International Journal of Innovative
Computing, Information and Control
Volume **13**, Number **6**, December **2017**

 ISSN 1349-4198
pp. 1887–1901

HIGHWAY ELEVATION DATA SMOOTHING USING LOCAL ENHANCEMENT MECHANISM AND BUTTERWORTH FILTER

Xinqiang Chen[1,4,*], Zhibin Li[2,4], Yinhai Wang[3,4], Chaojian Shi[1]
Huafeng Wu[1] and Shengzheng Wang[1]

[1]Merchant Marine College
Shanghai Maritime University
No. 1550, Haigang Ave., Shanghai 201306, P. R. China
[*]Corresponding author: chenxinqiang2005@163.com
{ cjshi; hfwu; szwang }@shmtu.edu.cn

[2]School of Transportation
Southeast University
No. 2, Sipailou, Nanjing 210096, P. R. China
lizhibin@seu.edu.cn

[3]College of Transportation Engineering
Tongji University
No. 4800, Caoan Road, Shanghai 201804, P. R. China

[4]Department of Civil and Environmental Engineering
University of Washington
101 More Hall, Seattle, WA 98195, United States
yinhai@uw.edu

Received May 2017; revised August 2017

Abstract. *Highway elevation is an important factor for performing quantitative and qualitative analysis of traffic safety, traffic operation and management. Google Earth provides public accessible highway elevation data which is deemed as a credible elevation source by many researchers and transportation practitioners. However, the raw elevation dataset, extracted from Google Earth, shows multiple types of outliers. Hence, we presented a local enhancement mechanism integrated with Butterworth filter to remove noises in the raw elevation data. Segment's elevation under local extrema grades was adjusted by the local enhancement mechanism before the Butterworth smoothing process. This operation helped reduce the side effect of over-smooth phenomenon for the Butterworth filter. Elevation data for three typical freeway segments, with some typical categories of outliers, has been analyzed and smoothed. The smoothing results provided evidence that the proposed smoothing model was good at reducing the noises in the raw elevation data. This research can help traffic regulators, administrative departments and field researchers construct trustable and large-scale elevation dataset for nationwide highways.*
Keywords: Denoising, Outlier smooth, Google Earth, Validation, Accuracy

1. **Introduction.** Highway elevation plays an important role in determining many transportation factors such as traffic congestion, freeway geometrical design, vehicle pollution measurement and prediction, and calculation of fuel consumption. In fact, more and more researchers have focused on the elevation and its derivatives' effect on traffic safety [1,2], fuel consumption [3,4] and speed [5,6]. Wang et al., performed several studies to predict and reduce the freeway crashes on the basis of elevation distribution [7,8]. Similarly, Banihashemi found that elevation fluctuation in the horizontal curvature is a significant factor for ensuring rural highways' safety [9]. In addition, Wood et al., analysed roadway grades' influence on vehicle's fuel consumption and speed. It showed that grade variation

caused additional 1% to 3% power consumption for light-duty vehicles [10,11]. Travesset-Baro et al., quantified the influence of the roadway gradient on fuel consumption for both electrical and non-electrical cars under mountainous freeways [12].

For the sake of implementing the above-mentioned studies, reliable elevation dataset is one of the prerequisites. Initially, transportation practitioners and researchers tried to collect elevation data by manually finding out the interested highway segments and drawing the corresponding elevation isolines. Recently, some new and easy accessible elevation sources were sought by researchers. These sources involved the National Elevation Dataset [13], the global 30 arc-second elevation dataset [14], the Global Digital Elevation Model [15], the Shuttle Radar Topography Mission [16] and the Light Detection and Ranging data [17], etc. However, these data sources play a limited role for the field researchers as the data resolution is low in some places. For instance, the resolution for the National Elevation Dataset is approximately 60 meters in Alaska area.

Meanwhile, transportation researchers have raised some methods to extract elevation from Google Earth (GE) for obtaining more reliable elevation data [18,19]. Factually, GE showed better data quality compared with other elevation datasets. However, the GE elevation may not reflect the complete ground-truth data on account of the following two reasons. First, as GE only records elevation for the buildings or structures on the top level, elevation data are prone to be inaccurate in the segments with overhead buildings or tunnels. Second, some researchers found that the images of the ground surface from GE may be distorted unexpectedly [20]. Thus, the elevation data for the corresponding segments are likely to be biased. Therefore, the applicability of GE elevation data may be largely confined for traffic practitioner, researchers and regulators. As a result, it is urgent to develop an efficient denoising method for the elevation dataset.

Data cleaning and denoising is one of the hot topics in many fields such as signal processing [21,22], image smoothing [23], multiple source data fusion [24]. The elevation data is one dimension data and outliers in the raw elevation data randomly scatter in the series. Such characteristic implies that smoothing algorithms for signal processing are more suitable for ruling out abnormal elevation data. Some popular signal denoising methods, involving empirical mode decomposition (EMD) and wavelet transform (WT), are potential choices for removing elevation noises. However, the EMD method may still leave some outliers after noise removal process because of mode mixing phenomenon [25,26]. The WT method is possible to introduce some unexpected fluctuation into the elevation data [27]. Therefore, the above-mentioned methods are not applicable for the smoothing purpose for GE elevation data.

Moving average (MA) is one of the widely-used methods for ruling out noises by averaging adjacent neighbours' values [28]. The method is efficient in smoothing out random and short-term intense fluctuation and maintain the longer tendency of the original data. The denoise logic of MA method suits for removing outliers in the raw elevation data. Hence, we employed MA model as one of potential workable scenarios for smoothing out the exceptional values in GE elevation. Additionally, Butterworth filter (BWF) is one of the best-known and popular infinite impulse response filters. The BWF was designed to achieve the smoothing effect by obtaining mild transition between stopband and passband as much as possible. The superiority of BWF is that the noises existing in the elevation data can be regarded as high frequency component and it is easier to distinguish the noise from the noise-free data [29-32]. Hence, the BWF is also the potential smoothing method in our study.

However, previous studies showed that the BWF has the over-smooth phenomenon. Specifically, some truly minor elevation variation is likely to be ruled out by the BWF method. To reduce such side effect and obtain the optimization smoothing result, we

integrated a local enhancement mechanism (LEM) with the digital BWF method. Specifically, the LEM calculated grades, from the raw elevation data, for all the segments and determined the local maximum and minimum grade values. Then, the elevation for the local extrema grades was proportionally revised by the LEM. Lastly, we employed the digital BWF to rule out outliers in the revised elevation data. The smoothing results by the proposed methodological framework showed that the BWF method successfully kept the truly smaller fluctuation in the raw elevation data. In addition, most obvious outliers and anomaly elevation data have been corrected. The research can help transportation administrative department and researchers build nationwide elevation databases for different road classes.

