Object-oriented Classification of Impervious Surfaces in Sierra Vista Watershed

Prepared for:

Upper San Pedro Partnership

Prepared by:

Zachary Sugg

School of Geography and Development

University of Arizona

David Goodrich

USDA-ARS SWRC

Object-oriented Classification of Impervious Surfaces in Sierra Vista Watershed

Project Summary

The objective of the project described in this report was to estimate the area of impervious surfaces within a large section of the Sierra Vista subwatershed of the larger Upper San Pedro River basin. This was accomplished using high-resolution satellite imagery and GIS-based object-oriented classification software. Object-oriented classification has been shown to effectively overcome some of the limitations of more traditional vegetation-based methods that are not well-suited to sparsely vegetated arid and semi-arid areas such as southeastern Arizona. This report summarizes the steps used to produce a map of impervious and pervious surfaces for the study area as well as an assessment of the accuracy of the classification. It describes the study area, data pre-processing steps, the generation of data layers used to enhance the classification, and the workflow of the classification process for the Feature Analyst software. The approach that was developed produced a good mean overall classification accuracy across the set of subdivided images (89.7% with a range of 82.9-96.0%) and high mean overall Kappa coefficient of agreement (0.84; Kappa = 1.0 for a perfect classification and 0 for no agreement).

The study area has experienced rapid urbanization and development in recent years, likely altering runoff and recharge patterns within the watershed. Improved estimates of the amount of impervious surface associated with this development are critical to more effectively modeling and predicting the changes in runoff due to urbanization, and by implication, better estimation of the larger impacts of development within this environmentally sensitive watershed. The results reported here represent a significant improvement over previous estimates of impervious surfaces in the watershed.

Contents

1. Project Overview

The objective of this project was to update previous estimates of the area of impervious surfaces in and around the Sierra Vista municipal area. Recent and continuing rapid urbanization in the Sierra Vista watershed has implications for movement of runoff, in particular the potential enhancement of recharge to groundwater (Goodrich et al, 2004, Coes and Pool, 2005; Kennedy et al., 2011). The substantial increase in runoff due to urbanization has been demonstrated by instrumental records in the La Terraza watershed and adjacent natural watershed on Ft. Huachuca. A more up-to-date estimate of the amount of impervious surface within the watershed can be used to determine the upper bound of urbanization for hydrologic modeling purposes to help understand the implications of urbanization for water cycling in the area. The general study area is the western portion of the Sierra Vista subwatershed of the Upper San Pedro river basin.

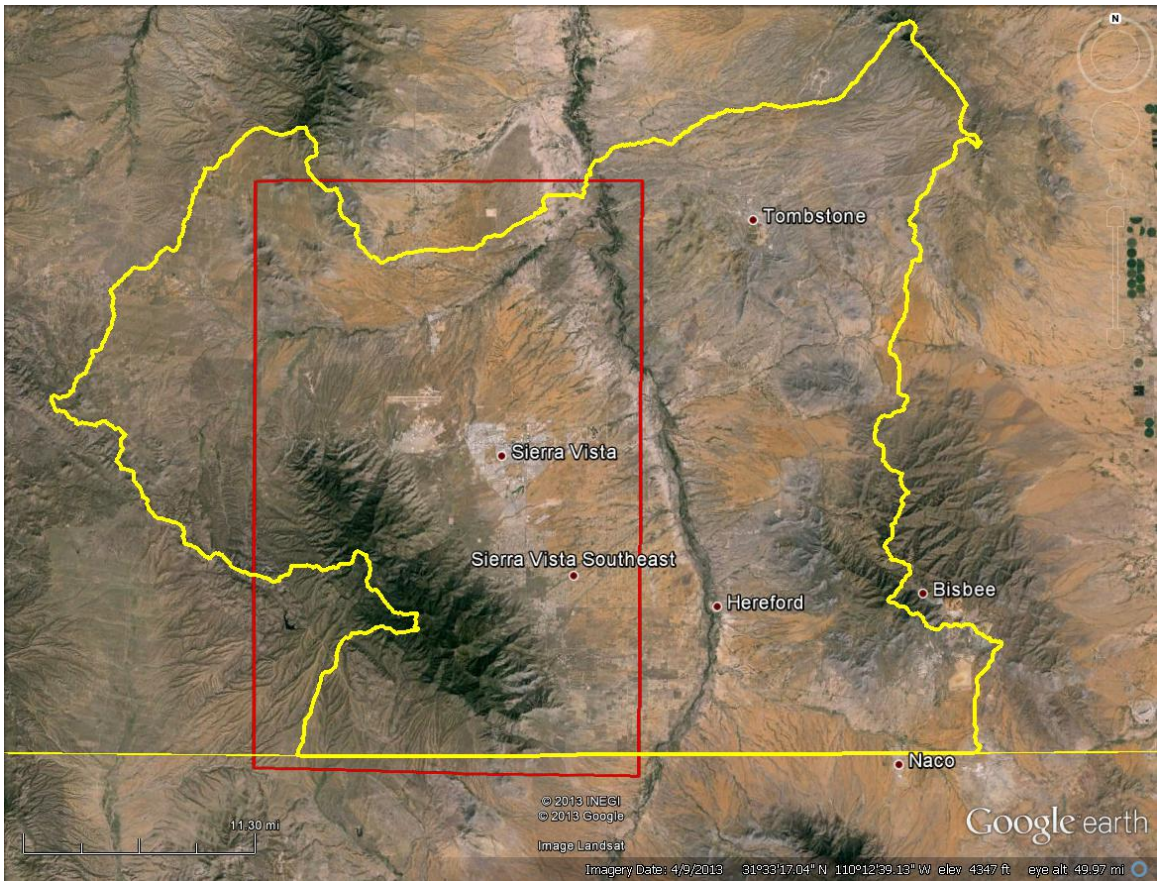
Traditional classification methods for discriminating between pervious and impervious surfaces have consisted generally of either manual delineation or automated classification based on spectral characteristics of imagery using such indices as Normalized Difference Vegetation Index (NDVI) (Finke et al., 2007). Both methods have important drawbacks; manual delineation is very labor intensive and impractical for larger areas, and spectrally-based methods are problematic in arid and semi-arid landscapes with a large amount of non-vegetated natural pervious surfaces. Recent work using object-oriented classification has shown to be effective at achieving desired accuracy without being prohibitively time-consuming for larger areas, even with higher-resolution imagery. Finke et al. (2007) developed a semi-automated method for detecting impervious surface for the relatively small (31.5 ac, 12.8 ha) La Terraza subdivision using the Feature Analyst object-oriented classifier with good results. That methodology was adapted to classify the present study area, which is many times larger

in spatial extent than the La Terraza subdivision. This report describes the process of adapting the methodology described by Finke et al. (2007) to achieve accurate classifications of pervious and impervious surfaces using QuickBird imagery of the Sierra Vista study area. A vector shapefile version of the impervious surface map layer generated by the study can be downloaded at http://www.tucson.ars.ag.gov/dap/.

