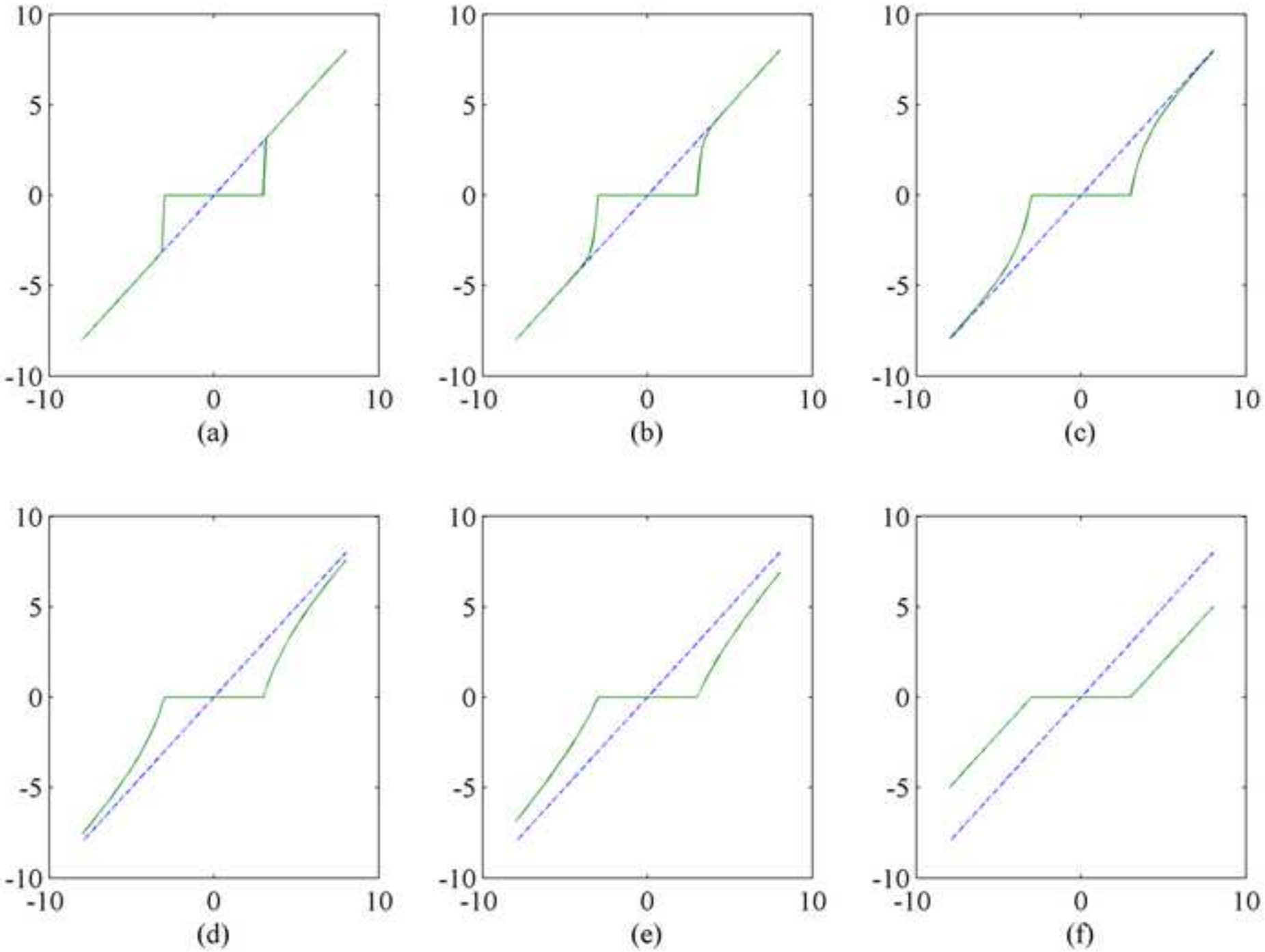


Figure 3
Click here to download high resolution image