The remaining part of the paper is organized as follows. Section 2 briefly introduces elevation extraction process and typical outliers in the raw GE elevation data are illustrated in Section 3. Section 4 describes methodologies used for this study and smoothing results are revealed in detail in Section 5. We conclude the paper and provide future work in Section 6.

2. Elevation Data Extraction.

GE stores worldwide ground elevation and the data can be extracted from application programming interface (API) of GE. Smart Transportation Applications and Research Laboratory (STAR Lab) of the University of Washington has developed a Google Earth Elevation Data Extraction System (GEEDES) tool for the purpose of extracting expressway's elevation data. The elevation extraction process mainly includes four steps which have been described in detail in our previous work [20]. The concrete steps for elevation extraction are briefly illustrated as follows.

- **Step 1:** Acquiring freeway points' coordinates based on geographic information systems (GIS) map. The GIS map containing centerlines for both traffic directions is a prerequisite for Step 1. With the available GIS map, we acquired the relative coordinates' set by disintegrating the centerlines into discrete points. We employed all roads network of linear referenced data (ARNOLD) map as the GIS map in our study.
- **Step 2:** Determining parameters of GE viewbox based on spatial information. The step is to assure that routes of interest will be shown in accordance with GE form. This is done by transforming geometric information of sampling points into GE relative coordinates.
- **Step 3:** Implementing the procedure of elevation extracting. After obtaining the sampling points' coordinates, we used GEEDES tool to extract elevation for interesting points. The shortage of this method is that only top-level elevation can be obtained. So, the routes with high-rise infrastructure or under tunnels may be recorded with inaccurate elevation.
- **Step 4:** Computing the link gradient. Freeway link grade is important for the transportation researches and is a key factor in designing roadway vertical curve. For a given link, its gradient is computed as Equation (1).

$$Grade(i) = \frac{E_{End}(i) - E_{Start}(i)}{Length(i)} \tag{1}$$

where $E_{Start}(i)$ is the elevation of beginning point for link i. The $E_{End}(i)$ is the ending point's elevation of the link. Parameter $Grade(i)$ means the grade for the link i while $Length(i)$ is the total length of link i. A positive value of $Grade(i)$ represents an uphill slope while a negative value for downhill segment. Sampling interval of GE elevation is 10 ft per point along freeways in this study.

3. **Outliers in Raw GE Elevation.** For the sake of finding outliers in the raw elevation data, the research team developed a process by checking the reasonable grade for single link as well as grade variation between neighbouring freeway links. American Association of State Highway and Transportation Officials (AASHTO) published A Policy on Geometric Design of Highways and Streets, known as Green Book, which provides principles for determination of link's reasonable grade scopes [33]. The gradient values, beyond the reasonable range in the Green Book, are labelled as outliers. Concretely, 7% is the maximum grade which can be available in some mountainous regions and speed limit in the area should be less than 60 mph (95 km/h). What is more, 6% is the generally maximal grade existing in some hilly downtown regions and mountainous segments.

Additionally, the Green Book provides some advice for the allowed grade changing range along freeways which mediate some sag vertical curves as well as crest. The curve's changing rate can be diverse under different freeway speed limits. For instance, with a given segment at speed limit of 60 mph, its length should not be less than 151 ft for 1% grade variation. In other words, 0.0065% is the largest permitted gradient change for 10 ft (i.e., 1% grade change for 151 ft). This principle serves as standard for determining outliers in the raw elevation data. After data quality control process, we have determined six types of errors in the raw GE elevation dataset. Typical outliers and reasons are discussed as follows.

(1) Spikes in the Elevation Curves

After checking with GE in 3D view, it was found that some overlapping infrastructures, traffic signs, large-size trucks are the primary causes for the spikes. Figure 1(a) shows example of spikes and the reason for that is the overhead bridge. As only elevation of top layer is recorded in the GE, the elevation for the underneath freeway segment is presented by the overhead bridge elevation instead of ground truth elevation.

(2) Sudden Dips in the GE Road Surface

Another typical type of anomaly is the sudden dips existing in the original GE elevation dataset which are demonstrated in Figure 1(b). After searching the GE 3D view of such outliers' segments, it was determined that flyover regions, bridges crossing rivers and valleys were more likely to have such outlier. The elevation around these areas was apparently different from its neighbours.

(3) Buckling Road Surface

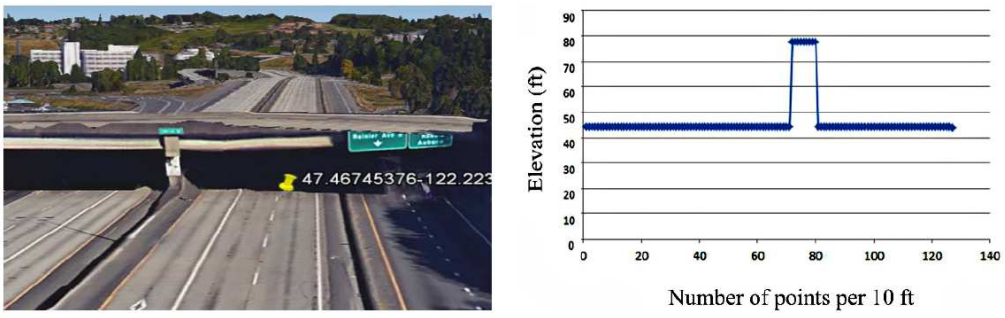
Some cases of buckling road surface are demonstrated in Figure 1(c). We have manually checked this category of outlier in different states and found that most of such outliers happened in hilly regions. As freeway street view from Google Map was recorded by onboard cameras, the images of buckling surface were taken as ground truth images. After inspecting a series of images for the outlier, we found that elevation recorded in GE is incorrect, which affirms the existence of outlier on buckling road surface.

(4) Surface Discontinuity at State and Road Boundaries

Elevation acquired from road surface that near boundaries of states may often include some errors when the boundaries are separated by rivers and valleys. As we do not have sufficient information about expressway elevation locating at the boundary of other States, it is not an easy task to rule out such errors. Similar problem also happens at road boundaries as they often begin at freeway intersections.

(5) Choppy Roadway Surface

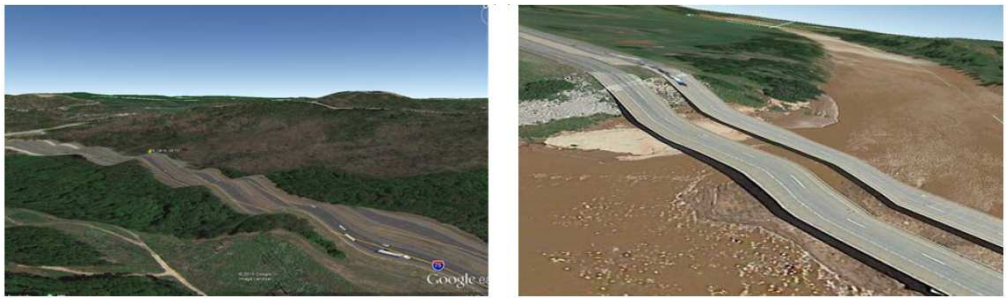
Some roadways appear in the form of successive choppy with small elevation fluctuations. Such examples have been shown in Figure 1(d). After checking many choppy segments in GE, we found that mountainous regions and bridges over rivers are the most likely areas containing such outliers. Elevation curve in these segments possesses some variations and abnormal grade change can be detected easily.



