Notice to Data Users:
The documentation for this data set was provided solely by the Principal Investigator(s) and was not further developed, thoroughly reviewed, or edited by NSIDC. Thus, support for this data set may be limited.

SMEX04 Land Use Classification Data

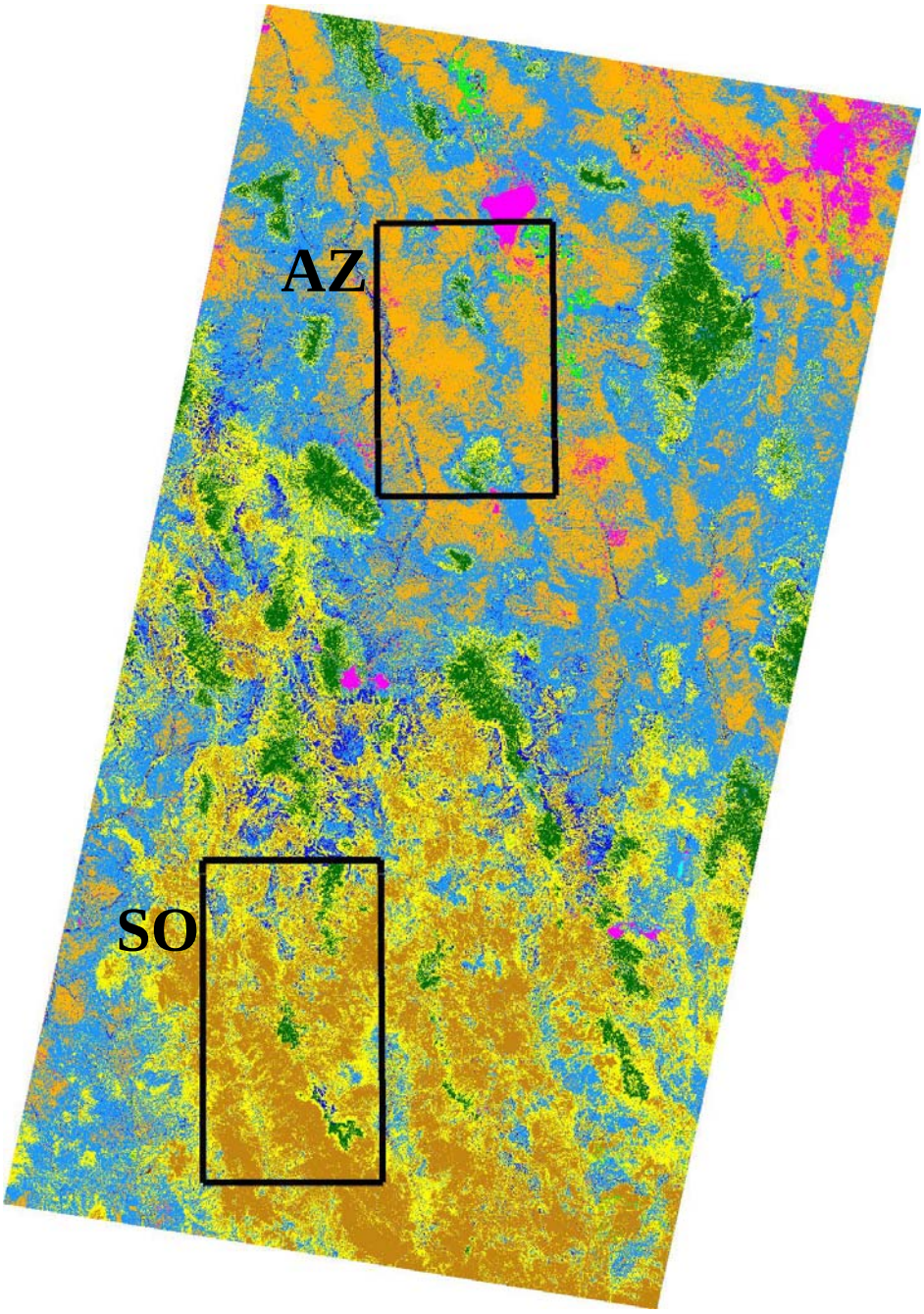


Figure 1. SMEX04 Regional Study Areas: Arizona (AZ) and Sonora (SO)

Summary

Land cover classification is necessary for modeling and scaling hydrologic variables of concern in the SMEX04 experiment. For the Arizona and Sonora study regions, six Landsat 5 Thematic Mapper TM scenes (Path 35/Row 38-39 for June 11, July 29, and August 30, 2004) were acquired and atmospherically corrected using MODTRAN. Ground truth information collected during the experiment was used for decision tree parameterization. Using these 6 scenes, high accuracy was achieved with respect to the test data set, and high qualitative accuracy was achieved when comparing the land cover to a SMEX04 classification image for the same regions (Yilmaz et al., 2007).