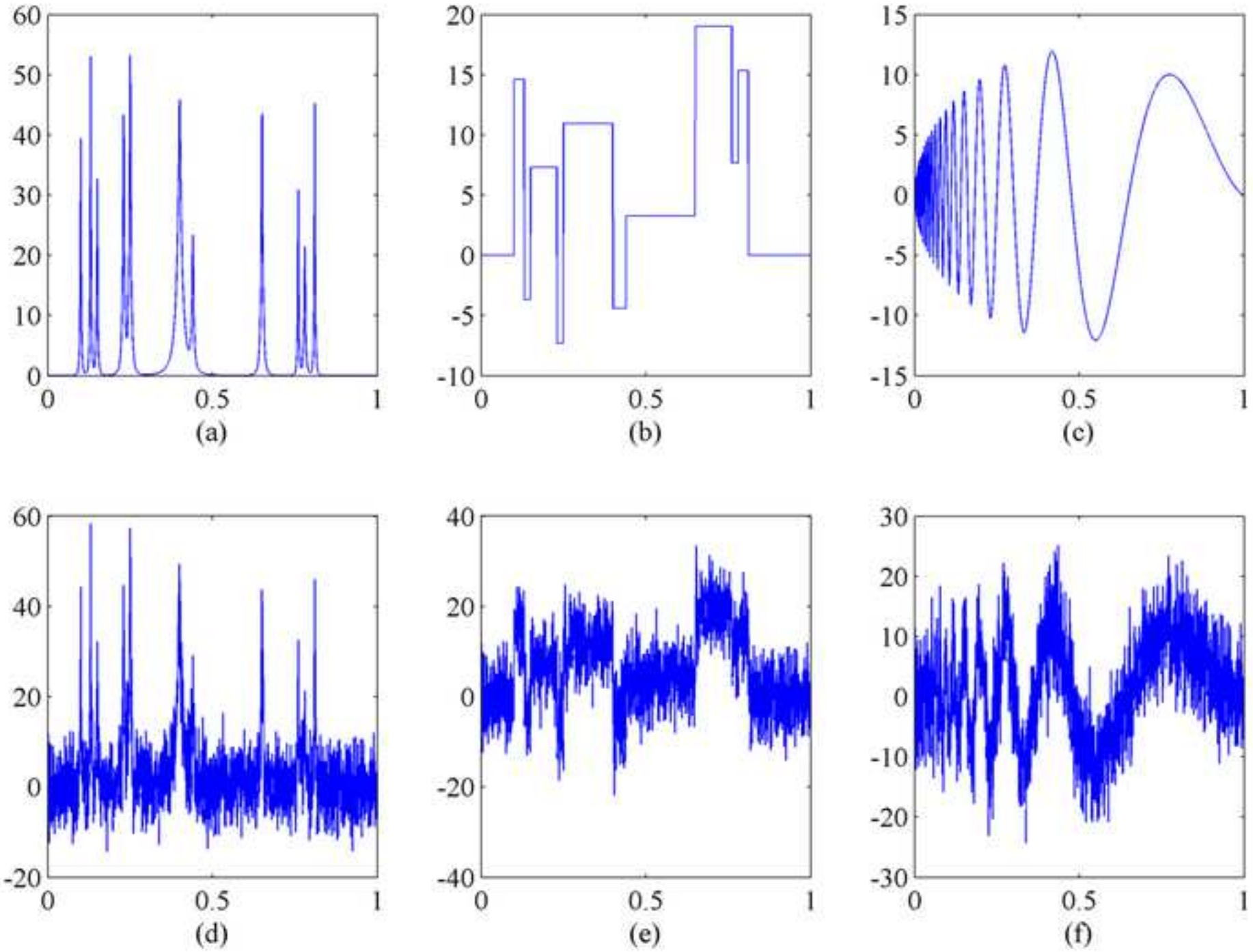


Figure 4
Click here to download high resolution image