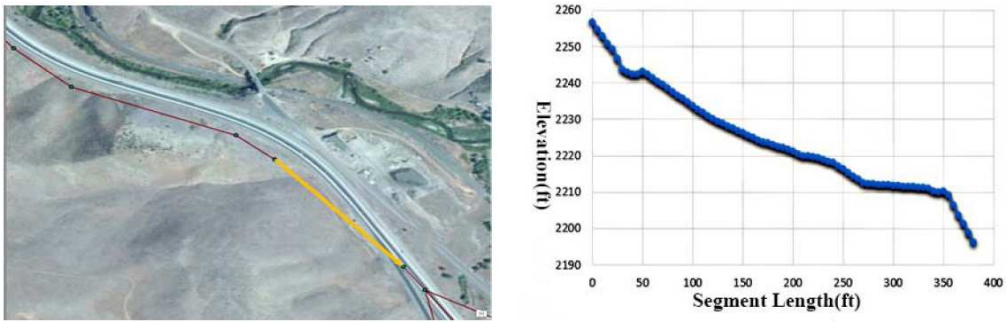
(a)

(b)

(c)



(d)



(e)

FIGURE 1. Examples of typical outliers in the GE highway elevation

(6) Quality Issue in the GIS Map

As we extracted freeway elevation based on roadway centerlines, data quality of the extracted elevation largely depends on GIS map marginally. It was found that some roadway centerlines do not match GIS roadway centerlines well in the ARNOLD GIS map (Figure 1(e)). In another word, elevation extracted from these areas may be not accurate.

4. **Methodology.**

4.1. **Elevation smoothing with LEM and BWF.**

4.1.1. *Elevation pre-process by the LEM.* Butterworth filter, proposed by a British engineer, is designed with the goal of obtaining smoothing frequency variation in the transition band. However, one of the drawbacks of the BWF is that higher order of BWF is the over-shoot phenomenon which degrades BWF performance [34]. In other words, true fluctuations in the elevation dataset may be wrongly replaced with flatted values by higher order of BWF.

To reduce the side effect of BWF, we integrated a local enhancement mechanism (LEM) with basic Butterworth filter to denoise the elevation data. Firstly, the LEM checked grade distribution for all segments. Secondly, the LEM determined local extrema gradient values and the corresponding elevation data. Thirdly, the segments' elevation was proportionally increased or decreased based on Equation (2). For instance, the original segment's elevation, with gradient between -0.02 and 0.02, is prone to be overfitted by the BWF. So, the LEM mechanism modified the corresponding elevation series so that the gradient was two-fold steeper than its initial grade at least. This operation intended to cancel out the over-smooth drawback of the BWF. The first calculation expression in Equation (2) demonstrated the process of obtaining the corresponding modified elevation data. Additionally, the rest expressions in Equation (2) showed specific calculation processes of substituting initial elevation for other grades. Segments, with the absolute grade values exceeding 7%, were labelled as significant outlier segment (O_{sig}). We kept the elevation data of O_{sig} segments as its original values.

$$Ele_{bw} = \begin{cases} (2 + 0.2 * rand) * grade(t) * \Delta mile + ele_{init} |grade(t)| \in [0, 0.02] \\ (1 + 0.1 * rand) * grade(t) * \Delta mile + ele_{init} |grade(t)| \in (0.02, 0.03] \\ (0.5 + 0.05 * rand) * grade(t) * \Delta mile + ele_{init} |grade(t)| \in (0.03, 0.05] \\ (0.1 + 0.01 * rand) * grade(t) * \Delta mile + ele_{init} |grade(t)| \in (0.05, 0.07] \\ ele_{init} grade(t) \text{ is } O_{sig} \end{cases} \tag{2}$$

where $grade(t)$ is the original gradient value for the tth segment. The length of the segment t is $\Delta mile$. The symbol $|grade(t)|$ represents the absolute value for the $grade(t)$. The ele_{init} is the initial elevation and Ele_{bw} is the corresponding modified elevation data.

4.1.2. *Elevation smooth by the BWF.* The above proposed LEM mechanism provided us the new elevation data where smaller fluctuation in the raw elevation data was intensified. With the help of the LEM mechanism, we can alleviate the shortage of the BWF method when smoothing the elevation series. The transfer function of the BWF was depicted in Equation (3).

$$H(z) = \frac{a_0 + a_1 z^{-1} + \cdots + a_k z^{-k}}{1 + b_1 z^{-1} + \cdots + b_k z^{-k}} \tag{3}$$

where a_i $(i = 0, 1, \ldots, k)$ and b_i $(i = 1, 2, \ldots, k)$ were the set of coefficients that decided frequency response of filter, k represented the order of filter. Equation (3) presented the

transformation function in the form of Z-domain from where we could find zeros and poles. In fact, zeros' points helped construct numerator and poles for building the denominator in transfer function. After obtaining the numerator and denominator, we could gain the smooth elevation data with formula shown in Equation (4).

$$Ele_{sm}(n) = \sum_{i=0}^{k} a_i * Ele_{bw}(n-i) - \sum_{i=1}^{k} b_i * Ele_{sm}(n-i) \tag{4}$$

The $Ele_{bw}(n-i)$ represented the above LEM modified elevation data and $Ele_{sm}(n-i)$ was the previous input smoothed elevation. Parameters of a_i and b_i were coefficients of filter in Equation (3). The $Ele_{sm}(n)$ was the output smoothed elevation data.