2. Study Area and Data

The study area consisted of a large section of the Upper San Pedro Watershed containing the developed areas of Sierra Vista and Ft. Huachuca (Figure 1). The scenes were bounded to the south by the U.S.-Mexican border and to the west and north by portions of the Sierra Vista subwatershed boundary. Six overlapping QuickBird scenes with an acquisition date of December 26, 2009 were acquired from Digital Globe. Swath width of the QuickBird satellite is 18.0 km^2, thus the nominal total area encompassed by the six scenes was 1,944 km^2; the actual total area is smaller due to image overlap. For each of the six scenes there were two images: one multi-band with a spatial resolution of 2.4 m and a 0.6 m panchromatic band. Multi-band images contained four bands: blue, green, red, and near-infrared (NIR). Twelve images in total were acquired.

Figure 1. The study area with the Upper San Pedro Watershed boundary and international border in yellow and the QuickBird imagery coverage in red.



3. Pre-processing

Pre-processing steps were done to increase the spectral resolution and positional accuracy of the images. First, the multi-band images were pan-sharpened using the panchromatic images to increase the resolution from 2.4 to 0.6 m to match the panchromatic images. Second, because all of the images were roughly georeferenced prior to acquisition (roughly 10 m accuracy) but not orthorectified, orthorectification was performed to improve positional accuracy. A 10 m digital elevation model (DEM) from the National Elevation Dataset was acquired from the USGS Seamless Server online and given the same horizontal projection as the QuickBird imagery. All twelve QuickBird images were orthorectified using the QuickBird orthorectification model in ERDAS Imagine, which incorporated the 10 m DEM, sensor settings contained in the rational polynomial coefficient files accompanying each QuickBird scene, and ground control points acquired from 2007 National Agriculture Imagery Program 1 m orthophotographs. The root mean squared error obtained for all images was less than 1.5 pixels with a range from 1.02 to 1.40 using an average of 15 control points per image.

4. Generation of Input Data Layers

Finke et al. (2007) also used QuickBird data and tested three different combinations of derived data sets for classification to determine which produced the best result. These were called “Vis,” “Pan/Vis/NIR-Plus,” and “PCA-Plus.” The “Pan/Vis/NIR-Plus” combination, which consisted of the four QuickBird multispectral bands, Normalized Difference Vegetation Index (NDVI), edge-enhanced images, and the panchromatic band produced the most accurate classification. Based on those results, this study replicated that combination of input bands. In order to enhance the classification by detecting vegetation, the NDVI was calculated for each multi-band image using the formula

$$NDVI = (DN_{NIR} - DN_{RED}) / (DN_{NIR} + DN_{RED})$$

where NIR is band 4 and RED is band 3. The Sobel filter was used to create a non-directional edge enhancement of the panchromatic band to further delineate more defined linear features associated with impervious surfaces such as buildings, roads, and parking lots. The data sets used for inputs for training the classifier included seven total data layers: the panchromatic band, the red, green, blue, and NIR bands, and the two derived layers which were the NDVI and edge-enhanced images.

All images overlapping the Sierra Vista subwatershed boundary were clipped to that boundary. Initial testing of the Feature Analyst classifier with the seven data layers resulted in processing times that were so lengthy as to be unworkable. In order to make processing manageable, the images were “diced” into smaller tiles. A 5 m “collar” was used during the dicing process to ensure the images overlapped to prevent the loss of edge pixels. Dicing all the QuickBird scenes resulted in a total of 21 individual tiles, each one to be classified individually.

5. Classification Workflow

This section summarizes the procedures used to classify impervious surfaces. The steps involved in a Feature Analyst classification workflow are outlined in the software documentation as follows:

A. Defining feature target examples for the Feature Analyst learner tool in a training set.
B. Running an initial learning pass.
C. Revise the classification:

- Refining the training set, as necessary.
- Training the Learner to recognize the targets through a sequence of Hierarchical Learning passes with adjustments, as necessary, to the input representation.
- Masking out troublesome features, if necessary.
- Removing clutter (noise) and finding missed features by correcting the Learner through the selection of correct, incorrect, and missed examples.
- Repeating the hierarchical learning process until satisfied with results.

D. Applying post-processing techniques such as smoothing and converting polygons to lines or points.

E. Saving results for hand-editing, batch classifying, or for use with Feature Modeler.

These procedures were used on 18 of the 21 tiles. The remaining three were classified manually because there were so few impervious features present in them.

5.A. Defining a Training Data Set

The first step in a supervised classification is to identify impervious features from which to create a training polygon data set. In order to train the classifier to distinguish between pervious and impervious surfaces, training polygons needed to capture both the edges of shapes and their internal spectral characteristics. Following Finke et al. (2007), a simple classification scheme of three classes was used: pervious, impervious, and shadow. Two sets of training polygons were created for each image tile: impervious and shadow. Training data sets for pervious were not created for reasons explained in section B below.

The goal was to capture the spatial and spectral diversity of shapes and their distribution throughout the scene. Initial training polygons were later refined following a round of accuracy assessments of the initial classifications in order to produce a more accurate result. Shadow pixels were to be later reclassified as pervious or impervious to produce a two-class classification for the final output. Features considered impervious were all paved or sealed surfaces that could be discerned from inspection of the images. These included objects such as paved roads, parking lots, sidewalks, airport runways, buildings, residential structures, driveways, and recreational areas with sealed surfaces such as running tracks. For the purposes of this classification, we considered everything that was not paved or sealed to be pervious; these features consisted mostly of vegetated surfaces (forested and shrub-dominated), undeveloped bare areas, unpaved roads and trails, agricultural plots, individual trees, playing fields, washes, lawns, parks, and road medians.

Dark asphalt was an obvious impervious surface, as were buildings and homes with bright, uniform reflectance. However, in this rapidly developing area where neighborhoods and subdivisions are often only partially developed, there can be a variety of surfaces that may or may not appear to be impervious depending on a particular band combination, or even regardless of band combinations. Consequently, roads and other surfaces in various stages of improvement are often problematic for the classifier to classify accurately. However, a few basic types of roads could be distinguished by the observer using contextual clues and ancillary data such as Google StreetView street-level photographs:

1) Dark asphalt. The “most developed” road surface, usually easily distinguishable based on clearly painted lines (usually dotted), presence of a median between lanes, and a very low reflectance that is also very homogeneous/uniform across the surface. Often found in parking lots.