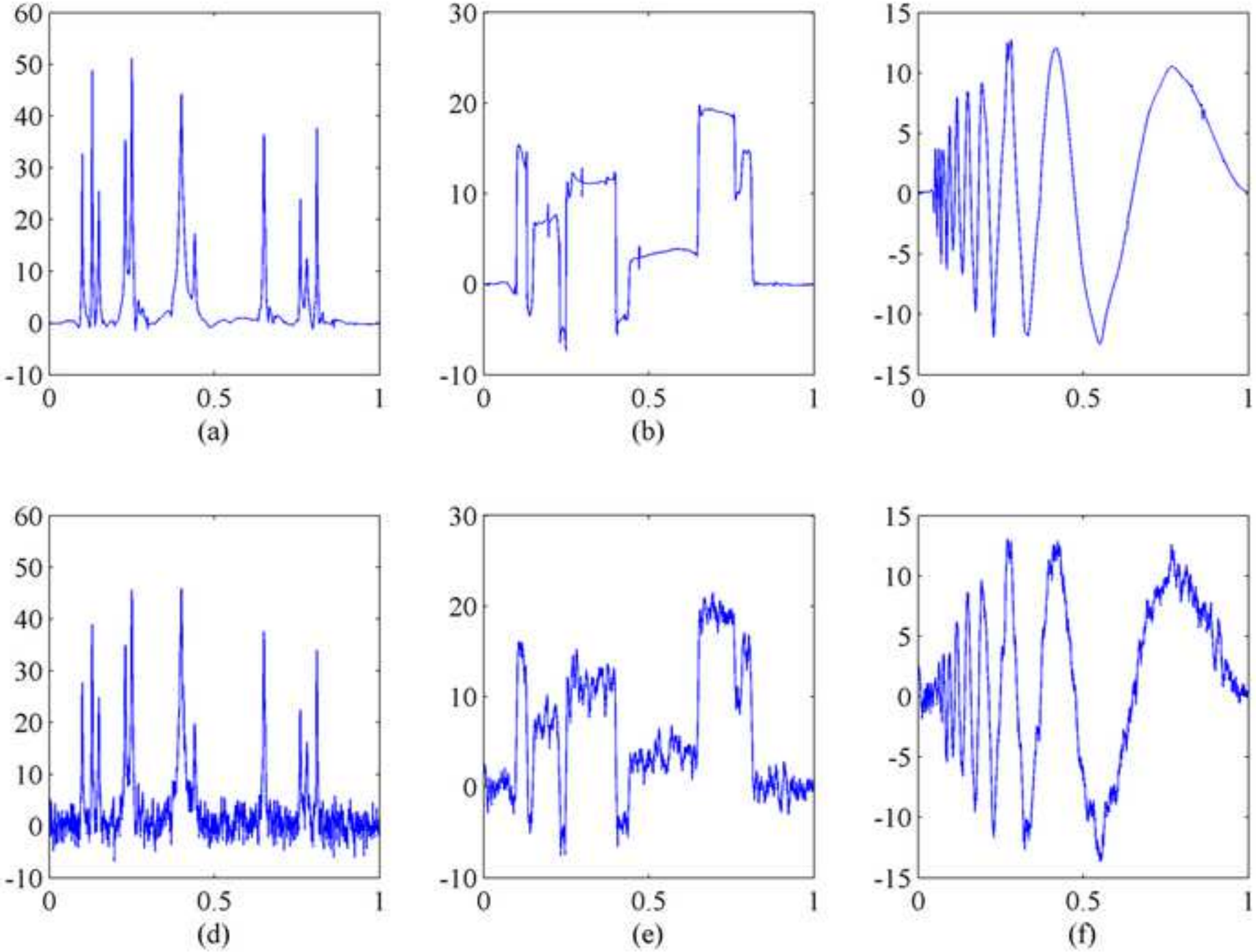


Figure 5
Click here to download high