Public geospatial datasets as an approach to maximizing efficiency in the collection of site covariates in wildlife–vehicle collision studies

JAMES A. VANCE, Department of Mathematics and Computer Science, The University of Virginia’s College at Wise, One College Avenue, Wise, VA 24293, USA jav6e@uvawise.edu
WALTER H. SMITH, Department of Natural Sciences, The University of Virginia’s College at Wise, One College Avenue, Wise, VA 24293, USA
GABRIELLE L. SMITH, The University of Virginia’s College at Wise, One College Avenue, Wise, VA 24293, USA

Abstract: Wildlife–vehicle collisions (WVCs) are a major research focus because of increasing human health and safety concerns and the potential for biological impacts on wildlife. A key component of both understanding the causes of WVCs and designing mitigation measures is the collection and analysis of environmental and roadway data at WVC sites. However, collecting these site data can be logistically challenging and potentially dangerous to researchers. We studied the feasibility and accuracy of using public geospatial datasets, particularly Google Earth and Street View, as an alternative approach to assessing WVC on-site covariates. We randomly selected 50 sites from a larger WVC study and measured the topography, habitat type, width of the road median, and presence of fencing at each site as representatives of typical WVC site covariates. We compared the measurements recorded in the field to estimates obtained from public geospatial datasets in the lab. We determined that median topography had the lowest overall accuracy (60%), followed by presence of fencing with accuracy at 75% of sites. By contrast, median habitat type was identified correctly in almost all comparisons (96% overall accuracy). The root mean squared error for median width was 1.15 m overall. Our results suggest that Google platforms may serve as viable alternatives to field data collection for site covariates related to coarse measures of habitat type and some characteristics of road topography, thus reducing time requirements and potential safety risks to researchers in the field. However, there are several crucial caveats to consider when using geospatial platforms, particularly as they relate to 3-dimensional depictions of roadway features. Thus, we urge caution when attempting to use digital platforms to collect data on these covariates.

Key words: data collection, Google Earth, safety, Virginia, wildlife–vehicle collision

Wildlife–vehicle collisions (WVCs) are a major area of research focus in both biodiversity conservation and public safety (Litvaitis and Tash 2008). These collisions may result in increased wildlife mortality in many areas and constitute a potential threat to public safety, especially when large mammals are involved (Sullivan and Messmer 2003). Wildlife mortality rates from WVCs, for example, can exceed 10 individuals/km per day in extreme cases (Aresco 2005). Wildlife mortality rates are often dependent on migratory behavior and other phenomena related to the life histories of individual species (Ashley and Robinson 1996).

Bissonette et al. (2008) reported that in addition to the increased risk of human injuries and fatalities, WVCs result in millions of dollars of property damage. Concomitantly, research to identify measures to mitigate WVCs has evaluated wildlife crossings and improvements to road signage to alert drivers to areas of high WVC frequency as well as provide safe passage for wildlife (Clevenger et al. 2001, Sullivan et al. 2004, Huijser et al. 2007, McCollister and Van Manen 2010).

Key to understanding the underlying causes of WVCs and designing effective mitigation measures is the collection of field data at WVCs sites. Both the location of WVCs and spatiotemporal data involving the context of the collision are essential in developing robust models seeking to explain variation in WVC frequency (Malo et al. 2004, Gunson et al. 2011). Abiotic factors including the location of the roadway in relation to various habitat features, the presence of structures such as...
culverts and fencing, and even the design of the roadway itself, have been correlated to high WVC frequencies (Seiler 2005, Ng et al. 2008, Danks and Porter 2010, Neumann et al. 2012). This information can both help to elucidate the proximate causes of spatial aggregations of WVCs and inform the design and placement of effective mitigation strategies.

However, collecting data on the environmental correlates of WVCs can be logistically challenging and may even place researchers at risk of harm in certain contexts. Because WVCs occur over broad areas, multiple seasons or weather conditions, and often in large numbers (Forman and Alexander 1998, Trombulak and Frissell 2000), the collection of WVC data is time- and labor-intensive, especially when multiple variables reflecting roadway design and the spatial contexts of collisions must be collected on-site (Olson et al. 2014). In some areas, features such as narrow or nonexistent road shoulders, high traffic volume, and steep slopes or cliffs may also make the collection of such data dangerous or impossible for researchers, making more efficient and safe methods for the collection of these data essential to performing WVC research.