2) Non-asphalt paved roads. Distinguishable by lighter coloration as associated with concrete, painted lines and low reflectance. Less homogeneous than dark asphalt. Also often found in parking lots.

3) Graded but unpaved roads. These tend to appear bright in the imagery due to presence of gravel, rock, and compacted soil, but not homogeneous like buildings. Often discernible due to spatial context, e.g. in mining, industrial, undeveloped/forested areas, and connection to other non-paved areas.

4) Dirt roads. These surfaces are distinguishable by a tan or reddish coloration characteristic of nonvegetated natural surfaces in natural color imagery.

For this classification, we considered the first two surface types to be impervious, and the rest pervious, and the classifier was trained accordingly.

The training data for shadows were designed primarily to capture at least the shadows cast by buildings and structures and shadows associated with natural surfaces, e.g. those by tall trees in forested areas. In the portions of the study area with a relatively high amount of relief, i.e. the southwest area, large areas were cast in shadow by ridgelines. However, as these areas were clearly pervious, we generally limited the final training data sets to those shadows cast by individual features only, and not entire areas of pervious area. This is also because these areas would later be converted to pervious classification for the final output anyway.

5.B. Running a Learning Pass: Classifier Settings

Following the creation of the training polygons, the second major step is setting up the parameters of Feature Analyst. Feature analyst settings were kept the same among the tiles in order to be consistent. Extensive testing and examination of the results of different combinations of classifier settings and training data sets helped determine an effective classification model that could be applied across all tiles. This process retained most of the same settings used by Finke et al. (2007) but also deviated in some ways due to the vastly greater spatial extent and diversity of features that had to be accounted for in order to optimize the accuracy of results.

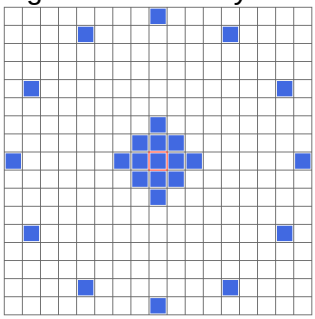
The initial round of classifications using the same settings as Finke et al. (2007) indicated that a key source of error resulted from the confusion of certain bare natural surfaces with impervious surfaces that had very similar shapes and reflectances. The common result was the erroneous detection or “false positives” of impervious shapes in natural areas with bare patches, which littered the undeveloped areas of the images. This effect could be minimized by altering the classification method such that rather than using three input training data classes (shadow, impervious, and pervious), only two were used (shadow and impervious), and then leaving the remainder of the image temporarily unclassified. Finke et al. (2007) used a “wall-to-wall” classification type where every pixel is classified into one of the three input classes. For this study, the wall-to-wall setting was turned off, permitting the classification of only the two input layers and leaving the rest as unclassified background area. Testing results showed that doing this minimized the confusion between pervious and impervious training polygons which seemed to be the primary source of this type of error. The reasoning behind the altered classification scheme was that everything that was not shadow or impervious (the “background” pixels) was pervious, therefore the two input training classes were used to

extract just those two classes initially, and the unclassified background areas (the pervious areas) could then be classified in a later step and renamed “pervious.”

Feature Analyst processing settings are in four groups as follows:

(1) The input representation settings determine the pattern used by the classifier to detect impervious features. Although certain preset patterns are recommended by Feature Analyst for certain types of features, e.g. buildings, trees, roads, etc., we needed to use just one pattern to detect all impervious shapes. We tested the custom pattern used by Finke et al. (2007) but found it tended to produce too many false positives in pervious areas. Comparison of the results from using different preset pixel patterns indicated that the “Bull’s Eye 4” pattern was more consistently effective than others at detecting both linear and rectangular/square impervious features with relatively lower amounts of misclassified false positive clutter in pervious areas. Consequently we used the Bull’s Eye 4 pattern with a width of 17 pixels (Figure 2). The inner portion of the pattern is designed to detect rectangular man-made shapes, while the outer ring of pixels helps to also detect linear shapes such as roads.

Figure 2: “Bull’s Eye 4” classification filter pattern.



(2) Input bands were the seven mentioned previously: the panchromatic band, edge-enhanced, NDVI, and the red, blue, green, and near infrared (NIR) bands from the multispectral image. All were set as reflectance bands. The resampling factor was set to 1 (no resampling). The option to create a wall-to-wall classification was left unchecked in order to return unclassified pixels (these were later to be converted to the “pervious” class). The option to “find rotated features” was selected in order to detect features such as buildings and roads oriented in all different directions. To reduce clutter, the classifier was set to aggregate small regions with a minimum of 100 pixels.

(3) The masking settings were used to ensure that the classified area was identical to that of the correct image tile extent.

(4) In the output options settings the output format selected was vector. The post-processing option to aggregate small regions is used to consolidate or eliminate result polygons that are smaller than a specified size. This option was selected with the minimum area set to 100 pixels, which eliminated smaller “island” features with a size of 36 m^2 or less. This helped reduce clutter.

The classifier model created using these settings was applied to 18 of the 21 tiles using two-class (impervious and shadow) training polygon sets devised for each tile. Three tiles (321, 411, and 6) contained so few impervious surfaces that it made more

sense to classify them manually. The output from the classification process is a two-class vector layer.

5.C. Revising the Classification: Hierarchical Learning, Post-processing, Cleanup

One of the main advantages of the Feature Analyst classifier is the array of tools used to iteratively refine the classifier until the desired result is produced. This process consisted of some or all of the following steps:

1) Refining the training set, as necessary.
2) Training the Learner to recognize the targets through a sequence of Hierarchical Learning passes with adjustments, as necessary, to the input representation.
3) Masking out troublesome features, if necessary.
4) Removing clutter (noise) and finding missed features by correcting the Learner through the selection of correct, incorrect, and missed examples.
5) Repeating the hierarchical learning process until satisfied with results.

Typically, the first step was to split the classified output vector layer into separate layers to be cleaned up individually as necessary based on visual inspection of the results. Based on this, one or both of the following two sets of tools in Feature Analyst were used iteratively to correct obvious misclassifications and to incorporate missed objects.

The Remove Clutter tools are used to tell the classifier what it did right and what it did wrong. It works by selecting whole polygons and digitizing portions of those features that were correctly and incorrectly classified in order to retrain the classifier more precisely. The classifier then runs another learning pass based on the new information and a new output layer is created. This may be repeated as many times as is necessary to achieve the desired result.

The Add Missed Features tool is best used once a good classification has been obtained, but some key features were still left out. The user draws polygons to capture examples of the missing features, and the classifier uses the information to retrain and produce an improved layer.