For the sake of finding a_i and b_i ($i = 0, 1, 2, \ldots k$) in Equations (3) and (4), we established similar transfer function as Equation (5). Then a relation between similar transfer function (see Equation (5)) and digital transfer function (see Equation (3)) was built through Equation (6) where C is a constant. Hence, we can get the coefficients of a_i and b_i by solving the simultaneous Equations (3), (5), (6) and (7).

$$G(p) = \frac{A_0 + A_1 p + \cdots + A_k p^k}{B_0 + B_1 p + \cdots + B_k p^k} \tag{5}$$

$$p = \frac{C\,(1 - z^{-1})}{1 + z^{-1}} \tag{6}$$

$$G(p) = H(z) \tag{7}$$

4.2. **MA method.** MA is a popular method to rule out singularities and maintain the trend of data series. The window size of MA is set differently according to different applications. The moving average method employed in our study for smoothing is an equal weighted method which calculates point elevation based on its neighbours. Given a freeway segment with the elevation series ele_{m-n}, $ele_{m-n+1}, \ldots, ele_{m+n}$, the elevation ele_{MA} was calculated as Equation (8). It is noted that smooth result is mainly affected by window size. Specifically, smaller window size cannot rule out the noises in raw elevation data completely while larger window size would neglect truly fluctuation details in the raw data.

$$ele_{MA} = \frac{ele_{m-n} + ele_{m-n+1} + \cdots + ele_{m+n}}{n} = \frac{1}{2n+1} \sum_{i=m-n}^{m+n} ele_i \tag{8}$$

4.3. **Goodness of fit.** To measure the denoising effect of BWF and MA, the smoothed elevation data has been compared with ground truth data under frequently used statistical indicators. As no ground-truth elevation data can be accessed in public, we took gradient data from Highway Safety Information System (HSIS) [35] as real data in the light of the following two reasons. Firstly, grade data for freeway segments is available in Washington State (WA). Second, the Washington State Department of Transportation (WSDOT) updates the freeway grade in the HSIS database every year since 2002. Thus, it is trustable to use the HSIS gradient data as ground truth data.

The indicators for evaluating the fitness effect included mean squared deviation (MSD), Pearson product-moment correlation coefficient (Pearson's r) and mean absolute deviation (MAD) [36-38]. For a given freeway segment t, $Grd_{smth}(t)$ and $Grd_{HSIS}(t)$ represented grade obtained from smoothed elevation and HSIS database, respectively. The mean value of $Grd_{smth}(t)$ and $Grd_{HSIS}(t)$ were denoted as $\overline{Grd_{smth}(t)}$ and $\overline{Grd_{HSIS}(t)}$ (see Equations

(9) and (10)). The expressions for MAD, MSD and Pearson's r were shown from Equations (11) to (13).

$$\overline{Grd_{smth}(t)} = \frac{1}{n}\sum_{i=1}^{n} Grd_{smth}(t) \tag{9}$$

$$\overline{Grd_{HSIS}(t)} = \frac{1}{n}\sum_{i=1}^{n} Grd_{HSIS}(t) \tag{10}$$

$$MAD = \frac{\sum_{t=1}^{n} |Grd_{smth}(t) - Grd_{HSIS}(t)|}{n} \tag{11}$$

$$MSD = \frac{\sum_{t=1}^{n} |Grd_{smth}(t) - Grd_{HSIS}(t)|^2}{n} \tag{12}$$

$$r = \frac{\sum_{i=1}^{n}\left[\left(Grd_{smth}(t) - \overline{Grd_{smth}(t)}\right) \times \left(Grd_{HSIS}(t) - \overline{Grd_{HSIS}(t)}\right)\right]}{\sqrt{\sum_{i=1}^{n}\left(Grd_{smth}(t) - \overline{Grd_{smth}(t)}\right)^2} \times \sqrt{\sum_{i=1}^{n}\left(Grd_{HSIS}(t) - \overline{Grd_{HSIS}(t)}\right)^2}} \tag{13}$$

5. **Results.** To evaluate the smoothing effects, we applied the proposed smoothing model and MA on the three typical highway segments which locate in WA. For simplicity, the proposed elevation smoothing framework is denoted as revised-BWF. The first highway segment is in the southbound of Interstate 5 with the milepost from 0 to 8.67 mile. The second segment is in the northbound of Interstate 405 from 0 to 10 miles. The last segment is in the westbound of Interstate 90 from 162 to 172 miles. The selected three segments possess different kinds of outliers which result in abnormal grades. Elevation of Interstate 5 has been modelled in detail for illustration of noise elimination, and Interstate 405 and 90 are then denoised for verification.

5.1. **Revised-BWF smooth result.** Two parameters are required for the revised-BWF algorithm which included cutoff frequency (CF) and order (Ord). In our study, elevation data sampling rate was settled as 1.90×10^{-3} which was calculated by 1 divided 528 (528 points were sampled per mile). After a series of experiments, we found that optimum CF was bounded with scope between 2.00×10^{-5} and 1.00×10^{-6} (with interval of 1.00×10^{-6}) while limits of Ord was from 1 to 8. As trial and error (TE) is the method which tries all possibilities to find optimum solution, it is suitable to find optimal revised-BWF parameter settings by searching all possible combinations of Ord and CF [39]. For each order of revised-BWF, we used TF rule to exhaust CF settings by comparing its smooth result against HSIS data. The revised-BWF settings and corresponding statistical indicators are shown in Figure 2 and Table 1.

It can be seen from Figure 2 that statistical indicators do not change obviously after Ord arrives 4. The first three rows of Table 1 demonstrated that the revised-BWF smoothed data is getting closer to HSIS data with increase of order. The best performance of revised-BWF is obtained when the Ord is 3 and CF is 1.40×10^{-5}. The minimum MSD was 2.93×10^{-5}, MAD was 4.27×10^{-3} and maximal Pearson's r was 0.98 which depicted that the revised-BWF smoothed elevation is close to HSIS data. The variation of statistical indicators, in Figure 2, confirmed our above analysis. Thus, it is reasonable to say that optimal revised-BWF smoothing elevation is acquired with parameter setting of Ord as 3 and CF as 1.40×10^{-5}.

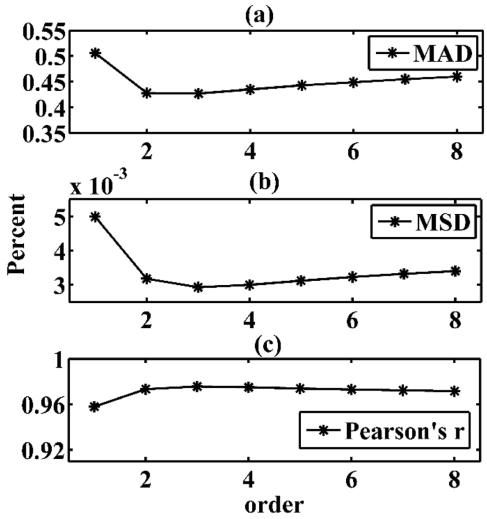


FIGURE 2. Tendency of statistical indicators of with different parameter settings for revised-BWF