Citing These Data

The following example shows how to cite the use of this data set in a publication. List the principal investigators, year of data set release, data set title, and publisher.

Hunt, Jr., E. R., T. J. Jackson, and M. Tugrul Yilmaz. 2009. *SMEX04 Land Use Classification Data.* Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center.

Overview Table

Category	Description
Data format 8-bit	Binary
Spatial coverage	32.7° N, 111.4° W to 29.3° N, 108.6° W
Temporal coverage	June 11 2004, July 29 2004, August 30 2004
File naming convention SMEX04_L AND_COVER.bin	
File size 100	MB
Parameter(s)	Land Cover Classification 0 Background 1 Water Body 2 Unvegetated (Bare Soil) 3 Shrubland 4 Grassland 5 Riparian Mesquite 6 Riparian Woodland 7 Sparse Woodland 8 Evergreen (Broadleaf +Needleleaf) 9 Subtropical Shrub 10 Agriculture
Procedures for obtaining data	Data are available via FTP.

Table of Contents

1. Contacts and Acknowledgments
2. Detailed Data Description
3. Data Access and Tools
4. Data Acquisition and Processing
5. References and Related Publications
6. Document Information

1. Contacts and Acknowledgments:

Investigator(s) Name and Title:

E. Raymond Jr., Hunt, Physical Scientist, Thomas J. Jackson, Hydrologist, Tugrul Yilmaz, USDA-ARS-Hydrology Lab.

Technical Contact:

NSIDC User Services
National Snow and Ice Data Center
CIRES, 449 UCB
University of Colorado
Boulder, CO 80309-0449 USA
phone: (303) 492-6199
fax: (303) 492-2468
form: Contact NSIDC User Services
e-mail: nsidc@nsidc.org

Acknowledgements:

The USDA ARS Southwest Watershed Research Center, the University of Arizona and the many graduate students and volunteers who collected the field data. Funding was provided by NASA grant NNG04GQ426 to Dr. Susan Ustin, University of California. We thank Susan Ustin, Vern Vanderbilt, Charles Fernandez, Alfredo Huete, Ed Glenn, Pam Nagler, Lyssa Goins, John Schroeder, Phil Valco, Dave Darling, Ho-Jin Kim, Pan Sirikul, Young Wook Kim, Lin Li, Jose Ramon, David Riano, Johny Kefauver, Mi-Young Jang, and Rocco Panciera.

2. Detailed Data Description:

Format:

Data consist of a single 8-bit binary file and an associated header file in ENVI format.

File Naming Convention:

There are two files:

SMEX04_L	AND_COVER.bin
SMEX04_L	AND_COVER.hdr

File Size:

File size is approximately 100 MB.

Spatial Coverage:

Southernmost Latitude: 29.3° N
Northernmost Latitude: 32.7° N
Westernmost Longitude: 111.4° W
Easternmost Longitude: 108.6° W

Samples: 8550 Lines: 12171 pixel size 30 meters
UTM Coordinates, Zone 12, WGS-84
xmin: 460890 E xmax: 737010 E
ymin: 3244404 N ymax: 3619794 N

Temporal Coverage:

June 11, 2004, July 29, 2004, August 30, 2004, and daily images from interpolation

Parameter or Variable:

Land Cover Classification

0 Background
1 Water Body
2 Unvegetated
3 Shrubland
4 Grassland
5 Riparian Mesquite
6 Riparian Woodland
7 Sparse Woodland
8 Evergreen (Broadleaf + Needleleaf)
9 Subtropical Shrub
10 Agriculture

3. Data Access and Tools:

Data Access:

The data are stored in binary format and were generated via IDL/ENVI. A header file is attached with geolocation information.

Software and Tools:

Most image software will be able to input the data.