These two tool sets were used to retrain the classifier as needed until a satisfactory classification was produced, as determined from visual inspection based on main criteria of relatively small/minimal amount of clutter, no obvious systematic or major misclassifications, and generally good capturing of impervious objects throughout the image.

During the re-classification runs, the aggregation minimum threshold for the shadow class was lowered to 25 pixels in order to pick up additional shadow effects that were excluded due to the higher 100 pixel threshold, mentioned above in section B4.

Finally, the cleaned-up impervious layer was recombined with the shadow layer, with settings adjusted such that (a) the cleaned-up impervious layer takes precedence for conflicting pixels, and (b) any pixels not included in either of the two classes during the union of layers are classified into polygons. These now-classified “background” polygons, i.e. the remainder of the image that was left unclassified before, form the pervious class. The resulting three-class polygon layer was converted to raster format and pixel values were changed using the Reclass command in ArcMap as necessary into 1 (impervious), 2 (pervious), and 3 (shadow) for consistency across tiles. Lastly, the raster image was clipped using the tiled panchromatic raster layer as a mask in order to remove any zero-value collar pixels. The image was then ready for accuracy assessment in ERDAS Imagine

6. Results and Classification Accuracy

The accuracy of each classified image tile was assessed in ERDAS Imagine software using a stratified random sampling approach with a minimum per-class sample size of 50 points to ensure that classes which comprised a very small geographic area in a tile were still adequately sampled. Although the impervious class was often a small part of the classified image compared to impervious in terms of area, it was extremely important; consequently, in most cases the impervious class was intentionally over-weighted in terms of number of sample points. Three images (321, 411, and 6) were classified manually because there was so little impervious area. Since the small impervious areas in these images were easily digitized and all remaining area was impervious, it was not necessary to create a shadow class.

In-field validation of sample points was not feasible due to the large number of total points needed to evaluate the accuracy of all 21 tiles. Alternatively, high resolution ancillary images can be used to determine the accuracy of sample points; in this case we used the highest resolution data available, which was a combination of the QuickBird panchromatic and pan-sharpened multispectral images, aerial photos and "streetview" photographs of locations on the ground available online through GoogleEarth and GoogleMaps Street View, and experiential knowledge gained through previous work (Finke et al. 2007) on the La Terraza subdivision. Using one or more of these data sets helped to distinguish similar surfaces such as pervious gravel and xeriscaped yards from fully impervious roads, parking lots, and roofs. There were 5,257 sample points total for all 21 scenes combined (Table 1). This is a much more extensive sample compared to a recent study using a similarly large amount of QuickBird data by Campos et al. (2010), who used a total of 1,735 accuracy assessment points.

Table 1. Sampling points for accuracy assessment

Image Tile ID	Number of Sample Points
111	275
112	225
121	275
122	200
211	475
212	275
221	275
222	275
311	275
312	275
321	175
322	275
411	175
412	275
421	225
422	275
511	200
512	275
521	200
522	200
6	175
Total:	5,275

Error matrices were computed for each classified image tile in ERDAS Imagine (example shown in Table 2) and used to calculate standard accuracy metrics of percentage pixels correctly classified and the Kappa coefficient of agreement (Congalton, 1991; Janssen and van der Wel, 1994). From the error matrices we calculated overall accuracy as a percentage correct out of all sample points, and producer's and user's accuracy of each individual class (Table 3). Producer's accuracy indicates how well a reference pixel was classified; user's accuracy indicates the degree of reliability that the different map classes are accurate representations of those categories on the ground (Jensen, 2004, p. 505-506). The Kappa coefficient of agreement, a metric for accuracy that factors in the probability the result could have been produced by chance, was also calculated for overall classification and for individual classes (Table 4).

Mean overall accuracy for the 21 classifications was 89.7% with a range of 82.9-100%. 100% accuracy was associated only with those images that were classified manually. Excluding those three images, the mean overall accuracy was 88.1% which was not significantly lower ($t = 1.213$; $p > 0.05$; $df = 34$). Mean producer's accuracy for the impervious class (94%) was higher than for the pervious class (85%) ($t = 4.344$; $p < 0.05$; $df = 33$). Mean user's accuracy for the impervious class was accurate at 84%, but significantly lower than the pervious class (97%) ($t = -5.568$; $p < 0.05$; $df = 26$). The mean Kappa coefficient was higher for the pervious class than the impervious class ($t = -5.087$; $p < 0.05$; $df = 31$).

Table 2. Sample error matrix for tile 221. The main diagonal (in bold) indicates the number of correctly classified pixels.

	Predicted Class			
Actual Class	Impervious	Pervious	Shadow	Row Total
Impervious	**75**	25	0	100
Pervious	2	**96**	2	100
Shadow	10	0	**65**	75
Column Total	87	121	67	**236**

Table 3. Classification accuracy in percentage correct. N/A indicates tile was classified manually.

Image Tile ID	Overall Accuracy (%)	Impervious Producer's Accuracy (%)	Impervious User's Accuracy (%)	Pervious Producer's Accuracy (%)	Pervious User's Accuracy (%)	Shadow Producer's Accuracy (%)	Shadow User's Accuracy (%)
111	89.8	98.8	82.0	79.4	100.0	98.5	86.7
112	84.4	91.5	72.0	77.5	100.0	97.3	72.0
121	90.9	91.6	87.0	85.8	97.0	98.5	88.0
122	96.0	100.0	96.0	92.6	100.0	100.0	88.0
211	91.2	88.4	95.0	92.2	94.0	98.2	73.3
212	88.4	91.0	81.0	87.4	90.0	86.8	96.0
221	85.8	86.2	75.0	79.3	96.0	97.0	86.7
222	83.6	91.4	74.0	72.4	97.0	98.3	78.7
311	82.9	91.0	71.0	73.3	96.0	92.4	81.3
312	89.8	87.1	88.0	86.5	96.0	100.0	84.0
321	100.0	100.0	100.0	100.0	100.0	n/a	n/a
322	86.6	87.1	81.0	80.5	99.0	98.3	77.3
411	99.4	100.0	98.7	99.0	100.0	n/a	n/a
412	87.3	96.2	75.0	76.3	100.0	98.5	86.7
421	85.8	100.0	74.7	76.3	100.0	97.4	74.0
422	91.3	90.0	90.0	87.6	99.0	100.0	82.7
511	89.0	100.0	70.0	84.8	95.0	90.6	96.0
512	89.1	94.6	87.0	84.6	93.0	89.0	86.7
521	85.0	100.0	74.0	84.3	86.0	77.1	94.0
522	89.0	93.6	88.0	87.9	94.0	87.0	80.0
6	99.4	100.0	98.7	99.0	100.0	n/a	n/a
Mean	89.7	94.2	83.7	85.1	96.8	94.7	84.0

Table 4. Overall and per-class Kappa coefficients (k) with confidence intervals for overall k.

