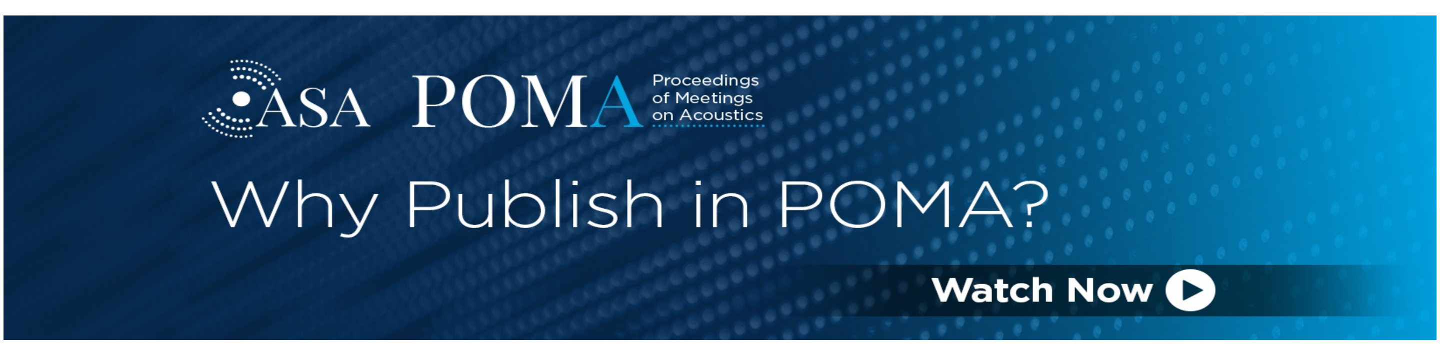

ASA POMA Proceedings of Meetings on Acoustics
Why Publish in POMA?
Watch Now

POMA Proceedings of Meetings on Acoustics

Volume 45 http://acousticalsociety.org/

181st Meeting of the Acoustical Society of America

Seattle, Washington

29 November -3 December 2021

Signal Processing in Acoustics: Paper 2aSP8

Toward improving road traffic noise characterization: A reduced-order model for representing hourly traffic volume dynamics

Mylan R. Cook, Kent L. Gee and Mark K. Transtrum
Department of Physics and Astronomy, Brigham Young University, Provo, Utah, 84602; mylan.cook@gmail.com; kentgee@byu.edu, mkt24@byu.edu

Shane V. Lympany and Matthew F. Calton
Blue Ridge Research and Consulting LLC, Asheville, NC, 28801; shane.lympany@blueridgeresearch.com; matt.calton@blueridgeresearch.com

The National Transportation Noise Map predicts time-averaged road traffic noise across the continental United States (CONUS) based on average annual daily traffic counts. However, traffic counts may vary significantly with time. Since traffic noise is correlated with traffic counts, a more detailed temporal representation of traffic noise requires knowledge of the time-varying traffic counts. Each year, the Federal Highway Administration tabulates the hourly traffic counts recorded at more than 5000 traffic monitoring sites across CONUS. Each site records up to 8760 traffic counts corresponding to each hour of the year. The hourly traffic counts can be treated as time-dependent signals upon which signal processing techniques can be applied. First, Fourier analysis is used to find the daily, weekly, and yearly temporal cycles present at each traffic monitoring site. Next, principal component analysis is applied to the peaks in the Fourier spectra. A reduced-order model using only nine principal components represents much of the temporal variability in traffic counts while requiring only 0.1% as many values as the original hourly traffic counts. This reduced-order model can be used in conjunction with sound mapping tools to predict traffic noise on hourly, rather than time-averaged, timescales. [Work supported by U.S. Army SBIR.]

Published by the Acoustical Society of America

1. INTRODUCTION

Road noise can be a significant component of total anthropogenic noise in many developed areas and can have a large impact on diverse acoustic environments. Not only humans are adversely affected by loud road noise,[1] but also other species.[2,3] Road noise cannot be effectively measured along every roadside, so accurate modeling of road noise is necessary for improving road noise characterization.

Because overall traffic noise is correlated with traffic volume rates, traffic noise characterization can depend heavily on characterization of traffic volume, along with other parameters such as vehicle speed, pavement type, road inclination, and land cover[4,5]. The National Transportation Noise Map published by the Bureau of Transportation Statistics uses annual average daily traffic (AADT) counts to predict an annually and daily averaged sound level near major roads across the continental United States (CONUS)[6]. While this map is useful for determining average sound levels, it lacks temporal variability, and so may not reflect the actual sound level for a particular time of the day or night.