resolution image

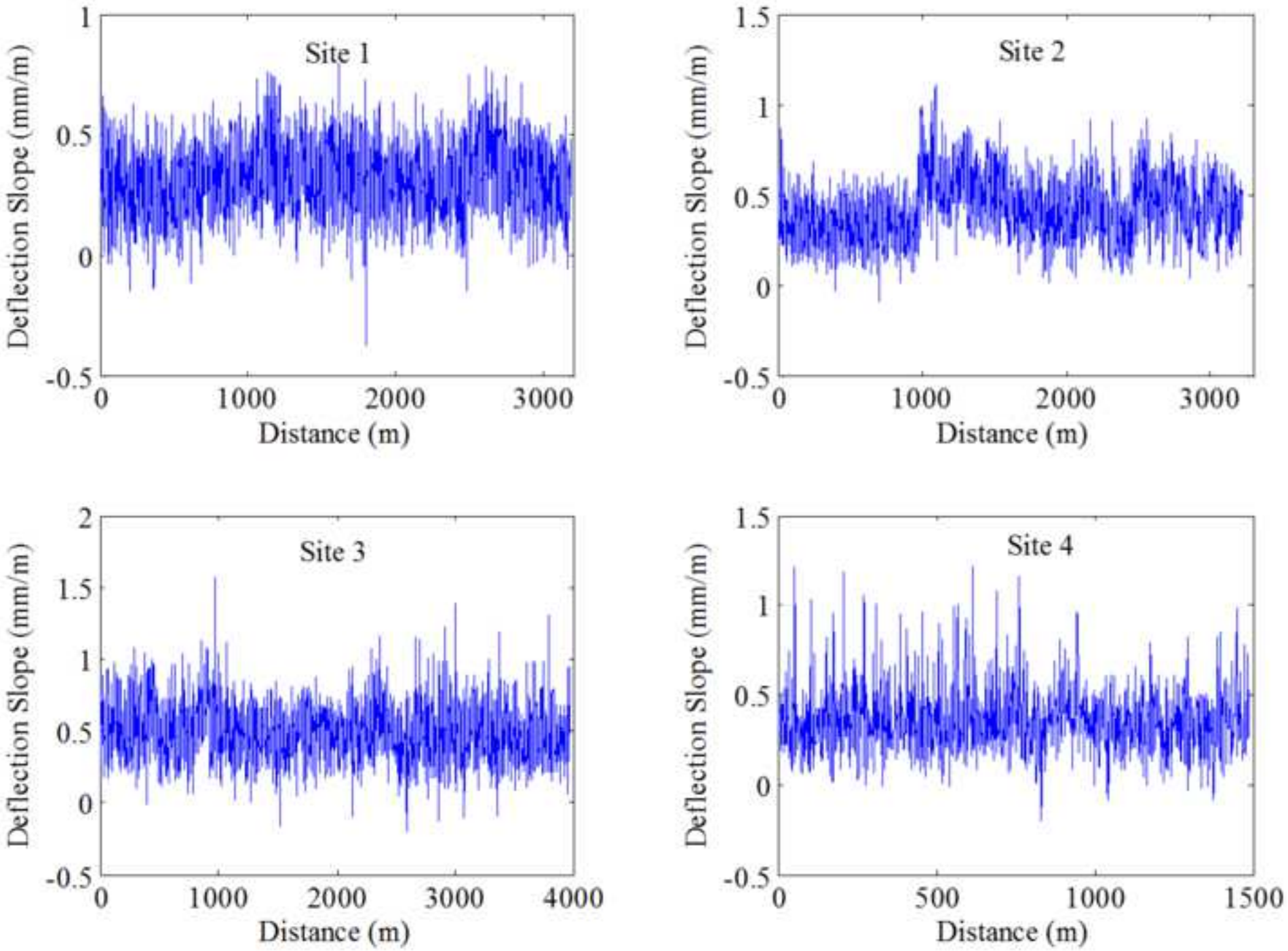
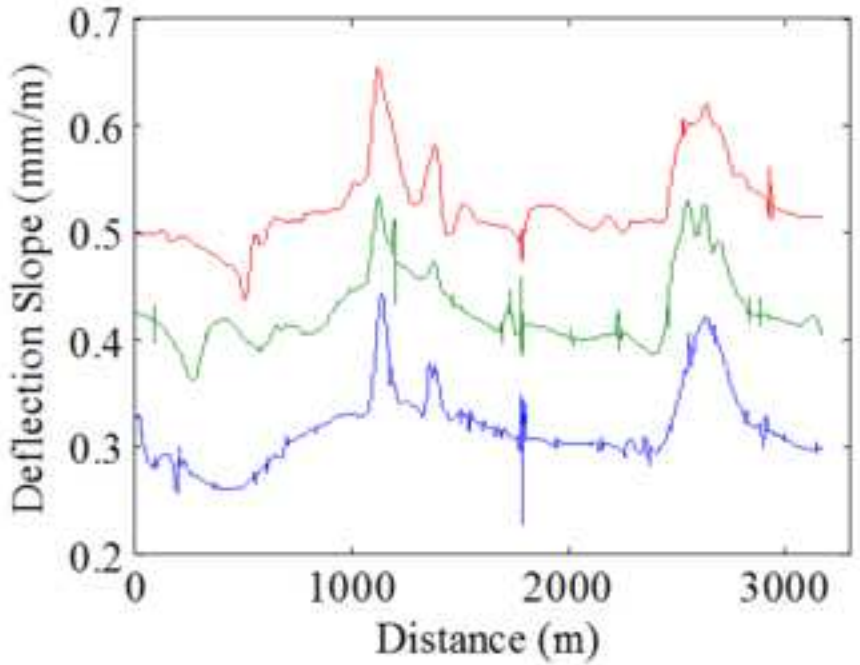
