Geostatistical Modeling of Primary and Secondary Automobile Traffic Volume as Ecological Disturbance Proxy Across the Contiguous United States

Sean L. McFall[1], Matthias Leu[2], Benoit Parmentier[3], Marco Millones[4]
Stuart E. Hamilton[5]

[1]The College of William & Mary
Center for Geospatial Analysis P.O. Box 8795
Telephone: (703) 969-3916
Email: sean.l.mcfall@gmail.com, mmmillones@wm.edu

[2]The College of William & Mary
Applied Conservation Ecology Lab P.O. Box 8795
Telephone: (757) 221-7497
Email: mleu@wm.edu

[3] The University of Maine
Sustainability Solutions Initiative
Telephone: (508) 221-7497
Email: benoit.parmentier@maine.edu

[4] Salisbury University
Department of Geography and Geosciences
Email: sehamilton@salisbury.edu

1. Introduction

Roads have a deleterious effect on terrestrial biological integrity (Trombulak 2000). As vectors, they often direct negative anthropogenic activities on natural landscapes. For example, highways and dirt roads alike provide corridors for seed banks and invasive plant seed dispersal (Parendes 2000). In addition, depending on traffic volume, roads act as barriers to dispersal and as death traps for animals (Alverson 1988).

In this study, we created a nation-wide traffic disturbance layer derived from AADT (Annual Average Daily Traffic) data collected by the Departments of Transportation in the 48 contiguous states (CONUS). Our product predicts disturbance from primary roads, such as freeways or interstates, and from less-traveled secondary roads, like regional highways. The traffic-based disturbance layer provides a spatial variable as input for landscape analyses and models concerned with habitat degradation, connectivity of ecological systems and animal populations.

The creation of this layer provides a unique geocomputational problem related to data preparation and automated processing of workflow. In contrast to datasets that contain evenly distributed and/or concentrated data points across a spatial extent, the AADT data often does not follow this pattern. For instance, interstates cover large areas and as a result receive fewer data collection points. When kriged, these datasets tend to have large areas of one interpolated value between data collection points. Accounting for this problem is necessary for a coherent dataset that can be readily used by researchers, agencies and policy makers.

This dataset not only covers the CONUS region, but we believe improves upon previous proxy for road-borne environmental disturbance. Through extensive data processing and automation we have created a traffic disturbance index that is novel in scale and potential use.

2. Data Collection and Processing

This study assembled datasets for the 48 CONUS states. With the exception of Arizona and Texas, all data were collected electronically through states respective Departments of Transportation (DOT). The temporal extent varied state by state and to normalize this difference in data we initially selected the 2005-2009 time period, but significant loss of data for select states required us to expand to the 2002-2010 to maximize temporal coverage (Table 1).

Category	**Description**
Temporal Scale	2002-2011; varies by state
Spatial Scale	Contiguous United States; HI and AK not included
Source	Respective State's DOT
Units	Annual average daily traffic (AADT); vehicles
Format	Point

Table 1. Summary of Data Properties

Another idiosyncrasy with the data set was identified early on: traffic volume transitions between states were often abrupt and discontinuous, thus unrealistic. Traffic volume estimations doubled or tripled on the same road when moving from one state into another. Different methods of DOT traffic volume data collection and estimation are most likely the reason for this abrupt shift. To account for this, we buffered each state and appended any points within that out-of-state extent to the in-state data set (Figure 2). The length of this buffer depends on the neighboring states range, which is generated from the neighboring states variogram. For instance, a state like Washington would have two distinct buffers, one for each respective bordering state (OR, ID). All buffer creation was completed before any interpolation of the point data was carried out.