4. Data Acquisition and Processing:

Six Landsat TM scenes (June 11, July 29 and August 30, 2004) and one DEM image were used to construct the land cover classification for the SMEX04 Arizona and Sonora study regions. There was very little cloud cover in the June 11, 2004 (<1%), July 29, 2004 (<5%), and August 30, 2004 (<1%) images.

1) Atmospheric Corrections of six TM images were performed using MODTRAN.

2) Ground truth in Arizona and Sonora were collected with WAAS-enabled GPS units (~5m spatial accuracy) and converted to Regions of Interest (ROI).

3) NDVI and NDWI of the Atmospherically-corrected images were calculated using these equations.

$$NDVI = \frac{\rho(band4) - \rho(band3)}{\rho(band4) + \rho(band3)} \quad \text{and}$$

$$NDWI = \frac{\rho(band4) - \rho(band5)}{\rho(band4) + \rho(band5)}$$

4) NDVI and NDWI statistics were extracted.

5) Topographic features like slope and maximum curvature were computed (within ENVI) using the DEM image. The original DEM file is in 3 second spatial resolution. In order to get better features, slope and maximum curvature were computed over the original 3 second image and then the image was converted to a 30 meter resolution image.

6) In order to decrease the execution time, some other files were created such as Maximum NDVI, Maximum difference of NDVI, Average NDVI, Maximum NDWI, Maximum difference of NDWI (between June-July-August).

7) By creating an imaginary flora boundary which coincided with the U.S./Mexico border, the natural distinction between subtropical and semi-arid climates can be accounted for. The distraction was based on species differences, because the area was transitional between 2 floristic provinces.

8) A decision tree was constructed for classification (All Nodes and expressions used in the decision tree are given below).

9) The cloud cover is minimal for these scenes and the cloudy pixels were repairable with the other images, hence, a cloud mask was not used for classification.

10) DN values of the Ground truth were calculated over this image and confusion matrix is computed between this image and ground truth.

Table 1. Confusion Matrix

		Ground								
		Grass-land	Shrub-land	Riparian Mesquite	Riparian Woodland	Ever-green	Agri-culture	Sparse Woodland	Total	User's Accuracy%
Classified image	Grass	4	19	1		2			26	15.4
	Shrub	4	89			2			95	93.7
	Riparian Mesquite		1						1	0.0
	Riparian Woodland			1					1	0.0
	Evergreen					2			2	100.0
	Agriculture						1		1	100.0
	Sparse Woodland	1		2	2		1		6	
	Total	9	109	4	2	6	2	0	132	
Producer's Accuracy		44.4	81.7	0.0	0.0	33.3	50.0		Overall Accuracy	72.7%

Since the number of shrub sites was dominant among that of other sites, the overall accuracy was mainly affected by shrubland. In the same way, since the number of sites sampled as Riparian Mesquite and Riparian Woodland was very low compared with the total number of sites, accuracy of these did not affect the overall accuracy significantly.

Decision tree image

The 12 variables used in the decision tree for land cover are:

Band 1 : DEM image (Digital Elevation Model)
Band 2 : Maximum Curvature Image
Band 3 : Slope
Band 4 : Subtropical-semiarid border
Band 5 : NDVI June 11
Band 6 : NDVI July 29
Band 7 : NDVI average
Band 8 : NDVI maximum
Band 9 : NDVI maximum difference
Band 10 : NDWI June 11
Band 11 : NDWI maximum
Band 12 : NDWI maximum difference

Average NDVI is simply calculated by adding NDVI values of three dates divided by 3 (within ENVI). Maximum NDVI or NDWI is the maximum NDVI or NDWI value of three dates. Similarly, the maximum NDVI or NDWI difference is the maximum NDVI or NDWI increase or decrease over three dates. Also, slope and maximum curvature are created within ENVI by using "Topographic modeling" tool with a kernel size of 55 pixels.