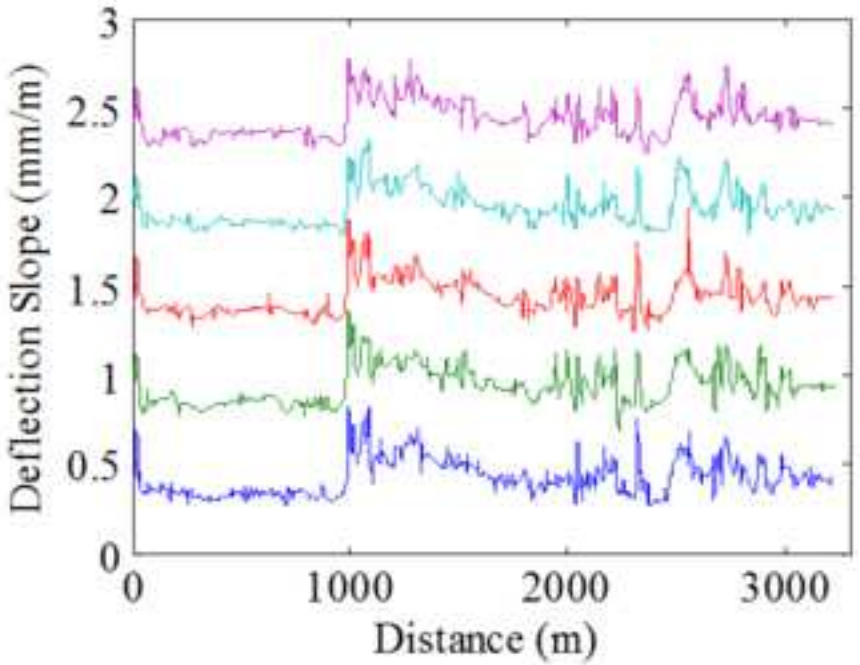
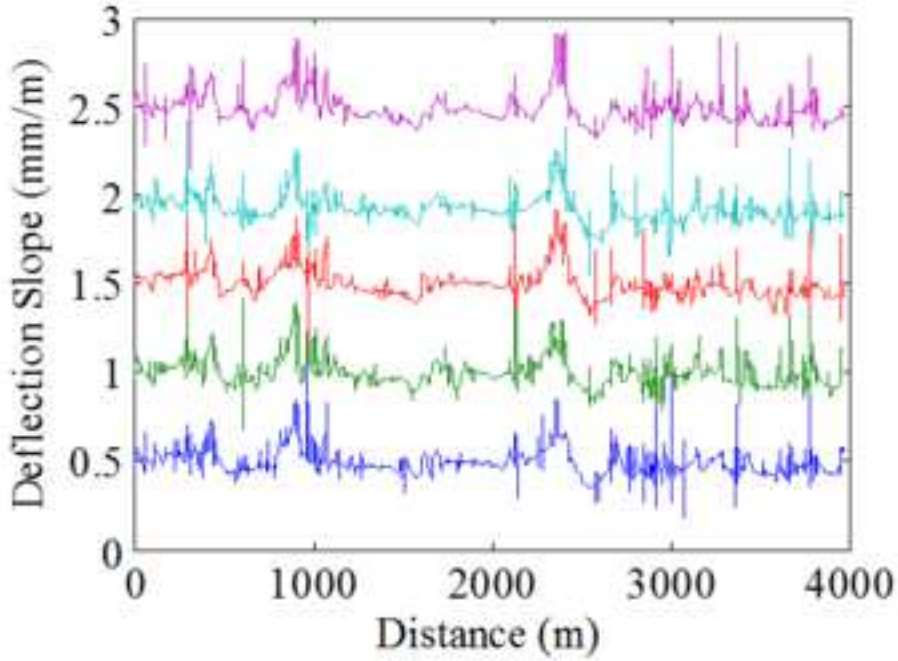