3. Methods

3.1 Method choice

We chose to model traffic disturbance using ordinary kriging models (Isaaks 1989). Traffic modeling estimations commonly utilize a linear networking approach to estimating traffic counts by assuming that traffic entered or exited the network between AADT counts. For example, if the average daily count is 10,000 vehicles on primary road X at point Y and the next count on primary road X at point Z is 8,000 vehicles, 2,000 vehicles are assumed to have left the traffic on secondary roads between point Y and Z on the primary road. In our case, however, while our main data source is traffic data, our goal is the creation of a proxy “disturbance” variable that can be generalized in all directions across many state boundaries. The disturbance index covers the whole study area and describes the negative effects of traffic volume (death traps for fauna, vectors for invasive flora, habitat destruction etc.). Presence and distance to roads are often used as an indicator of these negative effects, but without taking into account traffic amounts. We posit that this traffic disturbance proxy will be a stronger indicator of the negative effects of roads on ecosystems. This is not an effort to solely interpolate traffic. We are using traffic volume as a proxy variable for environmental disturbance in lieu of more intangible and difficult to measure variables. Additionally, a geostatistical approach helps overcome the issue of abrupt changes in traffic estimation at state boundaries.

3.2 Method Implementation

To develop a fully automated workflow for the CONUS region, we used an automated fit and selection of variogram models and its parameters. We used the autoFitVariogram and autoKrige functions implemented in the automap and gstat libraries in R (Hiemstra 2008, Pebesma 2004). However, the autoMap package has difficulties dealing with large areas when creating variograms. To address this shortcoming, we made some modifications to the variogram fitting procedures. Using a modified autoFitVariogram function (Koohafkan 2012), we were able to create variogram models that fit the data more reasonably. Figure 1 shows the sequence of tasks performed. The R code can be found online in a code repository (McFall 2015).