Traffic volume can show large variation, not only diurnally but also from weekday to weekend and from summer to winter. A dynamically varying representation of traffic noise would be useful for determining hourly sound levels across CONUS. Characterizing the dynamic nature of traffic volume is not only useful for determining the changes in sound levels at locations where traffic counts are known but can also lead to a model for predicting variable traffic volume at other locations, and therefore predicting sound levels across CONUS for particular time periods.

Modeling traffic volume can be done in various ways, though a simple model can be made using vehicle counts at various locations[7,8]. Each year, the Federal Highway Administration tabulates hourly traffic counts recorded at thousands of traffic monitoring sites across the United States. Each site can record up to 8760 traffic counts corresponding to each hour of the year (or 8784 for leap years). While this gives a detailed representation of hourly traffic volume for individual sites, this representation requires thousands of individual hourly counts for each location, and using vehicle counts to predict counts at other times or at other locations is not straightforward. Methods such as wavelet decomposition[9] can be used to include temporal variability. In this paper, another possible method that can be used to predict hourly vehicle counts, hereafter used synonymously with traffic volume, at other locations is considered.

Treating the vehicle volume as a time-dependent signal enables signal processing techniques to be applied. The approach explained in this paper uses Fourier analysis and principal component analysis to give a reduced-order model to represent and predict vehicle counts[10]. This representation requires a total of only nine variables at a site to determine hourly traffic counts for any day of any year, a total of 0.1% as many variables as the hourly counts for a single year.

The approach for this paper uses hourly-resolution vehicle counts, and does not consider shorter-term traffic volume behavior, though it may be possible to adapt this approach for shorter timescales.[11-13] This approach also creates a method to predict hourly vehicle counts at other locations using geospatial or location-specific features. This reduced-order model can be used in conjunction with sound mapping tools to predict traffic noise across the continent on hourly, rather than time-averaged, timescales.[14]