Image Tile ID	Overall k	Impervious k	Pervious k	Shadow k	95% C.I. upper	95% C.I. lower
111	0.85	0.74	1.00	0.82	0.90	0.79
112	0.75	0.62	1.00	0.66	0.81	0.69
121	0.86	0.80	0.95	0.84	0.91	0.81
122	0.94	0.95	1.00	0.85	1.00	0.87
211	0.86	0.91	0.89	0.70	0.90	0.81
212	0.82	0.72	0.84	0.94	0.88	0.77
221	0.78	0.63	0.93	0.82	0.83	0.74
222	0.75	0.63	0.94	0.73	0.82	0.69
311	0.74	0.60	0.92	0.75	0.81	0.67
312	0.85	0.81	0.93	0.79	0.90	0.79
321	1.00	1.00	1.00	n/a	1.08	0.92
322	0.79	0.71	0.98	0.71	0.85	0.74
411	0.99	0.98	1.00	n/a	1.07	0.91
412	0.81	0.65	1.00	0.82	0.86	0.75
421	0.77	0.66	1.00	0.69	0.84	0.70
422	0.87	0.84	0.98	0.78	0.92	0.82
511	0.82	0.64	0.89	0.95	0.89	0.75
512	0.83	0.80	0.88	0.82	0.89	0.78
521	0.76	0.68	0.71	0.91	0.84	0.68
522	0.82	0.84	0.87	0.74	0.89	0.75
6	0.99	0.98	1.00	n/a	1.07	0.91
Mean	0.84	0.77	0.94	0.80	0.90	0.78

7. Reclassification of Shadow Pixels

In order to produce a final classified image without any shadow pixels, we needed a way to reclassify them systematically and consistently. The problem this posed was that it is sometimes difficult to know for sure what kind of surface a shadow may be obscuring, whether impervious or pervious. We reasoned that a shadow feature cast in an undeveloped area would be surrounded by pervious pixels and therefore those surrounding pixels could be used to reclassify the shadow pixels. The same would apply to impervious surfaces in the developed areas, e.g. a shadow cast by a building over a parking lot would be surrounded by impervious pixels, which could be used to reclassify the shadow as impervious. The “neighborhood analysis” tool in ERDAS Imagine uses this basic logic to reclassify pixels of a particular class based on the surrounding pixel values. We set the operation to pass a 3x3 window over shadow pixels and reassign their value based on the majority value of the surrounding pixels in the “neighborhood.” The presence of somewhat large, contiguous shadow areas required us to replicate this process and run it iteratively in MATLAB to ensure that all shadow pixels were reclassified. Imagine did not have the capability to automatically repeat the neighborhood analysis. This process was run on every classified image that had any shadow pixels, producing a reclassified set of image tiles with only two classes, impervious and pervious.

8. Mosaic Creation and Final Image

The images resulting from the neighborhood analysis in MATLAB contained background values of 0 in the collar around certain images. These values were

converted to 'no data' to eliminate them to ensure a clean mosaic. The mosaicing process was done in ERDAS Imagine using the Mosaic Tool. The tool was set to generate weighted cutlines automatically for overlapping portions of images, except in two instances where this setting resulted in small holes of no data. In these cases geometric cutlines were used instead.

The final resulting mosaicked image was 3.59 GB in size with a pixel resolution of 0.6 m and a total area of 1,179.5 km^2. The total area is less than the original nominal area due to clipping to the watershed boundary and the elimination of overlapping areas among the six QuickBird images. Based on this image, the study area contains an estimated 18.6 km^2 of impervious surfaces, or 1.6% of the total area, mainly concentrated in the highly developed core of Sierra Vista, and to a lesser extent Ft. Huachuca.

9. Sources of Error

As stated in section 4 above, computing power was not sufficient to run the Feature Analyst classifier on any one of the six QuickBird files plus derived image files, much less a mosaic of the six scenes. This necessitated breaking them up into separate tiles which had to be classified individually using different training data sets. Although the calculated classification accuracies were satisfactory, the fact that each image tile was classified based on different training data unavoidably resulted in some discrepancies among the tiles which are evident from visual inspection. For example, roads in one image tile may be more conservatively classified than in another, resulting in edge effects in the final mosaic where the road is wider on one side of a cutline than on the other. Had it been possible to classify the entire study area at once, only one training data set would have been used and edge effects would not have resulted. Inconsistencies among classifications such as edge effects thus ultimately stem from limitations of processing power.

In each classified image tile, the goal was to strike a balance between accuracy and efficiency in terms of user time spent (e.g. preparing and iteratively revising training polygons) and computer processing time. With the iterative revising tools within Feature Analyst it is possible to eventually achieve a very high level of accuracy. However, we found that in many cases this would require a significant time investment on the part of the user. Further, after a few iterations of re-training the classifier, a point of diminishing returns was often reached where the gains in additional accuracy (based on visual inspection and comparison) were relatively small compared to the amount of time required to achieve them. Each image tile had its own unique set of features, and consequently this point of diminishing returns was different for each tile. A perfect classification is of course not possible, and in each tile some tradeoffs had to be made; for instance, in several cases achieving an extremely high degree of accuracy for roads in a given tile was not possible without producing a large amount of false positives for those natural surfaces with similar spectral and shape characteristics to impervious roads, which would lead to lower user's accuracy (i.e. greater errors of commission). In this case, creating a classification model that produced a balance between the two, i.e. adequate classification of roads without a high amount of false positives in pervious areas, meant necessarily sacrificing some accuracy for roads. Discrepancies that are apparent from visual inspection in the final image are the result of these unique tradeoffs that had to be made in each tile based on the variety of features that it contained. In sum, each classified tile reflects the best effort of the user to achieve a satisfactory level of accuracy (as quantified by the traditional metric of the error matrix) within a reasonable amount of time invested in refining the initial classification.