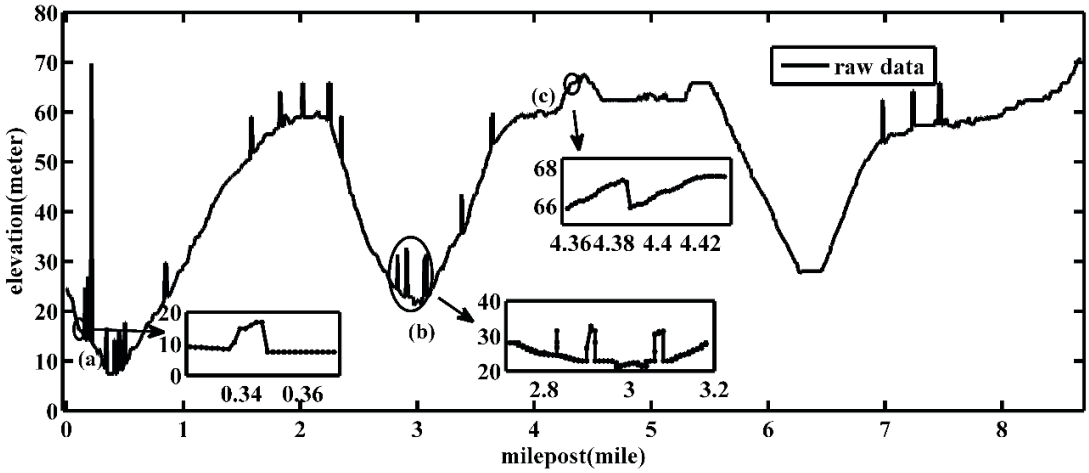
TABLE 1. Statistics indices for the revised-BWF results for Interstate 5 segment

Ord	CF	MAD	MSD	Pearson's r
1	1.00×10^{-5}	5.06×10^{-3}	4.99×10^{-5}	9.58×10^{-1}
2	1.30×10^{-5}	4.28×10^{-3}	3.18×10^{-5}	9.73×10^{-1}
3	$\mathbf{1.40 \times 10^{-5}}$	$\mathbf{4.27 \times 10^{-3}}$	$\mathbf{2.93 \times 10^{-5}}$	$\mathbf{9.76 \times 10^{-1}}$
4	1.40×10^{-5}	4.35×10^{-3}	3.00×10^{-5}	9.75×10^{-1}
5	1.40×10^{-5}	4.43×10^{-3}	3.12×10^{-5}	9.74×10^{-1}
6	1.40×10^{-5}	4.49×10^{-3}	3.23×10^{-5}	9.73×10^{-1}
7	1.40×10^{-5}	4.55×10^{-3}	3.32×10^{-5}	9.72×10^{-1}
8	1.40×10^{-5}	4.60×10^{-3}	3.40×10^{-5}	9.72×10^{-1}

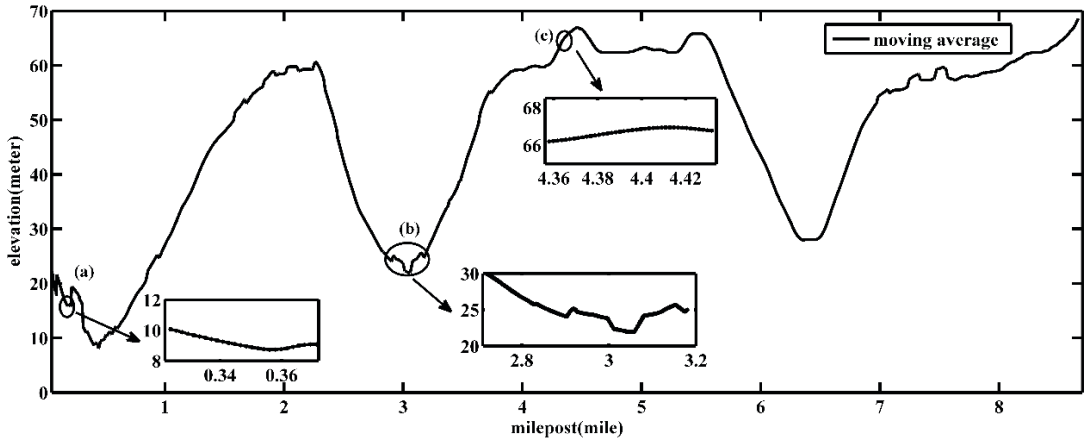
TABLE 2. Statistical performance of denoising with revised-BWF and MA for Interstate 5

Start point (mile)	End point (mile)	MSD		Pearson's r		MAD	
		MA	revised-BWF	MA	revised-BWF	MA	revised-BWF
0	8.67	1.81×10^{-4}	2.93×10^{-5}	8.54×10^{-1}	9.76×10^{-1}	9.00×10^{-3}	4.27×10^{-3}

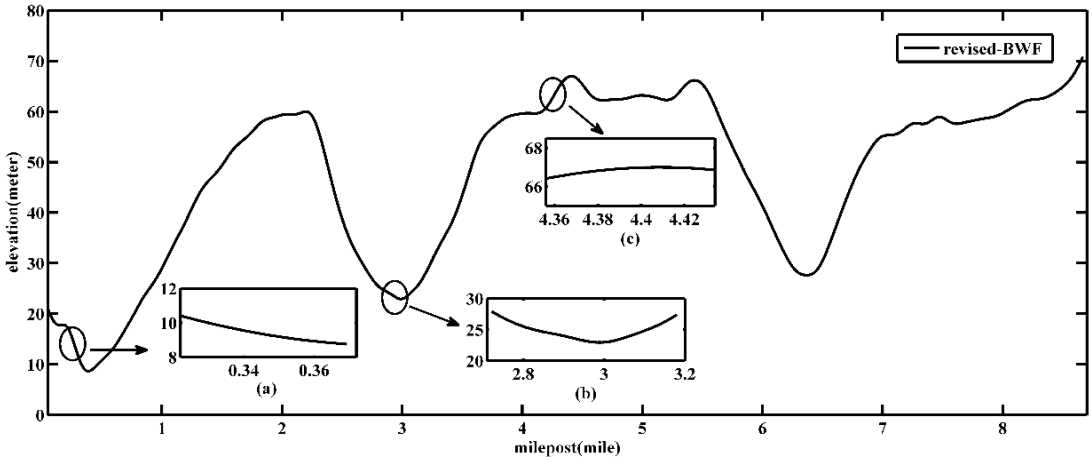
5.2. **Comparison between revised-BWF and MA.** The revised-BWF smoothing process illustrated above was used on Interstate 5 elevation data. In order to test validity of the revised-BWF, MA was used to denoise on the elevation data of the same Interstate 5 segment. We tested MA performance with window size changing from 50 ft to 2000 ft. The optimal MA result was acquired when the window size was set to 500 ft. The comparison of MA and BWF was displayed in Table 2 where the revised-BWF showed better performance. The statistical indicators for both revised-BWF and MA were displayed in Table 2. It showed that the revised-BWF model obtained better smoothing result than the MA method. Specifically, the MSD of the revised-BWF is 80% lower than the counterpart of MA while MAD for revised-BWF is less than half of the MA. In terms



(a) Raw elevation data



(b) Moving average smoothed elevation data



(c) Revised-BWF smoothed elevation data

FIGURE 3. Smoothing effects on elevation with revised-BWF and MA

of Pearson's r, elevation smoothed by the revised-BWF is 14% closer to HSIS data than MA-smoothed elevation.