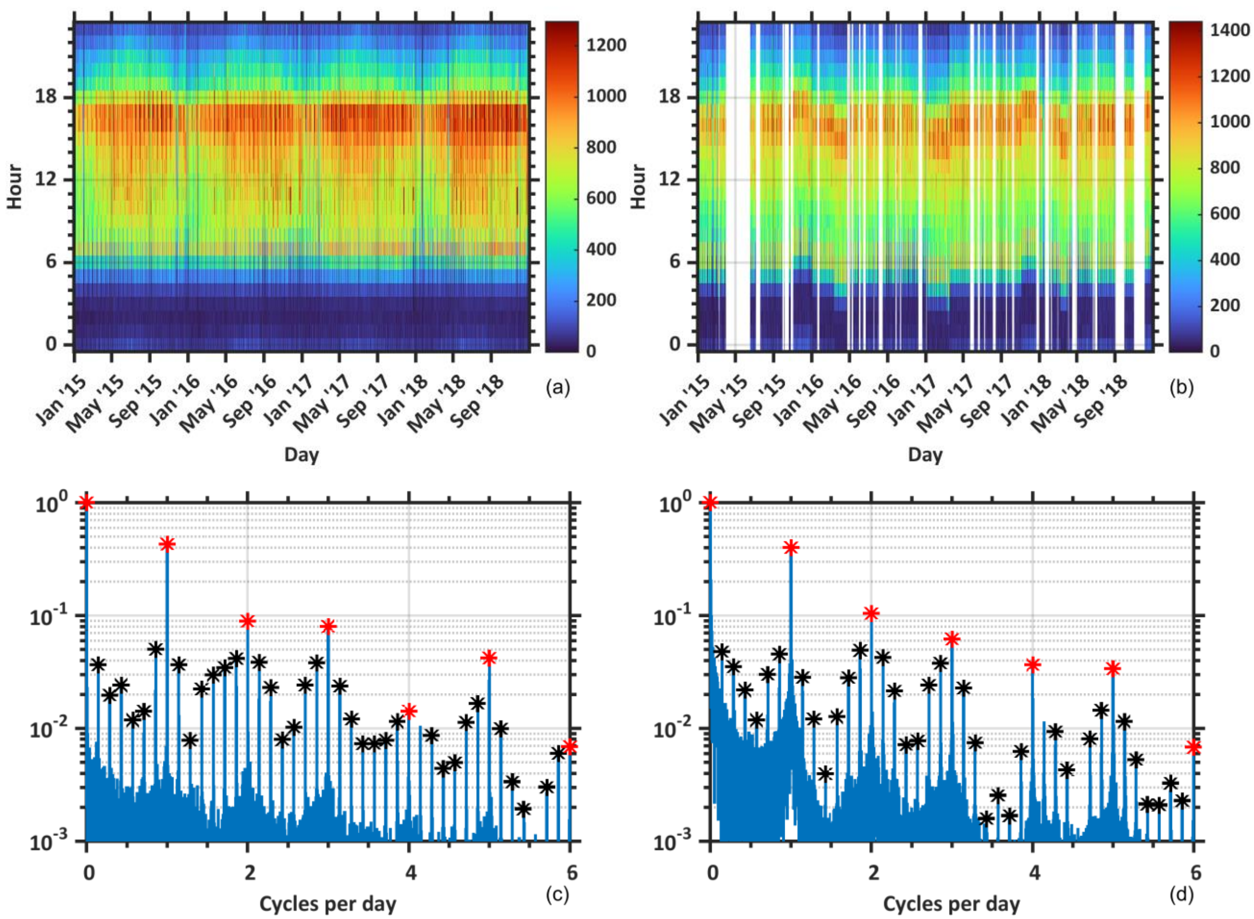
2. DATA USED FOR ANALYSES

The Federal Highway Administration obtains hourly traffic counts from thousands of traffic monitoring sites across the United States. Data reported between 1 January 2015 and 31 December 2018 are used in this paper. However, not all stations report counts for each hour during this time period. Additionally, some of the data reported contain erroneous counts. Initial automatic checks found probable errors in data reported at several sites, which were entirely removed from the dataset. The remaining data come from 5695 sites across CONUS and are further analyzed for useability.

Within the remaining data set, some sites can still contain possible erroneous counts, and site-reported data are not equally reliable. For sites that report data consistently, reported values are more reliable, while sites that only report counts intermittently can contain dubious values. This can be seen in the data counts shown in Figure 1. The data pictured in Figure 1(a) contain values for every hour of the 4-year period and reported counts show high consistency. For the data pictured in Figure 1(b), however, shows several gaps where days, weeks, or even months of hourly counts were not reported; additionally, the reported counts for some months seem to be shifted by up to a few hours. To mitigate doubts in data fidelity, and therefore in results, a further data validation step is necessary. A site data confidence weighting is therefore used.

Generally, weights can be assigned based on the percentage of data present for each site. Sites with half as many reported hourly counts could be given half the weight as sites where all four years of hourly counts were reported. However, for processing purposes explained further on, cyclically missing data, or data missing between times with reported data, can have a negative impact on analysis. Therefore, an adjusted weighting is made based not on the total number of hourly counts reported, but by the maximum number of *consecutive* hourly counts reported for each site. By weighting data in this manner, each site is given a relative weighting between zero and one.

Figure 1. Reported hourly traffic counts for two locations (top). The horizontal axis shows the day of the year, while the vertical axis shows the hours of each day, with the traffic count represented by the color. The normalized Fourier transform of the vehicle counts for both locations, explained in Section 3, is also shown (bottom). Amplitudes for integer multiples of daily cycles are marked in red, and amplitudes for integer multiples of weekly cycles are marked in black.



3. FOURIER ANALYSIS OF TRAFFIC VOLUME

While visual patterns are seen when viewing traffic volume in Figure 1(a-b) (such as the diurnal or day/night patterns), in order to make use of this information—such as to predict traffic volume at other locations, or to predict missing vehicle counts like the gaps seen in Figure 1(b)—it is necessary to characterize these traffic volume patterns. By treating hourly traffic counts as time-dependent signals, Fourier analysis is used to identify the temporal cycles present at each individual site[15]. However, because a fast or discrete Fourier transform generally requires equally spaced and non-missing data, and because several sites contain some missing vehicle counts, a non-uniform Fourier transform must be used.

A. NON-UNIFORM FOURIER TRANSFORM

Fourier transforms usually require equally spaced temporal data, with the sampling frequency determining the maximum number of cycles that can be found for a time period. Most discrete Fourier transform algorithms are ill-equipped to handle missing data, and so instead the non-uniform discrete fast Fourier transform is used to analyze the traffic counts at each site[16]. This yields a two-sided, complex-valued Fourier spectrum. The magnitude of the single-sided spectrum for two sites is shown in Figure 1(c-d).