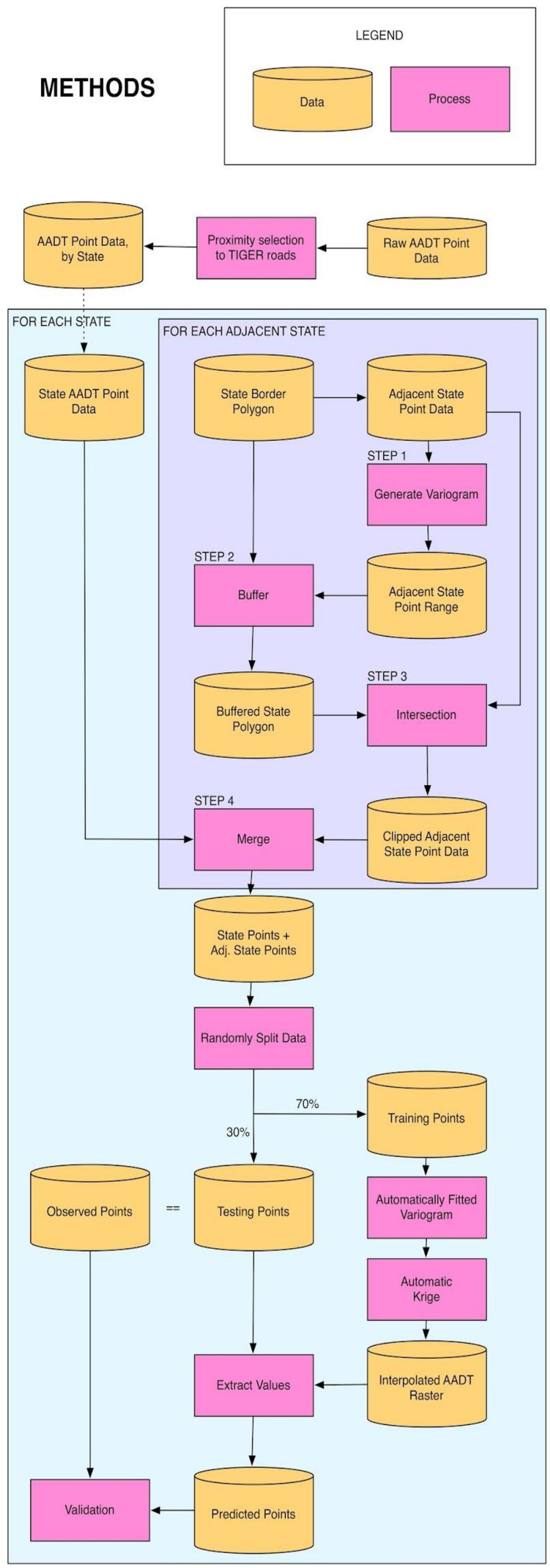

Figure 1. Methods for Traffic Disturbance Model

3.3 Validation

The validation of the kriged surface layer involves a split between training and testing samples that encompass the entire input AADT dataset. Only training values are incorporated in variogram creation and kriging for the interpolated AADT output raster layer. The observed values of the training samples are then compared to the predicted values of the raster layer, which are extracted from the same location as the training sample points. Accuracy metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are then incorporated to evaluate the accuracy of predicted values. Any unsatisfactory results will lead to adjustments in the variogram and kriging parameters. Such adjustments will lead to more satisfactory validation results.

4. Results

Here we provide the results for parts of the West Coast of the United States. The selected fitting model for each is the Stein variogram (Figure 2, Figure 3, Figure 4). Spatial patterns of the kriged surface show that disturbance is higher around city centers. Transitions between states are smooth for CA and AZ, though this is not the case with OR to WA. Oregon and Washington both contain large areas with similar values. Kriged values have been rescaled to 0-10 to avoid comparison to traffic values. (Figure 5)