The advent of public geospatial platforms such as Google Earth™ (Google Inc., Mountain View, CA, USA) have created a potential avenue for increasing the safety and efficiency of data collection for many WVC-related variables. These platforms provide high-resolution ortho-imagery and even ground-level imagery, such as Google Street View™, (Google Inc., Mountain View, CA, USA), that allows for features such as road design (e.g., slope, surfacing, etc.), habitat features, and the presence of structures and fencing to be determined ex situ rather than by researchers in the field (Anguelov et al. 2010, Yu and Gong 2012), provided that geospatial coordinates for the WVC site are available. These datasets remove researchers from potentially dangerous field conditions and allow for data from multiple, geographically-dispersed points to be collected simultaneously. The use of these platforms has previously been shown to serve as a viable alternative to field-based data collection for some site covariates outside of a WVC context in both civil engineering (Yan et al. 2013) and public health studies (Rundle et al. 2011, Odgers et al. 2012).

However, a potential tradeoff of this approach exists in the possibility that data from these digital sources may not accurately reflect actual field conditions found at WVC sites. Few studies have addressed this accuracy to date. To address this need, we compared environmental variables collected from the field and from public geospatial datasets at the same sites within an area with high WVC frequencies. We subsequently provide insights regarding how these public datasets may benefit and advance WVC research.

**Methods**

**Study area**

Beginning January 1, 2015, we sampled WVCs on an approximately 100-km road corridor from Richlands, Virginia (37.06696, -81.81630) to Wise, Virginia, USA (36.96314, -82.54373). This corridor includes 2 divided, 4-lane U.S. Highways (US-23 and US-58 Alternate; 92 km total) and 2 stretches of state-maintained secondary roads on its western (VA-706) and eastern (VA-609) termini. The study area occurs within the central Appalachian region of far southwest Virginia, an area with a high amount of biodiversity for many vertebrate taxa (Stein et al. 2000) and that traverses the northern end of the Upper Tennessee River watershed. Habitat types within the study corridor are diverse, ranging from mixed hardwoods and open agricultural land typical of the Valley and Ridge Physiographic Province on the corridor’s eastern end to primarily dense, mixed mesophytic hardwood forests characteristic of the Appalachian Plateau Province on its western end. Terrain across the study area is characterized by rolling hills and exceedingly steep plateau escarpments, with large tracts of publicly-owned conservation lands (Jefferson National Forest and Clinch Mountain Wildlife Management Area) adjoining much of the road corridor.

**Data collection**

This study was part of the larger WVC study described above and was conducted in 2015 by the authors. By the end of 2015, we recorded 1,837 WVCs at 1,804 sites. We conducted the present
analysis in October 2015, at which time we had 1,400 WVC sites. To conduct our analyzes, we randomly selected 50 of the 1,400 sites along our survey route to compare our field-based and public datasets. We chose this approach to: 1) avoid biasing our selection of points toward those easily accessible in the field, 2) randomize abiotic WVC covariates as much as possible within the scope of our dataset, and 3) eliminate any potential for bias in terms of selecting points that were easily visible in the Google Earth and Street View platforms. We chose a small number of randomly selected points relative to our overall WVC total to minimize the selection of sites with duplicate habitat and road covariates, as the WVC sites in our overall dataset were often spatially clustered at a high density.

Typical covariates measured at WVC field sites include road topography (Clevenger et al. 2003, Gomes et al. 2008), road width (Barrientos and Bolonio 2009, Gunson et al. 2011), local habitat (Jancke and Giere 2011, Barthelmess 2014), distance to cover (Finder et al. 1999, Clevenger et al. 2003), presence of ecotone (Farrell and Tappe 2007, Ng et al. 2008), presence of guardrail (Malo et al. 2004, Barthelmess 2014), and presence of fencing (Bashore et al. 1985, Seiler 2005). We recorded 4 covariates at each site: median topography (ditched, raised, level, slope up, slope down), median width (in meters), median habitat type (forest, shrub, herbaceous, hard surface), and presence of a fence (yes, no), which represent typical covariates in WVC studies. We measured widths using an electronic distance measuring tool (Johnson Level & Tool Manufacturing Company Inc., Mequon, WI, USA) with reported accuracy of 1.59 mm out to 50 m.

A small number of sites (n = 3) contained roadways with no median; we therefore recorded only the presence of a fence and the width of the road (instead of the width of the median) at these sites. At 5 additional sites, we recorded the width of the road instead of the width of the median due to obstructions in the median that prevented accurate measurement. These points were subsequently excluded from analyses involving comparisons of median characteristics. We were able to measure all covariates at all remaining sites.

We recorded these same measurements in the lab using Google Earth and Google Street View. We used the same categories for each habitat and median type mentioned above for field comparisons in our digital observations. We measured median width in meters using the measure distance tool in Google Earth. We observed median habitat type in Google Earth and then verified observations using Google Street View imagery. We recorded the presence of a fence and median topography by using Google Street View imagery to view the site from its nearest tile and all adjacent tiles, rotating through all possible angles within each tile.