Clear peaks in the Fourier spectra appear at several relevant frequencies. One important peak is that seen at 0 cycles per day, which represents the average number of measured vehicles per hour (the AADT divided by 24), the value of which has been used to normalize the plotted spectra. Other strong peaks are seen at integer multiples of 1 cycle per day (marked in red), which together represent the daily-repeating traffic volume pattern. More peaks are seen at integer multiples of 1/7 cycle per day (marked in black), and together represent the weekly-repeating traffic volume pattern. Though not visible on this scale, there are also peaks at integer multiples of one cycle per year.

B. DENOISING FOURIER SPECTRA

Peaks in the Fourier spectra on integer multiples of daily, weekly, and yearly time periods are often quite pronounced, but the spectra can also contain non-zero amplitudes at other time period cycles. This is because the traffic pattern does not repeat exactly every day, week, or year. Because the present research is concerned with average hourly behaviors, rather than a precise representation of the traffic counts, the small but non-zero amplitudes are treated as noise in the Fourier spectra, and so are removed or zeroed out.

A representation of the average traffic volume pattern for a site is obtained using the Fourier amplitudes at frequencies that are integer multiples of weekly and yearly cycles (daily cycle frequencies are captured using integer multiples of weekly cycles). Due to missing data or machine precision errors, peaks in the Fourier spectra can sometimes be found at frequencies that are adjacent to multiples of weekly and yearly cycles; when this is the case, the peak amplitude at that frequency is used instead of the Fourier amplitude at the integer-multiple frequency. This representation serves to decompose hourly traffic counts into a combination of sinusoidal traffic patterns that repeat on weekly and yearly cycles. This representation has the benefit of removing noise in traffic counts and requires a few hundred values instead of thousands of individual hourly vehicle counts.

C. PEAK VALUES REPRESENTATION

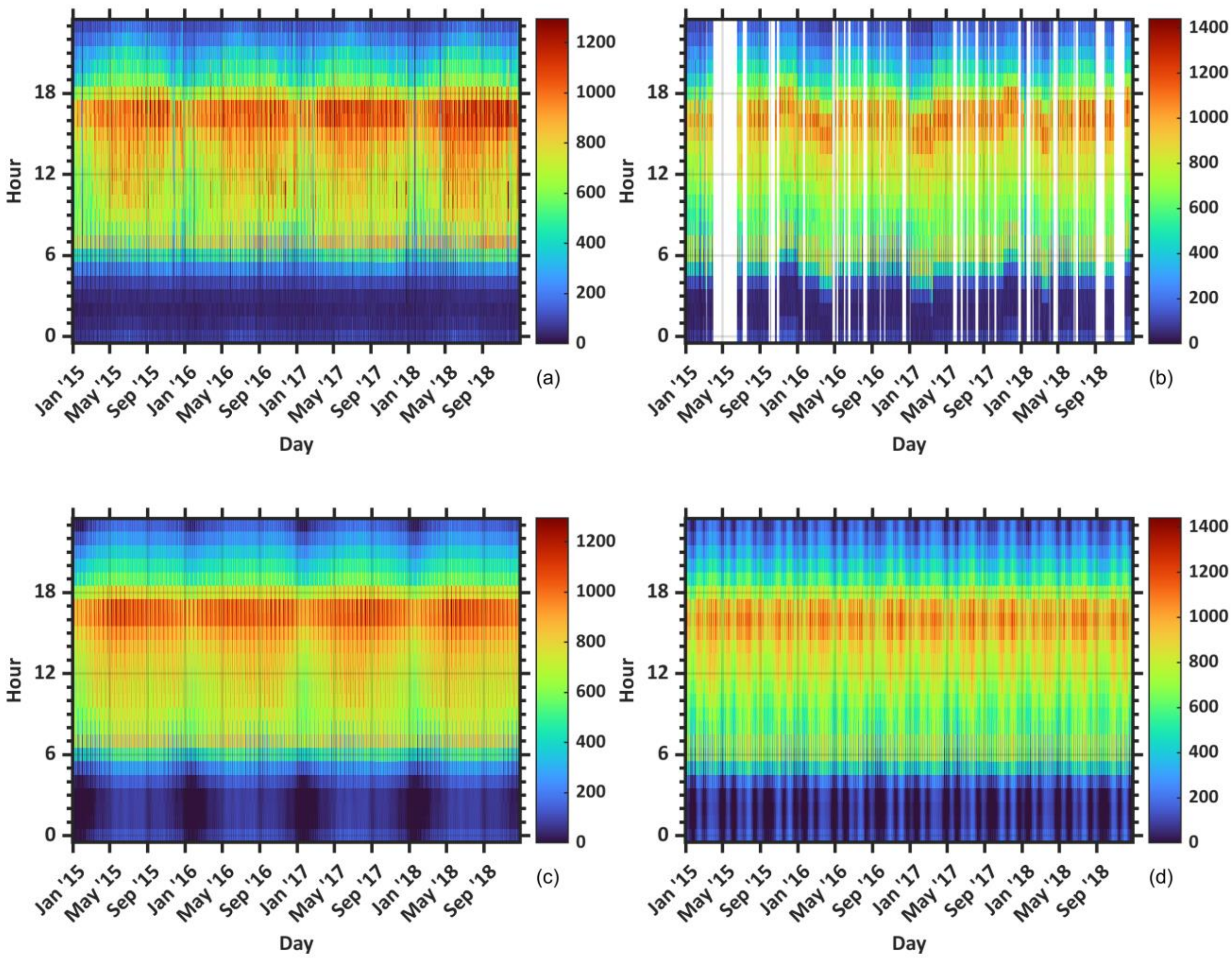
A smoothed, average hourly traffic volume pattern for each site is obtained by using an inverse Fourier transform on the denoised Fourier spectrum peak values. For the reported traffic counts shown in Figure 2(a), this approach yields a reasonable approximation as shown in Figure 2(c). However, for the reported counts shown in Figure 2(b), the representation shown in Figure 2(d) is not especially accurate. Some of the main reasons for the errors in the representation are considered below.