After examining the initial classification result, an important source of confusion was identified, which had to be addressed iteratively in post-classification cleanup procedures. This confusion was primarily between rectangular building structures with high reflectances and pervious (although likely highly compacted) natural surfaces with similar reflectances and shapes. In many cases it was evident from ancillary data that the natural surfaces were impervious rock. In this case, during accuracy assessment we counted it as correct if it was classified as impervious. However, there were also instances where the surface was not rock, but rather brightly-reflecting bare soil or a graded unpaved road with similar rectangular shape to buildings. When these areas were classified as impervious, they were considered a misclassification. Cleanup tools in Feature Analyst were useful for removing this clutter to a significant degree, but total accuracy was impossible. Additionally, roads and other surfaces are not spectrally homogeneous, and a training polygon in one part of a road may not be adequate to make the classifier capture all parts of a road feature. Training data sets took this into consideration, but could not always prevent gaps in road features. In many cases, creating a training polygon around one portion of a road led to more false positives in other spectrally similar non-road objects. When this limitation was found, we generally tried to be conservative, sacrificing some road accuracy in order to prevent false positives. Some amount of this type of error was found to be practically unavoidable, but could be improved to some extent using the hierarchical learning tools described previously.

Kappa values indicate that the classifier was more accurate for the pervious classes than the impervious classes, although impervious Kappa values were mostly very good, with a mean value of 0.77. In all tiles, the Kappa values show that overall and per-class accuracies were not due to chance. Percentage accuracy results show that while producer's accuracy was quite high for impervious classes, the main source of error was found in the user's accuracy. In other words, sample pixels that were actually impervious were classified at a high level of accuracy, while those pixels that were not impervious were more frequently classified incorrectly as impervious. Visual inspection of the classified results showed that this is to some degree related to the inability of the classifier to distinguish boundaries and edges between certain adjacent features. For example, in many cases the roads outside of the highly developed core of Sierra Vista appeared in the imagery to have a somewhat fuzzy gradient from the paved surface to the shoulder to the undeveloped ground. Bright road stripes and unpaved medians are separate objects, making training data creation difficult and consequently road classification difficult as well.

Appendix A

The procedures described above were designed to produce an estimate of the total amount of completely impervious surfaces, e.g., pavement, in the Sierra Vista subwatershed as of late 2009. However, highly developed land uses such as commercial and higher-density residential result in soil compaction even when the soil surface is not completely sealed with an impervious surface such as pavement. This may reduce permeability of soil substantially relative to undeveloped bare soils.

Previous estimates such as the Regional GAP Analysis Project estimated degrees of imperviousness at the scale of a 30x30 m grid cell, since most urban features are too small to be resolved at that resolution. Although the high resolution (0.6 m panchromatic, 2.4 m multispectral) of QuickBird imagery enables the resolution of most urban features, the classification method described above was designed to identify only completely impervious surfaces and consequently could not capture compacted soils which were undeveloped. In order to account for some of these areas, buffer zones of 1.8 and 3 m were created around the impervious surface polygons in the vector version of the map derived from QuickBird imagery. These buffer distances were based on the assumption that soils that are adjacent to completely impervious developed surfaces are likely also compacted to some degree relative to their undisturbed condition. The estimated area of totally impervious surfaces is 18.6 km^2. Generating a 1.8 m buffer zone added 9.7 km^2 of impervious or highly compacted surfaces, increasing the total impervious or compacted area in the subwatershed to 28.3 km^2. Creating a 3 m buffer zone added 15.4 km^2 to the original impervious area, for a total impervious or compacted area of 34 km^2.

Appendix B

The USPP Technical Committee requested data and analysis of changes in the amount and distribution of increases in impervious surfaces in the Sierra Vista area since 2001. This appendix reports results of two related comparisons. The first estimates the change in impervious surfaces in the Sierra Vista subwatershed, 2001-2009 based on National Land Cover Dataset (NLCD) products for 2001 and 2006 (Fry et al., 2011) and includes the most recent estimates obtained using the methods described in the main report for comparison. The second compares two estimates for the Coyote Wash watershed.

1. Increases in Impervious Surfaces in the Sierra Vista Subwatershed, 2001-2009

Data sets used in the 2001 to 2006 comparison:
- NLCD 2001 impervious surface area version 2.0[1] (Fig. 1)
- NLCD 2006 impervious surface area (Fig. 2)
- NLCD 2001-2006 percent change in impervious surface area
- Sierra Vista subwatershed boundary polygon

[1] USGS states that NLCD 2001 version 2.0 revisions were designed to make the 2001 maps directly comparable with the 2006 NLCD.

Figure 1. NLCD 2001 impervious surface map. Sierra Vista subwatershed boundary shown in yellow.