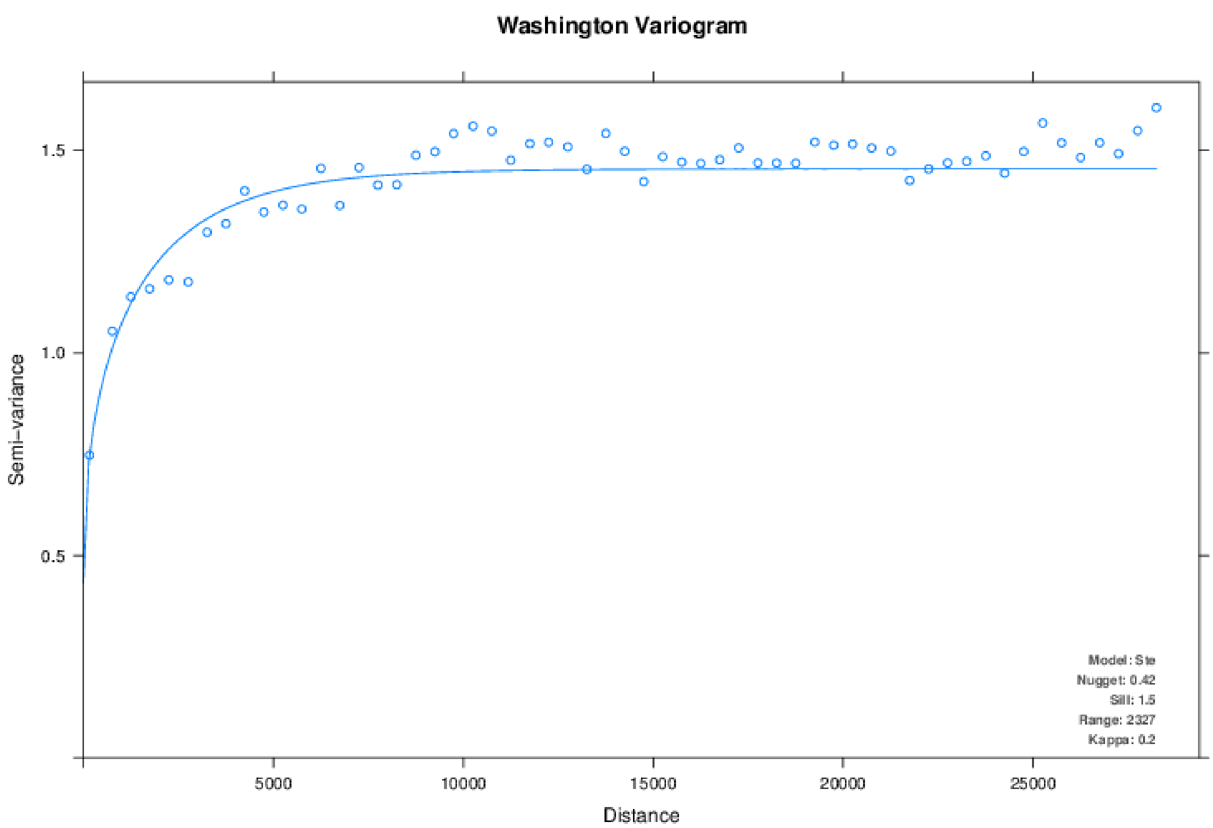

Figure 2. Variogram for log-transformed Washington State AADT data

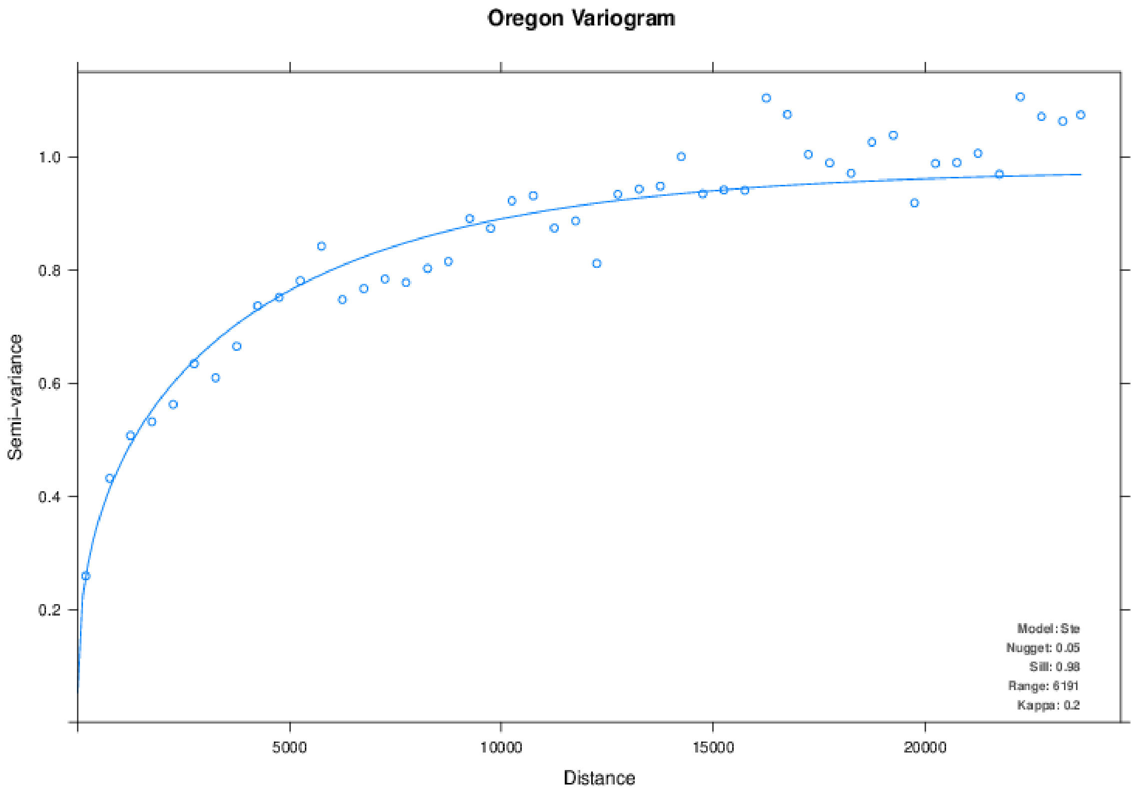

Figure 3. Variogram for log-transformed Oregon State AADT data

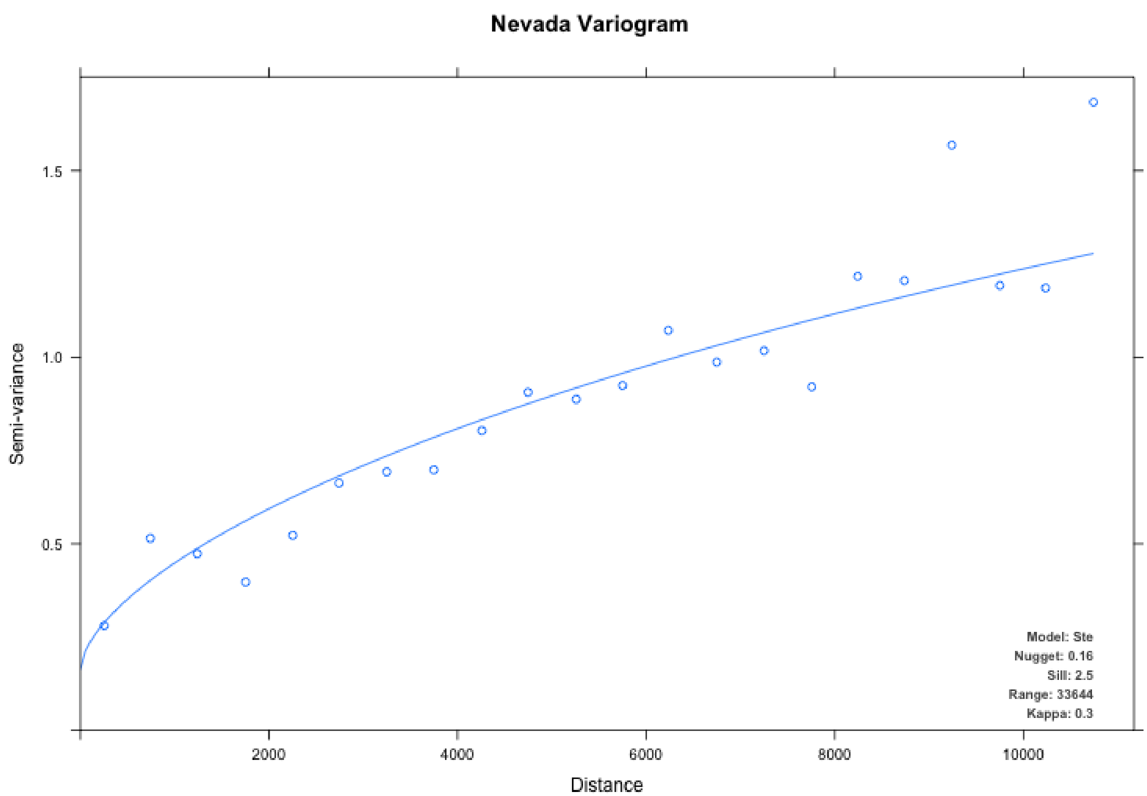

Figure 4. Variogram for log-transformed Nevada State AADT data