Figure 6
Click here to download high resolution image







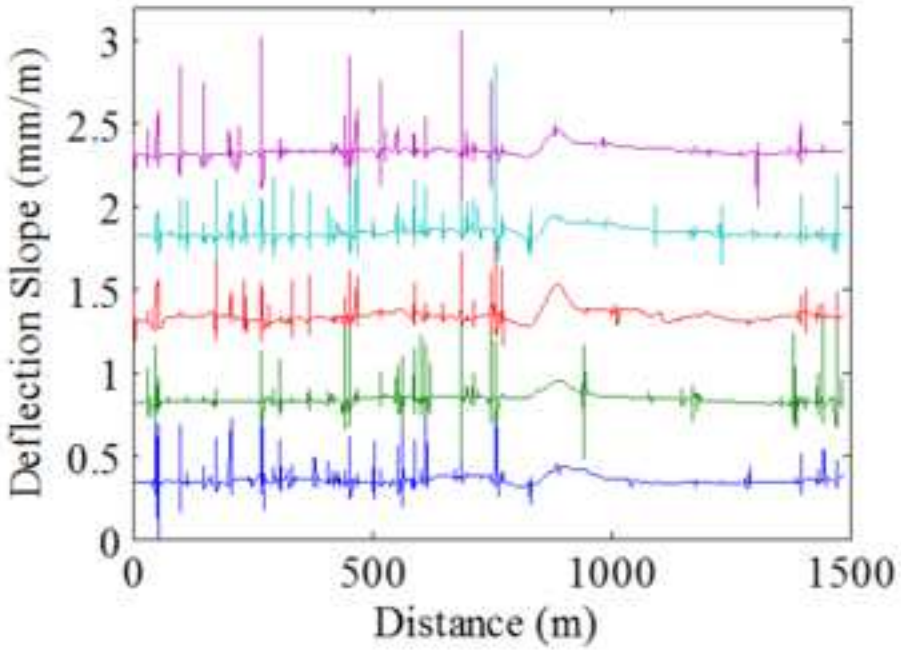


Figure 7
Click here to download high resolution image

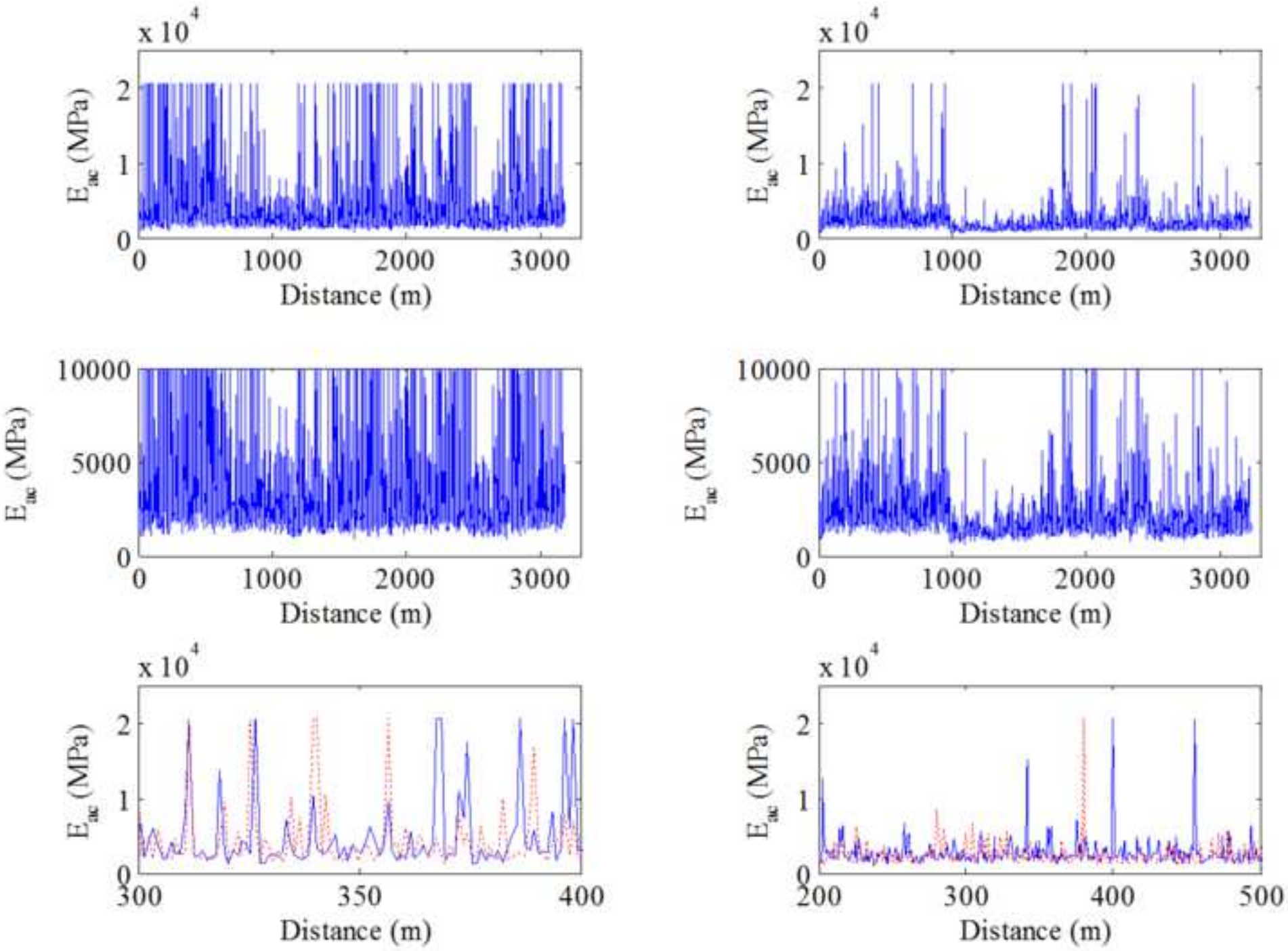


Figure 8
Click here to download high resolution image

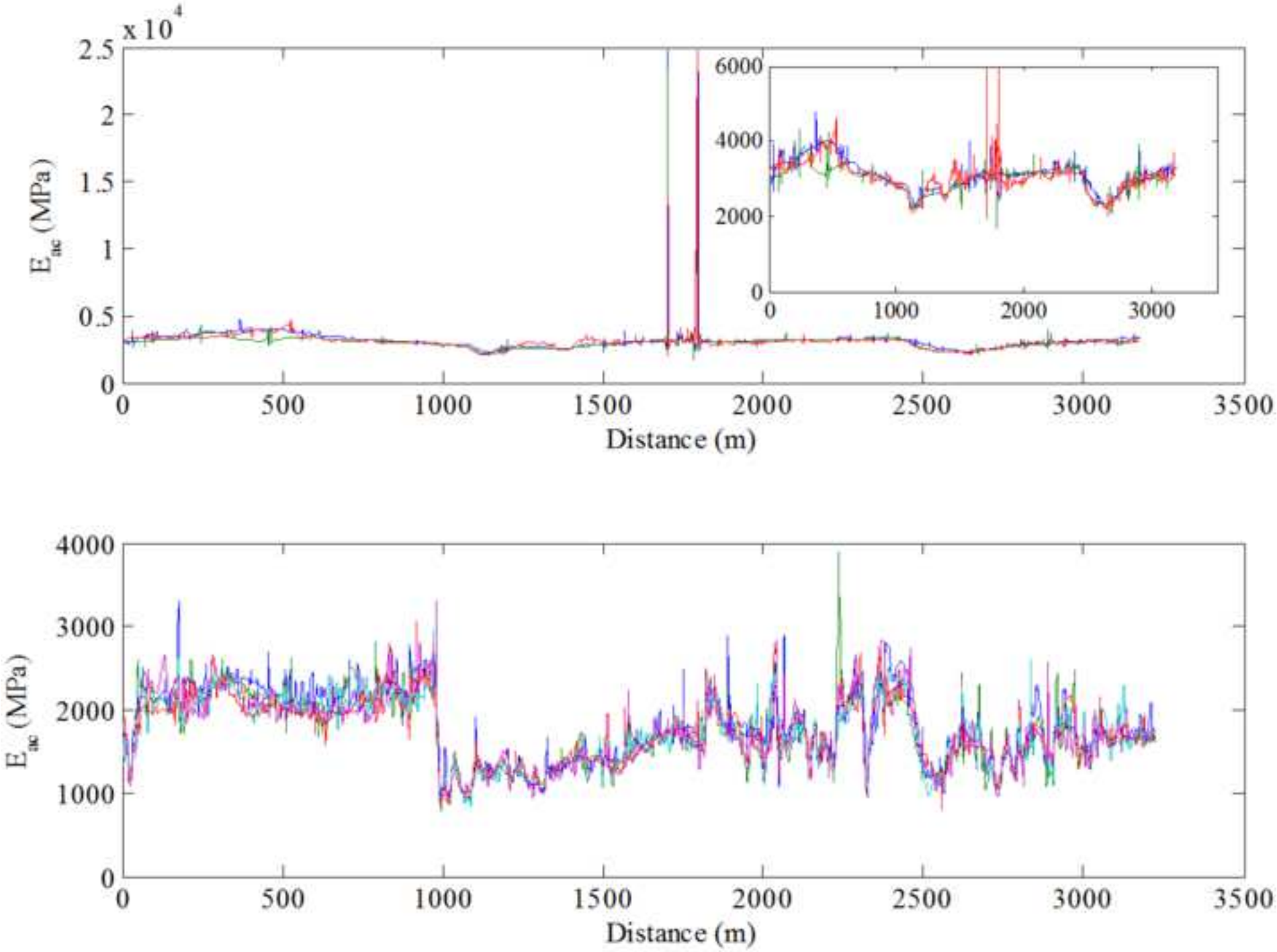


Figure Captions

Figure 1 Picture of TSD and computer generated schematic

Figure 2 Thresholding functions for a threshold T = 3 and different values of *r*; (a) $r = \infty$ (hard thresholding), (b) $r = 15$, (c) $r = 5$, (d) $r = 3$, (e) $r = 2$, (f) $r = 1$ (soft thresholding)

Figure 3 Test functions; Top: Original functions; Bottom: Noisy realization; Left: bumps; Middle: Blocks; Right: Doppler

Figure 4 Denoised functions; Top: Wavelet transform denoising with our proposed method; Bottom: Best moving average estimate; Left: Bumps; Middle: Blocks; Right: Doppler

Figure 5 Deflection slope measurements (noisy) at four test sites

Figure 6 Denoised deflection slope measurements at the four tested sites; replicate measurements are shifted vertically to allow for easier comparison; top: Site 1 (left), Site 2 (right); bottom: Site 3 (left), Site 4 (right)

Figure 7 Calculated asphalt layer modulus from noisy measurements; Left: site 1; Right: site 2; Top: Full measurements; Middle: intermediate detail; Bottom: close detail

Figure 8 Calculated asphalt layer modulus from denoised measurements; Top: site 1, inset figure shows details; Bottom: site 2