One reason for inaccuracy in the representation is caused by missing hourly counts at individual sites. That is why Figure 2(d) is not an accurate representation of the data in Figure 2(b). Missing data causes noise in the Fourier spectrum, which can alter the amplitude of the Fourier peaks which typically occur on integer multiples of weekly and yearly time cycles. The site weighting discussed previously, while not altering the behavior at an individual location, is important when looking at trends across sites, and will be discussed further on.

Another reason for inaccuracy is because, while traffic patterns contain both weekly and yearly repeating patterns, the days of the week do not match the day of the year, e. g., the first day of 2015 was a Thursday while the first day of 2017 was a Sunday. For generality, and to allow for traffic volume prediction for any day of the week of any year, the weekly traffic pattern is separated from the yearly traffic pattern. This is done by separating each Fourier spectrum into two separate data sets, one containing peaks on integer multiples of one cycle per year (up to 12 cycles per year, as after this the peaks are effectively in the noise floor), and the other containing peaks on integer multiples of 1/7 cycle per day (including the peak at 0 cycles per day so the average vehicle count is not removed). This ensures that the former will result in a yearly-repeating traffic pattern, and the latter in a weekly-repeating traffic pattern. These two patterns can then be combined to represent the traffic volume for any specific day of the week and day of the year.

A third reason for inaccuracy is that peak amplitudes in the Fourier spectrum are not independent of one another, and so noise in the Fourier spectrum that affects a single peak value, even by a small amount, can fundamentally change the overall temporal pattern found after performing an inverse Fourier transform. To state this in another way, a weekly-repeating traffic pattern is not determined solely by a single Fourier peak amplitude, and any irregularities in the data can cause changes in a single Fourier peak amplitude. For this reason, representing traffic volume solely as the sum of independent sinusoidal patterns is not the best representation. A better and more concise way to represent traffic volume is by using principal component coefficients of cyclic traffic patterns.

Figure 2. A comparison of the reported traffic volume for two sites (top) with the traffic volume calculated using the Fourier peak values representation (bottom).



4. PRINCIPAL COMPONENT ALANYSIS OF FOURIER SPECTRA

Principal component analysis (PCA) is used to find a lower-dimensional basis which can represent the majority of multi-dimensional data points of a set. For the current research, each dimension of a data point consists of the Fourier amplitude at a particular frequency. By splitting the Fourier peaks into separate weekly and yearly cycles, we create two separate data sets, each containing a point in a high-dimensional space for each traffic measurement site. Because the Fourier peak amplitudes are interdependent, principal component analysis is used to find a simpler basis to represent the most common combinations of cyclic traffic patterns found across sites. This serves to reduce irregularities in the traffic pattern at a particular site by using the common traffic patterns seen across multiple sites. What this means is that by using this approach, the weekly or yearly traffic pattern at any particular site can be represented as a combination of the most common cyclic weekly or yearly traffic patterns found across all sites.