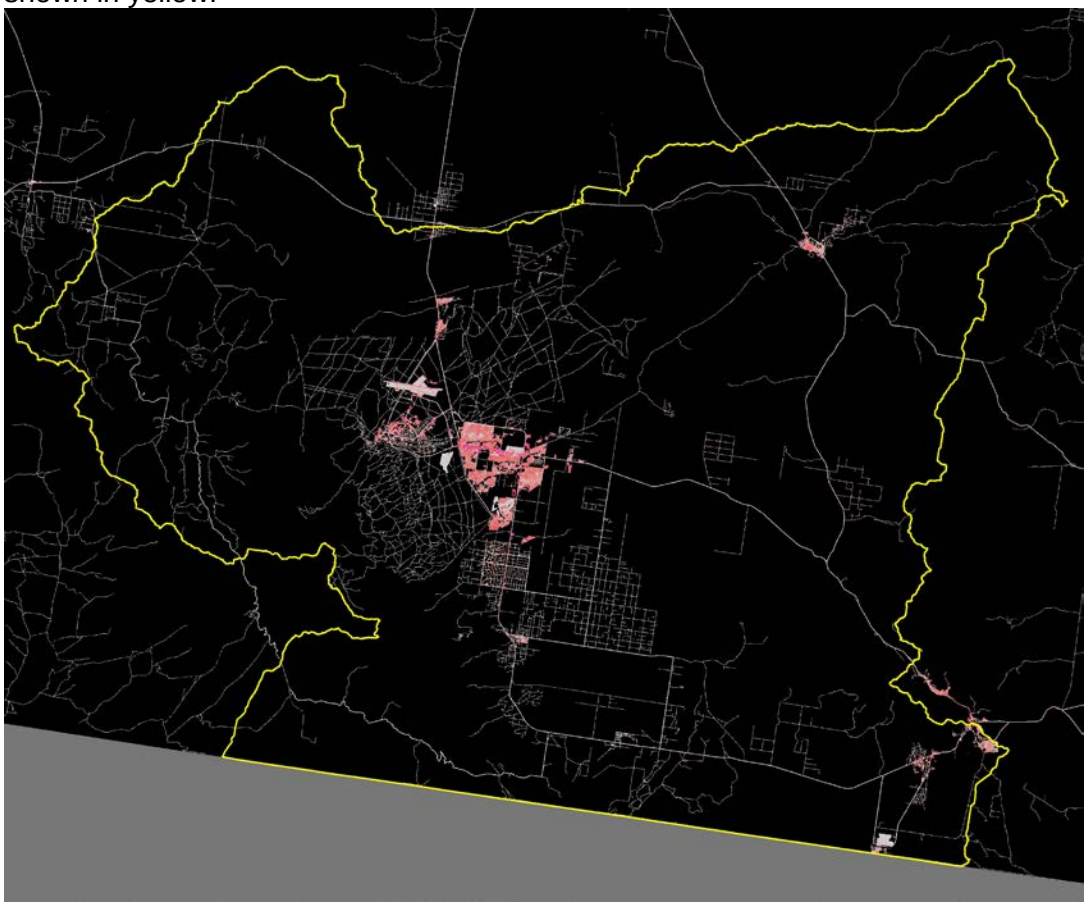
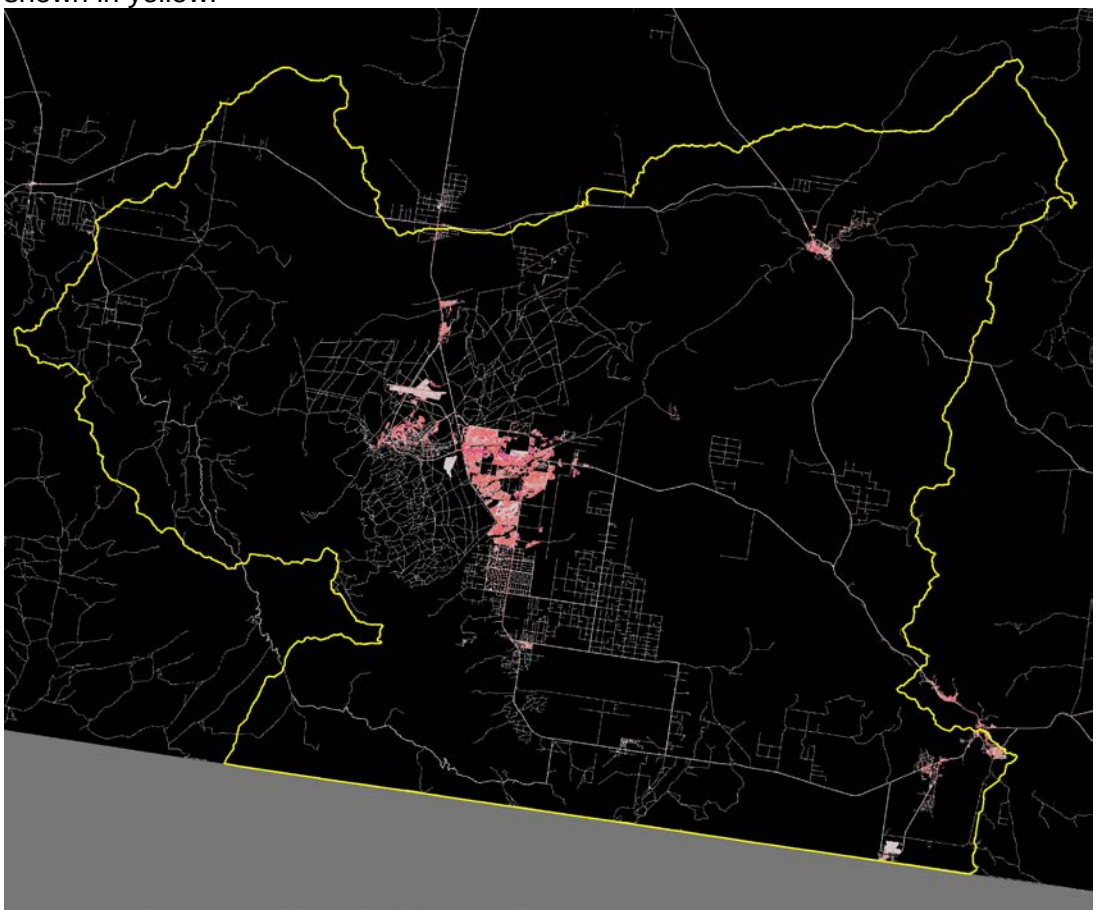


Figure 2. NLCD 2006 impervious surface map. Sierra Vista subwatershed boundary shown in yellow.



Processing Steps:

1. The Sierra Vista subwatershed boundary polygon was reprojected to match the NLCD data sets
2. All three NLCD raster data sets were clipped by the subwatershed boundary polygon extent.

In the 2001 Version 2 and 2006 NLCD imperviousness data sets, each pixel value is assigned a percentage. This is included in the raster attributes table in integer form in a column titled "value." Also included is the count of grid cells with that percentage of impervious area. To find the total area in the clipped images, the counts (number of 30m x 30m cells) were multiplied by 900 (the area of each cell in square meters) and summed. To find the estimated impervious areas, the "value" integer was divided by 100 to create decimal values, which were then multiplied by the total area in m^2 associated with each impervious value and then summed over all pixel values (0-100%). Total area and total impervious area were converted to km^2. The total area of the subwatershed boundary is 2,671.17 km^2. Results are shown in Table 1. Added to these results is the estimated impervious surface area obtained from Feature Analyst object-oriented classification of 2009 QuickBird imagery covering most of the developed part of the

subwatershed. This area of 18.66 km^2 is an increase of 2.55 km^2 (16%) compared to 2006 and 4.52 km^2 (32%) compared to 2001.

Table 1. Three chronological estimates of impervious surface areas in the Sierra Vista subwatershed.

Estimate Year and Source	Impervious Surface Area (km^2)	Increase Over Prior Estimate (km^2)
2001 NLCD v.2.0 (USGS)	14.14	-
2006 NLCD (USGS)	16.11	1.97 (14%)
2009 QuickBird (USDA-ARS SWRC)	18.66	2.55 (16%)

The NLCD impervious surface change data set shows where the increase in impervious surface between 2001 and 2006 happened, and where it was most pronounced. The zoomed out view in Fig. 3 shows that nearly all the additional impervious surfaces were in Sierra Vista, with some around the Ft. Huachuca area. The most substantial contiguous developments were located in the southern end of Sierra Vista. Figure 4 shows the Sierra Vista area in closer detail, with the 2009 Quickbird-based estimates created as described in the main report above superimposed to provide context. In both images the magnitude of change from 2001-2006 is represented by the darkness of coloration, with gray pixels denoting relatively small increases in impervious surfaces and dark red indicating more significant increases.

Figure 3. Location and intensity of increases in impervious surfaces in Fort Huachuca and Sierra Vista between 2001 and 2006 based on NLCD land cover change products.