Data analysis
We compared the presence of a fence, median habitat type, and median topography using the percent accuracy of lab estimates compared to field observations for each class (e.g., type of median, median topography, etc.) within a covariate and overall for each covariate. We compared the accuracy of the median widths by computing the root mean squared error between field and lab measurements.

Results
The collection of both field data and corresponding data in Google Earth were possible at all of our randomly selected WVC sites, excluding those special cases previously mentioned. Median topography had the lowest overall accuracy when comparing measurements from the field and public datasets. Estimates of median topography from Google Earth and Street View were accurate 60% of the time. Estimates for sloped and ditched medians were the least accurate across all possible median topographies, with 14% and 52% accuracy, respectively. The presence of fencing was our next most accurate habitat characteristic, with the presence or lack of fencing correctly identified in Google Earth and Street View at 74% of sites. Errors were more common for sites with fencing (59% accuracy) versus those without (86% accuracy).

By contrast, median habitat type was identified correctly in almost all comparisons (96% overall accuracy). Habitat was misclassified at 2 locations: 1 site containing a hard, paved surface identified as having herbaceous habitat from Google Earth and Street View and another site with shrub habitat classified as herbaceous in public platforms.
The root mean squared error for median width was 1.15 m with a 4.48-m maximum and 0.02-m minimum absolute difference between field and Google Earth measurements. Just over half \((n = 26)\) of the width measurements from the field were smaller than their corresponding lab measurements.

**Discussion**

Our results suggest that Google Earth and Street View platforms can serve as viable alternatives to WVCs site data collection, with some important caveats. The characterization of median habitat was the most accurate measurement in our study, with Google Earth and Street View estimates proving accurate >95% of the time. One benefit of characterizing median habitat is that it can be reliably assessed from both aerial ortho-imagery available in Google Earth and street-level imagery provided in programs such as Google Street View. This feature, coupled with the relative ease of visually comparing forested versus shrub or herbaceous habitat types, likely explains the high accuracy of this measurement.

However, it is important to note that the regional context of our study site may help to enhance the ease of our visual habitat comparisons. Forest habitats within the upper Tennessee Valley of southwest Virginia are typically composed of mixed hardwood forests interspersed with dense aggregations of eastern redcedar \((Juniperus virginiana)\) on associated glades or barrens that form a stark visual contrast to shrub or herbaceous habitats (Smalley 1984, Ludwig 1999). Further comparisons of field-based and public datasets, similar to our study, performed in regions with different (and possibly less visually apparent) habitat characteristics may help to further provide support for the measurement of this variable using publicly accessible platforms.

By contrast, median topography and the presence or absence of fencing were less accurate variables in our comparisons. Both of these covariates are reliant on accurate 3-dimensional imagery of WVC sites, because both refer to the presence of raised (or potentially raised) features within the larger landscape. While some fences or guardrails may be visible in high-resolution aerial ortho-imagery, these features may be obscured when located beneath the forest canopy or within thick, unmaintained herbaceous cover, leaving street-level imagery such as Google Street View the sole data source for these estimates in most cases.

When superimposed on the surrounding landscape, street-level imagery is rendered into a 3-dimensional “bubble” by stitching multiple georeferenced images into a seamless, 360-degree panorama using an image processing algorithm to smooth boundaries and transitions between images (Vincent 2007). This process may occasionally lead to errors in geolocation or, for accurately geolocated sites, phenomena where raised features and edges become slightly distorted or foreshortened (Kopf et al. 2010, Tsai and Chang 2012). These features may lend street-level imagery to being less accurate in these cases, a factor illustrated in our median topography dataset.

Our 2 least accurate median topographies (ditched and sloped), for example, involve topographic characteristics of medians that are reliant on an accurate 3-dimensional rendering to classify correctly. Slightly sloped medians may lose relief when imagery is rendered, as can medians whose centers are slightly ditched. Most misidentified median topographies in our dataset, in fact, were those that were actually ditched or raised yet appeared level to the observer when viewed using street-level imagery. Accuracy was much higher for other median topographies that did not rely on this type of 3-dimensional accuracy for correct interpretation. These results suggest caution if applying public datasets to the assessment of median topography in WVC studies.

The presence or absence of fencing is subject to similar concerns when viewed using public geospatial datasets. If fences are not rendered in street-level imagery in a way that allows for them to be distinguished from surrounding vegetation and terrain, it is possible that a bias may be introduced in terms of a high rate of false negatives (fences being estimated as absent when they are actually present). Our data supported this prediction, as inaccuracy was much higher when fences were present in the field as compared to when fences were absent (41% to 14% inaccuracy, respectively). As with median topography, these results
indicate that public datasets should be viewed cautiously in terms of indicating the presence of fencing at WVC sites. More broadly, it is likely that these same issues apply to most covariates that may be reliant on accurate 3-dimensional rendering in street-level imagery for detection, and we urge caution if attempting to use digital platforms to remotely collect data on such 3-dimensional covariates in WVC studies.