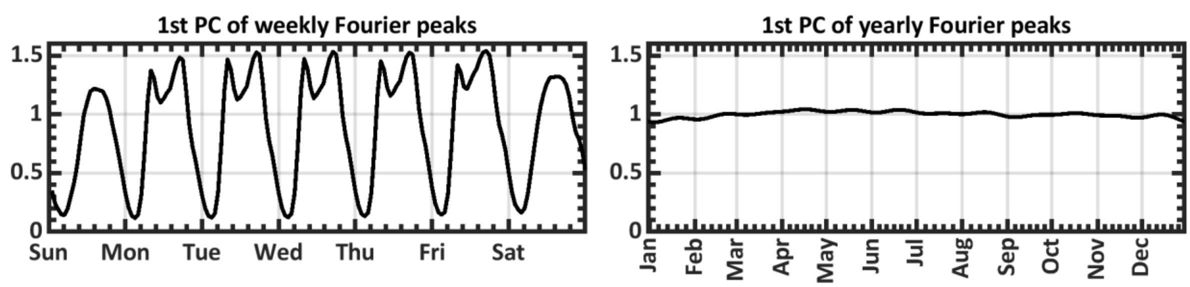
As mentioned in Section 2, the hourly counts from all sites are not equally reliable, and therefore the Fourier peak amplitudes at one site are not as accurate as they are for another site. To avoid propagating these errors, a weighted PCA is used to improve the accuracy of results, where the weighting used is described in Section 2. The analysis is performed separately for the weekly and yearly cycles. The analysis returns principal components, which each represents a linear combination of Fourier amplitudes or, by using an inverse Fourier transform, a specific traffic volume pattern. The principal component coefficients, which are simply numeric values, give the linear combination of these traffic volume patterns.

The first principal component vectors, one for the weekly data and another for the yearly data, are shown in temporal space in Figure 3, and give the weighted average traffic volume pattern seen across sites. The yearly pattern shows little variation across the course of a year. The weekly pattern shows the average hourly traffic pattern found across CONUS, namely one where weekends show a smooth hourly variation during daytime hours, while the weekdays show an increase in morning and evening hours higher than that during the middle of the day, with less traffic activity during the nighttime hours. This type of traffic pattern is common for several urban locations, and in particular shows high similarity to data shown in Fu et al. (2017).[17]

To represent different traffic patterns seen at other sites, especially rural locations, a few more principal components. By using an elbow analysis on the eigenvalue of the principal components, it was determined that six principal components should be used to represent the weekly traffic pattern, and four principal components to represent the yearly traffic pattern. In this manner, 73% of the weekly data is represented, and 83% of the yearly data. The unique traffic pattern at a site can then be represented using just 10 principal component coefficients (PCCs).