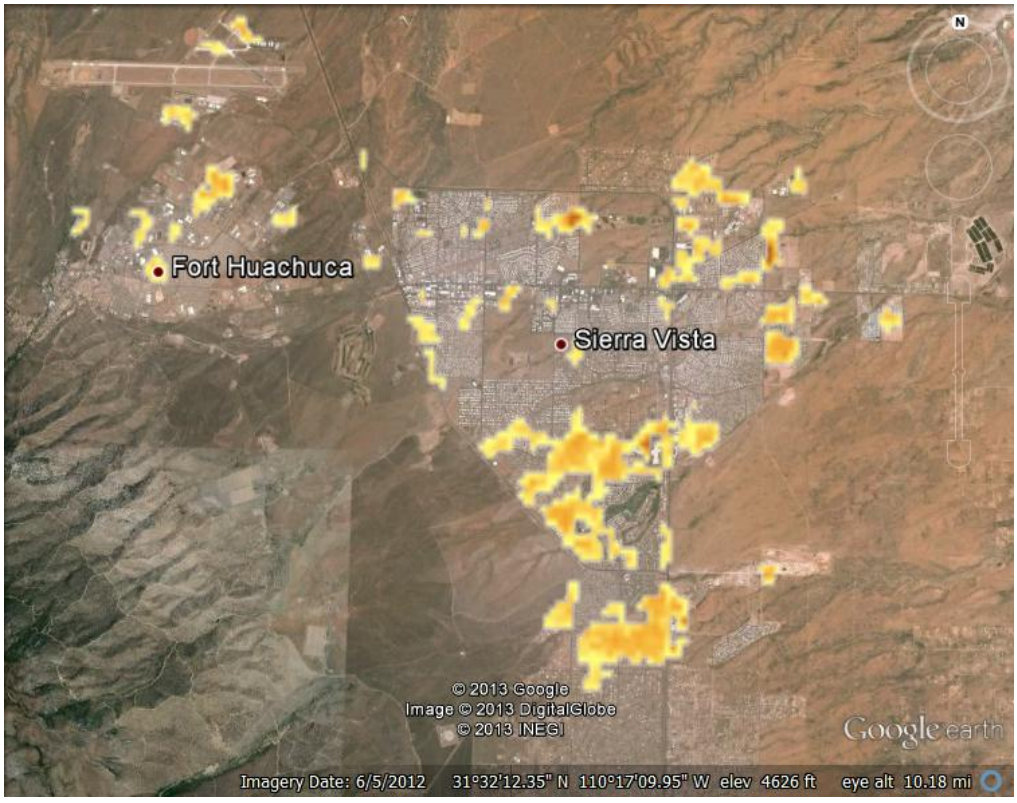


Figure 4. Increase and distribution of impervious surfaces in Sierra Vista based on NLCD land cover products. Gray and red coloration represents impervious surfaces added in Sierra Vista between 2001 and 2006, with darker red coloration indicating more intense development. The most recent estimates of total impervious surfaces based on Dec. 2009 QuickBird imagery are overlaid in yellow to provide context.

2. Impervious surfaces in Coyote Wash, 2001 and 2009.

A previous analysis by Levick and Goodrich (2005) used land cover data from the 2001 Southwest Regional GAP Analysis Project (GAP) and 1997 North American Land Cover (NALC) Classification map to estimate the amount of developed land within the Coyote Wash watershed. They found that 6.2 km^2 (13.6%) of Coyote Was was classified in the GAP map as “developed, open space – low intensity” and 14.2 km^2 (31.1%) was classified as “developed, medium – high intensity.” NALC 1997 estimated 18.9 km^2 of “urban” (high and low density).

The area of impervious surfaces inside Coyote Wash based on the map layer created for the present study was calculated for comparison. This produced an estimated 5.4 km^2 of completely impervious surface within Coyote Wash watershed, or 11.8% of the total area. With a 1.8 m buffer added to account for both impervious and highly compacted soils, the area increases to 8.2 km^2 (18% of Coyote Wash); a 3 m buffer covers an area of 9.8 km^2 (21% of Coyote Wash). Although this is actually lower in area than the proxies for imperviousness in the GAP and NALC maps, this should not be construed as demonstrating that impervious surfaces have decreased between 2001 and 2009. This is related to the very different spatial resolutions and classification schemes used in the present study compared to NALC and GAP maps, which represent developed areas as percentages of 900 m^2 grid cells, rather than an either/or classification.

10. References

Campos, N., Lawrence, R., McGlynn, B., & Gardner, K. (2010). Effects of LiDAR-Quickbird fusion on object-oriented classification of mountain resort development. *Journal of Applied Remote Sensing*, *4*(043556), 043556.

Coes, A.L., and Pool, D.R. (2005). Ephemeral-stream channel and basin-floor infiltration and recharge in the Sierra Vista subwatershed of the Upper San Pedro Basin, southeastern Arizona: U.S. Geological Survey Open-File Report 2005–1023, 67 p.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, *37*(1), 35–46. doi:10.1016/0034-4257(91)90048-B

Finke, T., Moran, S.M., Yool, S.R. (2007). Object-oriented classification to map impervious surfaces for hydrologic models. Unpublished manuscript.

Fry, J., Xian, G., Jin, S., Dewitz, J., Homer, C., Yang, L., Barnes, C., Herold, N., and Wickham, J. (2011). Completion of the 2006 National Land Cover Database for the Conterminous United States, *PE&RS*, Vol. 77(9), 858-864.

Goodrich, D.C., Williams, D.G., Unkrich, C.L., Hogan, J.F., Scott, R.L., Hultine, K.R., Pool, D., Coes, A.L., Miller, S. (2004). Comparison of methods to estimate ephemeral channel recharge, Walnut Gulch, San Pedro River Basin, Arizona. In: Groundwater Recharge in a Desert Environment: The Southwestern United States, J.F. Hogan, F.M. Phillips and B.R. Scanlon (eds.), Water Science and Applications Series, Vol. 9, American Geophysical Union, Washington, DC, pp. 77-99.

Janssen, L. L. F., & Van der Wel, F. J. M. (1994). Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, *60*(4), 419–426.

Jensen, J. R. (2004). *Introductory Digital Image Processing* (3rd ed.). Prentice Hall.

Kennedy, J. R., Goodrich, D. C., & Unkrich, C. L. (2013). Using the KINEROS2 modeling framework to evaluate the increase in storm runoff from residential development in a semi-arid environment. *Journal of Hydrologic Engineering, 18,* 698-706.

Levick, L.R. and Goodrich, D.C. (2005). 2005 Revised Urban Recharge Estimates Using Urban Area Calculations for the Upper San Pedro Basin (Tombstone Gage to US/Mexico Border) using NALC 1997 and Southwest GAP 2001 Land Cover Maps. USDA-ARS SWRC, Tucson, AZ.