The error in median width was relatively small, with no pattern of either larger or smaller measurements from the field compared to geospatial platforms. Accuracy using the measurement tool in Google Earth was complicated by shadows from overhanging tree limbs in aerial ortho-imagery and by the lack of an ability to consistently distinguish the boundary between the edge of the road and start of the median. This was not a major impediment to data collection in our study, although these complications may become more pronounced when applying our approach to areas with different road and/or habitat characteristics.

However, remote data collection using geospatial platforms may be the only approach available to researchers in some contexts. For example, it may be impossible to obtain measurements of median width in the field using electronic devices at sites with raised or forested medians. Also, field measurements using traditional devices like tape measures may be impossible due to steep drop-offs or sheer rock escarpments in the median. Other complications to width measurements in the field include the presence of water, amount of road traffic, and personnel requirements. The use of public geospatial platforms to measure median width, then, may be a viable alternative to field measurements, especially when the aforementioned limitations are not at play and/or actual field measurements are difficult to impossible to perform.

It is important to note that, while our results do support the use of public geospatial datasets in characterizing WVC covariates, there are several key caveats to keep in mind with respect to the applicability of these datasets. One potential risk relates to observer bias, because many of the features represented by aerial and street-level imagery are subject to more interpretation than when these same features are viewed in person. The shape of a median or the type of habitat present at a WVC site, for example, may be substantially more ambiguous when viewed using street-level imagery than when viewed in the field, potentially leading different observers to categorize these types of structures inconsistently. This type of bias, in fact, has long been an issue in studies that involve the characterization of habitat variables in field comparisons (Gotfryd and Hansell 1985, Block et al. 1987) and likely extends to the collection of these same variables when viewing habitat remotely.

While our study did not explicitly examine any potential impacts from observer bias in the use of public datasets (e.g., a single observer characterized covariates in the lab), this is a key consideration that should be investigated in future work if multiple researchers will be classifying covariates from public data sources. In addition, the geographic coverage of public geospatial datasets may limit the use of these data sources in WVC studies in some areas. Our study area was well-suited to our approach since the entirety of the region is covered by and visible when using both Google Earth and Street View, but this is not always the case. While technological advances in street-view imaging technology are steadily increasing the coverage of public datasets, many areas still lack street-level imagery or occur on smaller, county-maintained routes or roads maintained primarily for forest management that are not traveled by street-level imaging vehicles (Anguelov et al. 2010, Blitz 2012). In addition, some areas may lack both types of data when street-level imagery is not available and aerial ortho-imagery coverage is absent, sampled at an insufficient resolution, or obscured by vegetation or terrain. In these cases, field-based methods are likely the only source available to WVC researchers.

Management implications

Overall, however, our results demonstrate that public datasets may be a viable alternative to field-based data collection for at least some covariates in WVC studies, particularly covariates related to coarse measures of
habitat type and some characteristics of road topography. Although researchers will likely never be able to rely exclusively on ex-situ data sources due to the need to visit WVC sites to obtain data on the location of collisions and species of wildlife killed by vehicle collisions, this use of the ortho-imagery platforms we studied may help to streamline field data collection and reduce the safety risks to researchers. With new technological approaches to data collection for both the location of wildlife–vehicle collisions and WVC covariates continuing to advance, this applicability will likely only grow and enhance our ability to ascertain the causes, impacts, and prevention of wildlife–vehicle encounters.

Acknowledgments

Funding for this project was provided by the Department of Mathematics and Computer Science and the Department of Natural Resources at The University of Virginia’s College at Wise. The Virginia Department of Transportation issued a land use permit (#104-38295) for this project.

Literature cited


JAMES A. VANCE is an associate professor of mathematics and wildlife resources at the University of Virginia's College at Wise. He received his Ph.D. degree from Virginia Tech in 2006, where he investigated omnivory models. His recent work includes applications of mathematics and statistics to engineering, ecology, and wildlife management.

WALTER H. SMITH is an assistant professor of biology at the University of Virginia's College at Wise. He received his Ph.D. degree from the University of Alabama in 2011, where he investigated the impacts of prescribed fire on herpetofauna. His recent work has centered around the impacts of ongoing socioeconomic change in rural Appalachia on amphibian taxa.

GABRIELLE L. SMITH is a senior at the University of Virginia's College at Wise. She is completing her B.S. degree in mathematics along with the Virginia Teacher Licensure. Her position as a research assistant in the Department of Mathematics and Computer Science and her senior research project investigated the spatial distribution of roadkill in Southwest Virginia. She will pursue a Ph.D. degree in bioinformatics starting in Fall 2017.