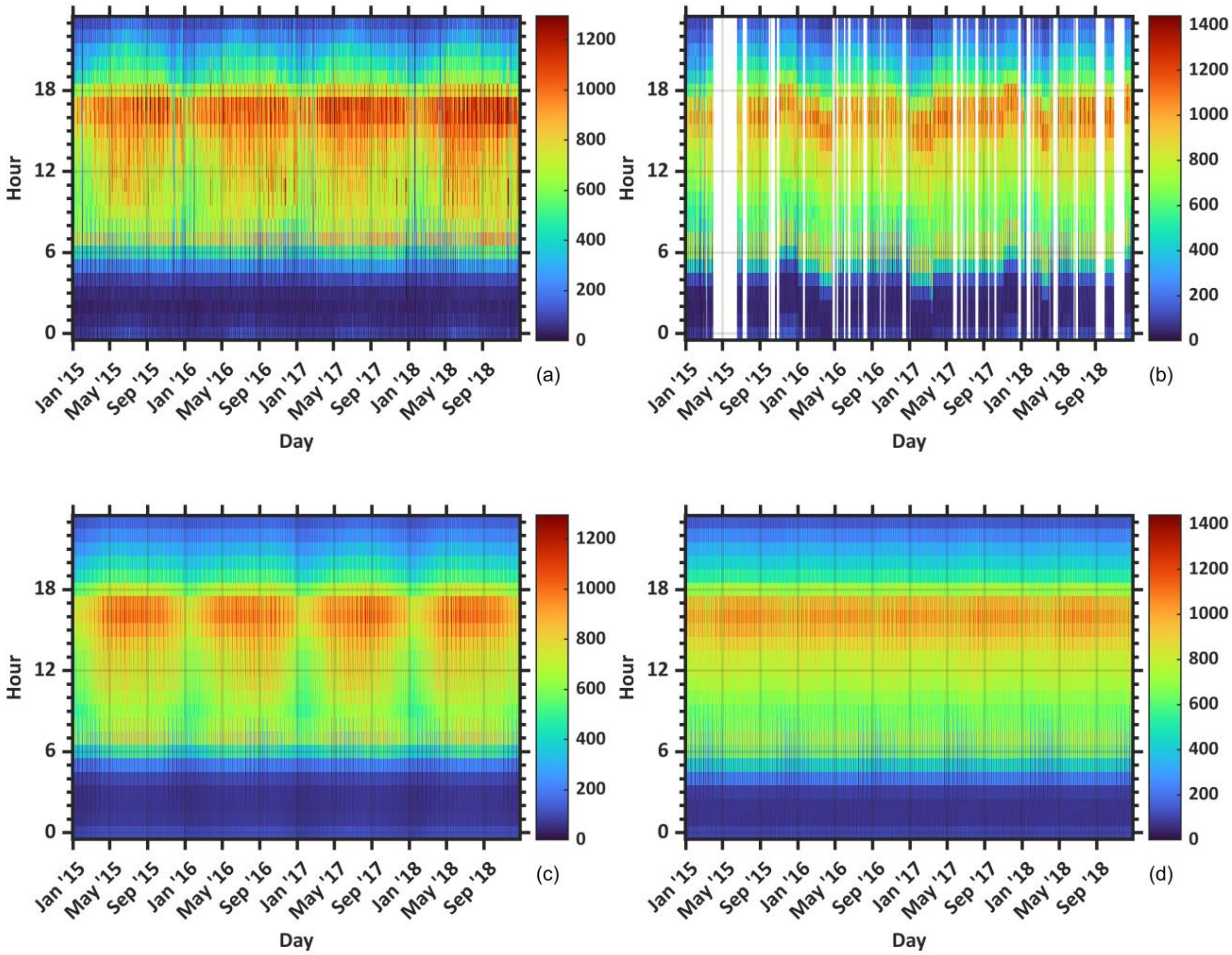
The first weekly and yearly principal components, as mentioned before, give the weighted average traffic pattern seen across sites, and are shown in Figure 3. The coefficients for the first principal component, both weekly and yearly, show high linear correlation with the AADT. This leads to a normalization approach based on the first PCC. What this means is that rather than needing 10 PCCs to represent traffic volume at a site, five normalized weekly PCCs and three normalized yearly PCCs, along with the AADT, can represent the total vehicle counts for a site. This normalized approach is more useful, as representing traffic volume for any hour of the week at a site, relative to its normal average number of vehicles, requires only five values, and only three for any day of a year, and predictions can be made that match the overall AADT of a site when it is known. The final reduced-order model therefore requires just nine total coefficients to represent hourly traffic counts for any site, 0.1% as many variables as the total number of hourly counts for a single year.

Figure 3. Normalized principal components that represent the most common normalized traffic patterns found across CONUS.



By using just these nine coefficients for a site, modeled traffic volumes are obtained, examples of which are shown in Figure 4(c-d) alongside the reported traffic counts in Figure 4(a-b). A comparison between the reported and modeled traffic counts shows that the PCA representation effectively smooths out inconsistencies in the raw data, while maintaining overall temporal trends. Additionally, because of the data weighting, missing data do not adversely affect the modeled traffic counts. By requiring only nine coefficients, the PCA representation allows for a simple way to both represent traffic volume data, which can not only be used to estimate traffic volume at sites when numbers are not reported, but also—by way of predicting PCCs—to predict traffic volume at other locations.

Figure 4. A comparison of the reported traffic volume for two sites (top) with the traffic volume calculated using the PCA representation (bottom).



5. CONCLUSION

The number of vehicles on a road can change drastically from one time period to another. While knowing the average number of vehicles is necessary to predict average noise levels caused by road traffic, a temporally varying model of road traffic is more beneficial as it can be used to predict not just a general average number of vehicles, but average numbers of vehicles for any time period of interest, whether that be an average Tuesday evening in springtime or the noise level for a particular hour, day of the week, and date.

By using traffic counts reported at thousands of locations across CONUS, a simple model is created to represent and predict traffic volume. This uses Fourier analysis to find temporal patterns in traffic counts at each site individually, and the principal component analysis to find the most common combinations of temporal patterns across sites. The model requires only nine coefficients to represent the hourly-dynamic nature of traffic volume for most locations.