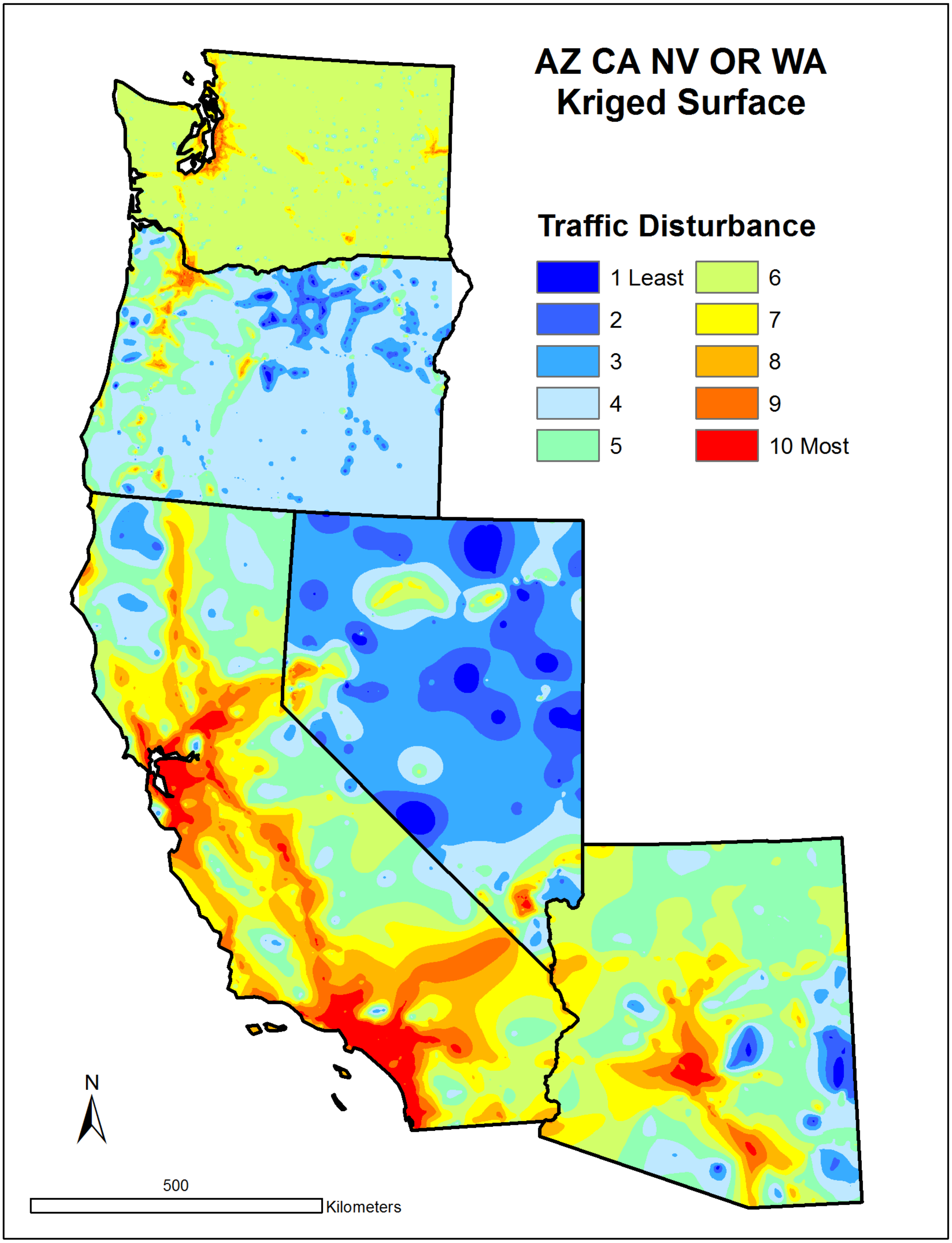

Figure 5. Kriged Predicted Surface for AZ, CA, NV, OR and WA

State	Raw AADT Data				TRAINING				TESTING			
					Raw		%		Raw		%	
	# years	# of points	Min Value	Max Value	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
AZ	5	11857	4.03	12.59	0.386	0.543	3.1	4.3	0.423	0.59	3.4	4.7
CA	3	13347	2.99	12.81	0.396	0.545	3.1	4.3	0.448	0.623	3.5	4.9
NV	2	1120	2.3	12.36	0.375	0.604	3.0	4.9	0.839	1.205	6.8	9.7
OR	5	10842	2.3	12.03	0.329	0.534	2.7	4.4	0.407	0.69	3.4	5.7
WA	5	15551	3.4	12.39	0.597	0.84	4.8	6.8	0.649	0.938	5.2	7.6