In addition, we compared the performance of the two algorithms in removing the outliers in the elevation data for the Interstate 5. Elevation distribution of original data, MA smoothed data, the revised-BWF smoothed data are depicted from (a) to (c) of Figure 3.

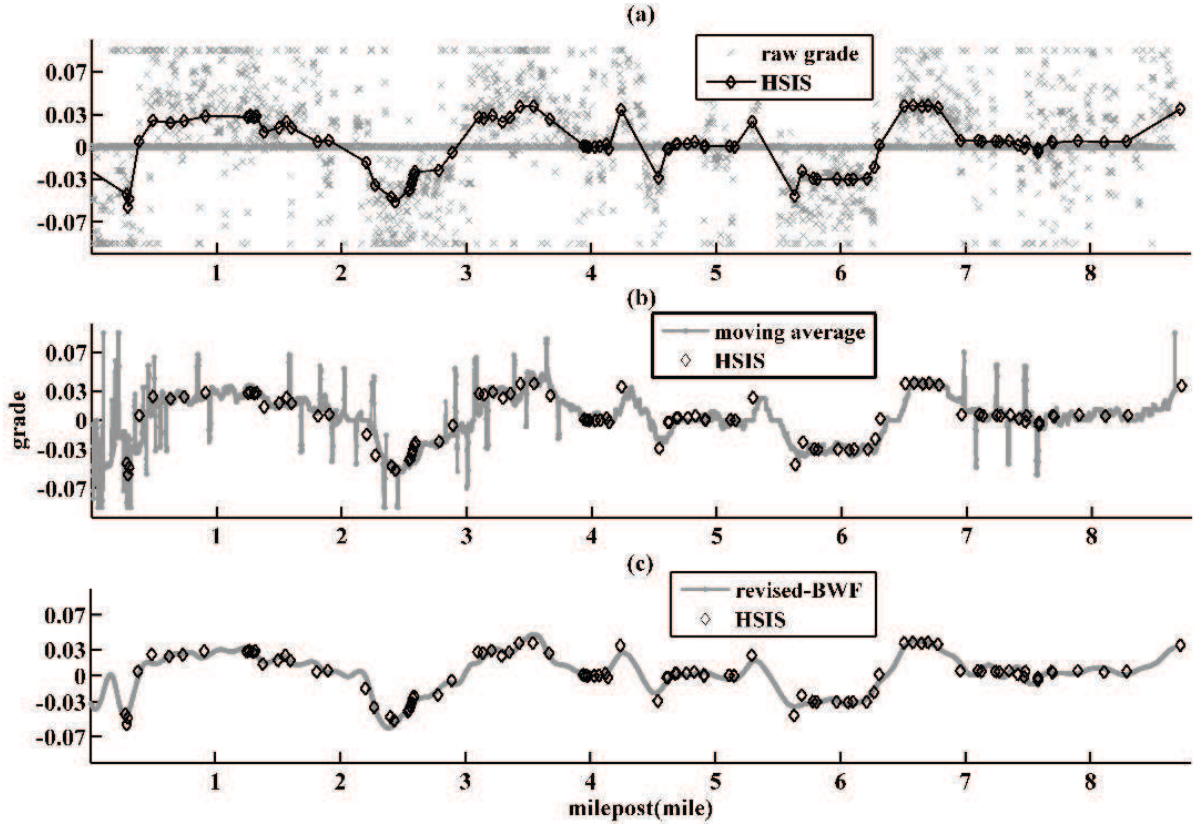


FIGURE 4. Smoothing effect on Grade with revised-BWF and MA

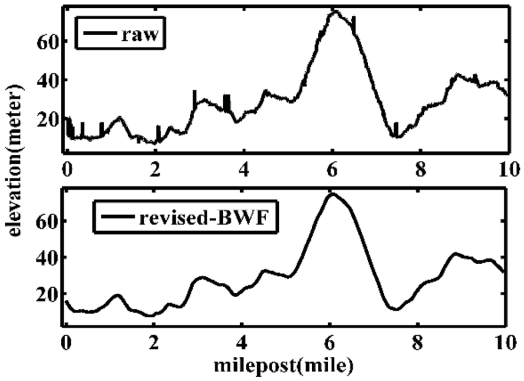
Different types of errors such as dips, choppy surface and spikes are found in the original elevation data in Figure 3(a). After smoothing process with MA algorithm, the elevation curve (see Figure 3(b)) is much flatter in comparison with Figure 3(a) and major outliers were removed. However, part of elevation curve still presents some abnormal elevation variation. With milepost ranging from 2.8 to 3.2 miles, the MA-elevation data still showed abnormal fluctuation. However, Figure 3(c) provided evidence that the revised-BWF discarded these anomalous data and substituted them with smoothing and reasonable elevation data.

What is more, we further verified the performance of revised-BWF and MA by examining the grade curves of revised-BWF and MA depicted in Figure 4. Positive grade in Figure 4 represents for uphill road segment while negative means downhill segment. For simplicity, grade exceeding 9% or lower than -9% will be set as 9% and -9% respectively. It can be seen that the grade distribution calculated by raw GE data do not confirm to HSIS data (ground truth data). The grade curve obtained from MA elevation data mainly fits with HSIS data. However, spikes can still be seen in the grade curve, from MA-smoothed elevation data. Some grade values are still beyond reasonable ranges. In another word, MA method fails to rule out all outliers in elevation data. Figure 4(c) shows grade curve obtained from revised-BWF smoothed elevation data which indicates the same data variation as HSIS data. Such grade distribution means the revised-BWF method performs better in eliminating anomalies in elevation data than MA algorithm.

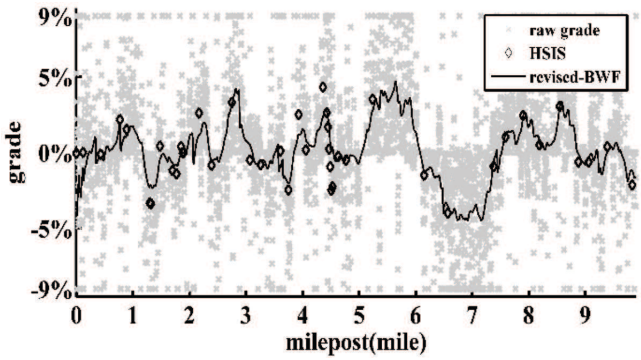
5.3. **Validation of smoothing result.** The revised-BWF has been applied to Interstate 405 and Interstate 90 for the purpose of performance comparison. We made use of the TE rule to choose optimal parameter settings of revised-BWF for Interstate 405 and Interstate 90 too. After testing all combinations, we determined CF and Ord as 1.10×10^{-5} and 1

for the sake of smoothing Interstate 405 elevation. The corresponding parameter settings for Interstate 90 were 9.00×10^{-6} and 8 respectively. The smoothed elevation and grade curves for both Interstate have been shown in Figure 5.