This simplified model not only creates a concise way to represent traffic volume patterns, but also enables simple prediction of traffic volume when counts are unknown. Because only nine coefficients are needed, further methods can be created to predict these coefficients for locations where traffic counts are unknown. This remains a topic of interest and will be explored further.

By better representing and predicting traffic volume, further improvements in the prediction of road traffic sound levels can be made. While annual average expected traffic noise levels are important, increasing the temporal variability to enable hourly-expected noise levels, without drastically increasing complexity, can greatly increase accuracy and reliability in predicting sound levels caused by road traffic.

Future research will include not just temporal variability of traffic volume as a whole, but also the temporal variability of different traffic class types, such as large trucks, which have different expected spectral characteristics and sound levels than do smaller vehicles. This increased traffic variability model can then be used to predict sound levels and spectral characteristics of traffic noise across the continent on hourly, instead of average, time scales.

ACKNOWLEDGMENTS

This work was supported by a U. S. Army Small Business Innovation Research contract to Blue Ridge Research and Consulting, LLC, with Dr. James Stephenson as the technical monitor.

REFERENCES

[1] E. Öhrström, A. Skånberg, H. Svensson, A. Gidlöf-Gunnarsson, “Effects of road traffic noise and the benefit of access to quietness,” Journal of Sound and Vibration, Volume 295, Issues 1-2, August 2006, pp. 40-59.

[2] G. Shannon, L. M. Angeloni, G. Wittemyera, K. M. Fristrup, K. R. Crooks, “Road traffic noise modifies behaviour of a keystone species,” Animal Behaviour, Volume 94, August 2014, pp. 135-141.

[3] K. M. Parris and A. Schneider, “Impacts of Traffic Noise and Traffic Volume on Birds of Roadside Habitats, Ecology and Society, Vol. 14, No. 1, June 2009.

[4] D. R. Johnson and E. G. Saunders, “The evaluation of noise from freely flowing road traffic,” Journal of Sound and Vibration, Volume 7, Issue 2, March 1968, pp. 287-288, IN1, 289-309.

[5] S. K. Tang and K. K. Tong, “Estimating traffic noise for inclined roads with freely flowing traffic,” Applied acoustics, Volume 65, Issue 2, February 2004, pp. 171-181.

[6] National Transportation Noise Map by the Bureau of Transportation Statistics, https://www.bts.gov/geospatial/national-transportation-noise-map, accessed January 2021.

[7] A. D. May, “Traffic Flow Fundamentals,” Prentice-Hall Incorporated, 1990.

[8] F. Kessels, “Introduction to Traffic Flow Modelling,” Springer Link, 2018.

[9] H. Xiao, H. Sun, B. Ran, and Y. Oh, “Fuzzy-Neural Network Traffic Prediction Framework with Wavelet Decomposition,” Transportation Research Record 1836, no.1 (2003), pp. 16-20.

[10] D. S. Dendrinos, “Urban Traffic Flows and Fourier Transforms,” Geographical analysis, Volume 26, Issue 3, July 1994, pp. 261-281.

[11] Y. Xie, Y. Zhang, and Z. Ye, “Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition,” Computer-aided Civil and Infrastructure Engineering, Volume 22, Issue 5, July 2007, pp. 326-334.

[12] W. Min and L. Wynter, “Real-time road traffic prediction with spatio-temporal correlations,” Transportation Research Part C: Emerging Technologies, Volume 19, Issue 4, August 2011, pp. 606-616.

[13] L. Li, S. He, J. Zhang, and B. Ran, “Short-term highway traffic flow prediction based on a hybrid strategy considering temporal-spatial information,” Journal of Advanced Transportation, Volume 50, Issue 8, December 2016, pp. 2029-2040.

[14] K. Pedersen, M. K. Transtrum, K. L. Gee, S. V. Lympany, M. M. James, and A. R. Salton, “Validating two geospatial models of continental-scale environmental sound levels,” JASA Express Letters 1, 122401 (2021).

[15] P. Sun, N. AlJeri and A. Boukerche, "A Fast Vehicular Traffic Flow Prediction Scheme Based on Fourier and Wavelet Analysis," 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1-6.

[16] MATLAB function “nufft”, Natick, Massachusetts, The MathWorks Inc., copyright 2019.

[17] M. Fu, J. A. Kelly, and J. P. Clinch, “Estimating annual average daily traffic and transport emissions for a national road network: A bottom-up methodology for both nationally-aggregated and spatially-disaggregated results,” Journal of Transport Geography, Vol 58, 2017, pp. 186-195.