Table 2. Training and testing analysis for five Western States

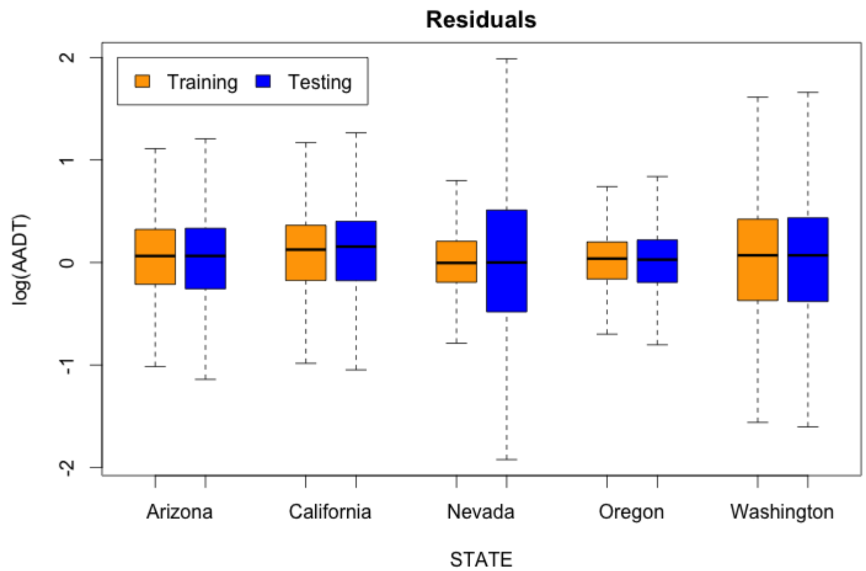

Figure 6. Validation residuals for Training and Testing data for five states

The circular and patterned nature of the kriged surface is expected. For kriging, uniformly distributed data collection points is the norm. Given a large distance between points, variation of the kriged surface will stagnate around the mean of the entire data set. Spatial patterns are highly affected by the distribution of the sample locations along the road network.

Given the inconsistent data collection and spatial distribution of points between states (e.g. North Carolina has a factor more points than many other states of equal area) we expect a certain level of individual adjustment for each states variogram and raw data.

We report in Table 2 the accuracy and error for training and testing using a 70 to 30 percent ratio. Raw error metrics and the proportion of error to the datasets max value are included. The wide range of residual values for Nevada in Figure 6 is presumably due to the low number of points within the state, at least when compared to the four other states.

Steps for improvement include the spatial selection of points using more up to date TIGER roads, (specifically the 2014 dataset), expansion of the kriging areas outside of state political boundaries (fauna has no knowledge of the state boundaries artifacts such as the Four Corners), and comparison of our traffic disturbance index to other commonly used proxies for ecological disturbance such as distance to roads.

The final poster will include more figures and tables. These will include all analysis results for all 48 states.

6. Acknowledgements

Marie Pitts, Corbett Wicks

7. References

Alverson, William S 1988, ‘Forests Too Deer: Edge Effects in Northern Wisconsin’, *Conservation Biology,* vol. 2, no. 4, pp. 348-358.

Hiemstra, P.H. 2008, ‘Real-time automatic interpolation of ambient gamma dose rates from the Dutch Radioactivity Monitoring Network’, *Computers & Geosciences*

Isaaks Edward H. and Srivastava R. Mohan, 1989, ‘An Introduction to Applied Geostatistics’, Oxford University Press, New York

Koohafkan, Michael 2012 ‘Know your variograms: working with the automap package for R’, *https://github.com/mkoohafkan/UCBcode-R/blob/master/Borr-kriging/borr-autofitmod.R*

McFall, Sean 2015 ‘Traffic-Volume-Interpolation: large scale traffic volume research project’, *https://github.com/slmcfall/traffic-volume-interpolation*

Parendes, Laurie A 2000 ‘Role of Light Availability and Dispersal in Exotic Plant Invasion along Roads and Streams in the H. J. Andrews Experimental Forest, Oregon.’ *Conservation Biology* vol. 14, no. 1, pp. 64-75.

Pebesma, E.J. 2004 ‘Multivariable geostatistics in S: the gstat package’ *Computers & Geosciences*, vol. 30, pp. 683-691

Trombulak, S. C. 2000, ‘Review of Ecological Effects of Roads on Terrestrial and Aquatic Communities’ *Conservation Biology* vol. 14, no. 1, pp. 18-30.