The raw Interstate 405 elevation curve, shown in Figure 5(a_1), has many spikes which were corrected by the revised-BWF method completely. Grade distributions of Interstate 405 obtained from raw GE elevation, revised-BWF and HSIS data shown in Figure 5(a_2) demonstrates that grade acquired from revised-BWF smoothed data is consistent well with HSIS data which is considered as ground truth data. Similar result can be drawn from Figure 5(b) which are smooth elevation and grade curves of Interstate 90. The smoothed elevation and grade data for Interstate 405 and 90 show that revised-BWF method can provide reliable elevation data.

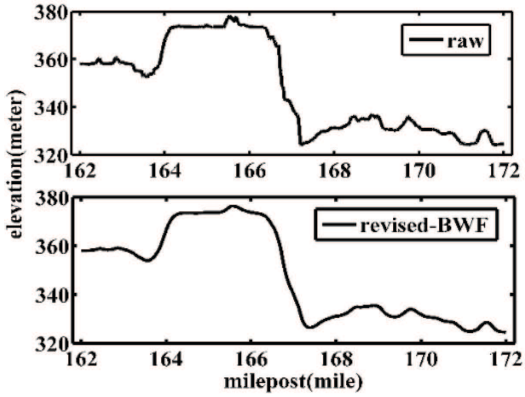


(a_1) Initial and revised-BWF smoothed elevation

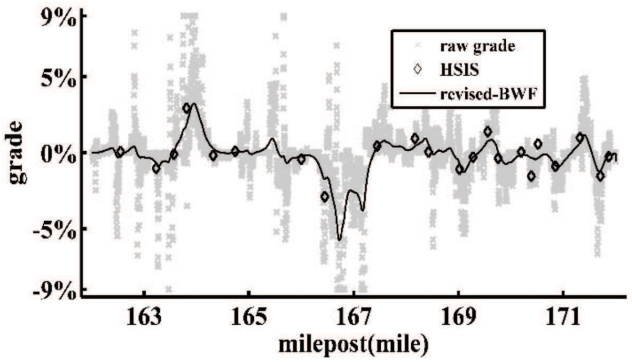


(a_2) Initial, HSIS and revised-BWF smoothed gradient

(a) Revised-BWF smooth result comparison for Interstate 405 segment



(b_1) Initial and revised-BWF smoothed elevation



(b_2) HSIS and revised-BWF smoothed gradient

(b) Revised-BWF smooth result comparison for Interstate 90 segment

FIGURE 5. Smoothed results comparison for Interstate 405 and Interstate 90 segments

The experimental results show the robust and satisfied performance for the proposed smoothing method. Roughly speaking, the revised-BWF ruled out sudden dips, bumps and other outliers successfully in the raw elevation data for the three freeway segments. Alternately, the denoised elevation data series, by the moving average method, still revealed some obvious noises which is shown in the subplot (b) of Figure 3. The reasons for the above phenomenon can be ascribed to the following aspects. First, the Butterworth filter in the proposed smooth framework is good at eliminating the elevation outliers. Factually, the revised-BWF tried to achieve the denoise result by obtaining the maximum

smoothing elevation data series. Second, the moving average method employed adjacent elevation to predict and obtain the smoothing result. Hence, the smoothing elevation is still the noisy series when minor outlier exists in the neighbouring elevation data.

What is more, the grade distribution, obtained from revised-BWF smoothed elevation, for Interstate 5, 90 and 405 segments agreed well with the ground truth data (HSIS data). Although, the potential drawback of the basic BWF is that the gentle elevation fluctuation may be ruled out incorrectly. The grade distributions demonstrate that the proposed LEM mechanism alleviate the drawback in the basic BWF filter effectively. Specifically, the LEM sharpened elevation variation for the segments' grades at local extrema values. The corresponding elevation values were enlarged or diminished by the LEM operation based on the grade values. This step preserved minor elevation fluctuation from being over-smoothed. In sum, the proposed elevation smooth framework achieved the aim of getting rid of abnormal elevation data without losing the elevation details.

6. **Conclusions.** This research assessed the performance of LEM-aided BWF for ruling out outliers of GE elevation data. The raw elevation data, extracted from the GE, showed obvious typical outliers and anomalies. These outliers limit further usage of GE elevation for both transportation engineers and researchers. Hence, we proposed a revised-BWF smoothing algorithm to denoise the elevation data. The LEM was introduced to modify the raw GE elevation data to avoid the BWF's over-smooth drawback. The optimal smoothing parameters for BWF were determined by the train-and-error rule. The revised-BWF performance has been evaluated on typical freeway segments of Interstate 5, 405 and 90. For the sake of contrast, moving average was used to smooth the same segments as those of the revised-BWF. The HSIS data was employed as ground truth data for performance comparison. The results showed that statistical indicators of MSD and MAD for the revised-BWF were much smaller than the counterparts of moving average. The Pearson's r for the revised-BWF were much bigger than that of the moving average. The results suggest that BWF can remove noises existing in raw elevation data efficiently. The method proposed by this research can help traffic management departments and field researchers obtain high quality elevation dataset. With such data, future research could concentrate on evaluating the influence of elevation on fuel consumption, traffic flow capacity and so on.

Acknowledgement. The authors appreciate the project support from the Federal Highway Administration (TOPR 33-01-14034), China Postdoctoral Science Foundation (Grant No. 2015M581585), and the National Natural Science Foundation of China (51579143, 51379121, 61304230, 51508094, 51508122), Shanghai Shuguang Plan Project (No: 15SG44) and the Natural Science Foundation of Jiangsu (BK20150612) and China Scholarship Council (CSC).

REFERENCES

[1] P. Lu and D. Tolliver, Accident prediction model for public highway-rail grade crossings, *Accident Analysis & Prevention*, vol.90, pp.73-81, 2016.

[2] S. Chen, F. Chen and J. Wu, Multi-scale traffic safety and operational performance study of large trucks on mountainous interstate highway, *Accident Analysis & Prevention*, vol.43, pp.429-438, 2011.

[3] W. Zhu, B. Wright, Z. Li, Y. Wang and Z. Pu, Analyzing the impact of grade on fuel consumption for the national interstate highway system, *Transportation Research Board 95th Annual Meeting*, 2016.

[4] K. Boriboonsomsin and M. Barth, Impacts of road grade on fuel consumption and carbon dioxide emissions evidenced by use of advanced navigation systems, *Transportation Research Record: Journal of the Transportation Research Board*, pp.21-30, 2009.

[5] R. Luque and M. Castro, Highway geometric design consistency: Speed models and local or global assessment, *International Journal of Civil Engineering*, vol.14, pp.347-355, 2016.
[6] Z. Jiang, K. Jadaan and Y. Ouyang, *Speed Harmonization – Design Speed vs. Operating Speed*, Illinois Center for Transportation/Illinois Department of Transportation 0197-9191, 2016.
[7] L. Wang, Q. Shi and M. Abdel-Aty, Predicting crashes on expressway ramps with real-time traffic and weather data, *Transportation Research Record: Journal of the Transportation Research Board*, pp.32-38, 2015.
[8] L. Wang, M. Abdel-Aty, Q. Shi and J. Park, Real-time crash prediction for expressway weaving segments, *Transportation Research Part C: Emerging Technologies*, vol.61, pp.1-10, 2015.
[9] M. Banihashemi, Is horizontal curvature a significant factor of safety in rural multilane highways?, *Transportation Research Record: Journal of the Transportation Research Board*, pp.50-56, 2015.
[10] E. Wood, E. Burton, A. Duran and J. Gonder, Contribution of road grade to the energy use of modern automobiles across large datasets of real-world drive cycles, *SAE Technical Paper 0148-7191*, 2014.
[11] E. Wood, E. Burton, A. Duran and J. Gonder, Appending high resolution elevation data to GPS speed traces for vehicle energy modeling and simulation, *National Renewable Energy Laboratory Technical Report NREL/TP-5400-61109*, 2014.
[12] O. Travesset-Baro, M. Rosas-Casals and E. Jover, Transport energy consumption in mountainous roads. A comparative case study for internal combustion engines and electric vehicles in Andorra, *Transportation Research Part D: Transport and Environment*, vol.34, pp.16-26, 2015.
[13] D. B. Gesch, M. J. Oimoen and G. A. Evans, Accuracy assessment of the US geological survey national elevation dataset, and comparison with other large-area elevation datasets: SRTM and ASTER, *US Geological Survey 2331-1258*, 2014.
[14] USGS (1996), *Global 30 Arc-Second Elevation (GTOPO30)*, https://lta.cr.usgs.gov/GTOPO30, 2014.
[15] T. Tachikawa, M. Kaku, A. Iwasaki, D. B. Gesch, M. J. Oimoen, Z. Zhang et al., *ASTER Global Digital Elevation Model Version 2 – Summary of Validation Results*, Technical Report, 2011.
[16] T. G. Farr and M. Kobrick, *The Shuttle Radar Topography Mission*, DTIC Document2000, 2000.
[17] S. E. Reutebuch, H.-E. Andersen and R. J. McGaughey, Light detection and ranging (LIDAR): An emerging tool for multiple resource inventory, *Journal of Forestry*, vol.103, pp.286-292, 2005.
[18] D. Potere, Horizontal positional accuracy of Google Earth's high-resolution imagery archive, *Sensors*, vol.8, pp.7973-7981, 2008.
[19] S. C. Benker, R. P. Langford and T. L. Pavlis, Positional accuracy of the Google Earth terrain model derived from stratigraphic unconformities in the Big Bend region, Texas, USA, *Geocarto International*, vol.26, pp.291-303, 2011.
[20] Y. Wang, Y. Zou, K. Henrickson, Y. Wang, J. Tang and B.-J. Park, Google earth elevation data extraction and accuracy assessment for transportation applications, *PloS One*, vol.12, p.e0175756, 2017.
[21] X. Dong, D. Thanou, P. Frossard and P. Vandergheynst, Learning laplacian matrix in smooth graph signal representations, *IEEE Trans. Signal Processing*, vol.64, pp.6160-6173, 2016.
[22] K. Harikumar, S. Athira, Y. C. Nair, V. Sowmya and K. Soman, ADMM based algorithm for spike and smooth signal separation using over-complete dictionary, *International Conference on Communications and Signal Processing (ICCSP)*, pp.1617-1622, 2015.
[23] J. Jiang, J. Ma, C. Chen, X. Jiang and Z. Wang, Noise robust face image super-resolution through smooth sparse representation, *IEEE Trans. Cybernetics*, 2016.
[24] J. Chen, K. H. Low, Y. Yao and P. Jaillet, Gaussian process decentralized data fusion and active sensing for spatiotemporal traffic modeling and prediction in mobility-on-demand systems, *IEEE Trans. Automation Science and Engineering*, vol.12, pp.901-921, 2015.
[25] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, pp.903-995, 1998.
[26] N. E. Huang, M.-L. C. Wu, S. R. Long, S. S. Shen, W. Qu, P. Gloersen et al., A confidence limit for the empirical mode decomposition and Hilbert spectral analysis, *Proc. of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, pp.2317-2345, 2003.
[27] B. Qiu, M. Feng and Z. Tang, A simple smoother based on continuous wavelet transform: Comparative evaluation based on the fidelity, smoothness and efficiency in phenological estimation, *International Journal of Applied Earth Observation and Geoinformation*, vol.47, pp.91-101, 2016.

[28] D. Krige, Two-dimensional weighted moving average trend surfaces for ore-evaluation, *Journal of the South African Institute of Mining and Metallurgy*, vol.66, pp.13-38, 1966.
[29] I. W. Selesnick and C. S. Burrus, Generalized digital butterworth filter design, *IEEE Trans. Signal Processing*, vol.46, pp.1688-1694, 1998.
[30] S. Butterworth, On the theory of filter amplifiers, *Wireless Engineer*, vol.7, pp.536-541, 1930.
[31] K. Henderson and W. Kautz, Transient responses of conventional filters, *IRE Trans. Circuit Theory*, vol.5, pp.333-347, 1958.
[32] P. Mitros, Filters with decreased passband error, *IEEE Trans. Circuits and Systems II: Express Briefs*, vol.63, pp.131-135, 2016.
[33] American Association of State Highway, Transportation Officials, *A Policy on Geometric Design of Highways and Streets*, 6th Edition, 2011.
[34] T. Khanna and D. K. Upadhyay, Design and realization of fractional order butterworth low pass filters, *International Conference on Signal Processing, Computing and Control (ISPCC)*, pp.356-361, 2015.
[35] *HSIS*, http://www.hsisinfo.org/index.cfm, 2016.
[36] *Mean Absolute Deviation (MAD)*, https://en.wikipedia.org/wiki/Average_absolute_deviation, 2016.
[37] *Mean Square Deviation (MSD)*, https://en.wikipedia.org/wiki/Mean_squared_error, 2016.
[38] *Pearson Correlation Coefficient*, https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient, 2016.
[39] *Trial and Error*, https://en.wikipedia.org/wiki/Trial_and_error, 2016.