Decision tree structure

Rule 1 Yes Rule 2
No Background (Class 0)

Rule 2 Yes Rule 6
No Rule 3

Rule 3 Yes Rule 5
No Rule 4

Rule 4 Yes Unvegetated (Class 2)
No Water Body (Class 1)

Rule 5 Yes Shrubland (Class 3)
No Grassland (Class 4)

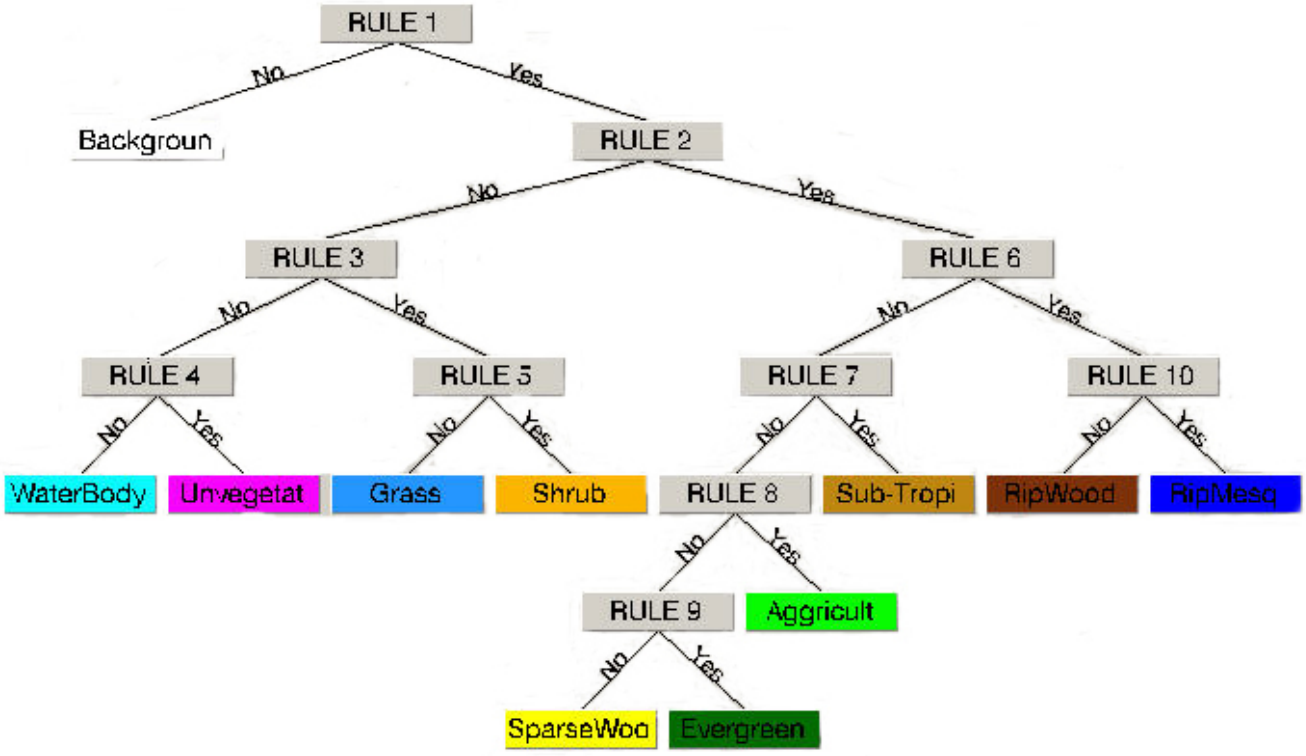
Rule 6 Yes Rule 10
No Rule 7

Rule 7 Yes Subtropical Shrub (Class 9)
No Rule 8

Rule 8 Yes Agriculture (Class 10)
No Rule 9

Rule 9 Yes Evergreen (Class 8)
No Sparse Woodland (Class 7)

Rule 10 Yes Riparian Mesquite (Class 5)
No Riparian Woodland (Class 6)



Decision tree rules

Rule 1 Band 1 > 200

Rule 2 Band 7 > 0.27 AND [Band 7 > 0.32 OR (Band 6 - Band 5) > 0.195]

Rule 3 Band 7 > 0.13

Rule 4 Band 7 > -0.05

Rule 5 Band 7 < 0.22 AND Band9 < 0.08

Rule 6 Band 12 < 0.26 AND Band3 < 5 AND Band 7 <0.53 AND (Band 1 <1550 OR Band 2 <30.3)
AND (Band 6-Band 5)<0.20 AND (Band 8-Band 5) < 0.275

Rule 7 Band 4 EQ 1 AND Band 5 < 0.31 AND (Band 6-Band 5) >0.23 AND Band 2 < 30.3 AND Band 7 > 0.40 AND Band 1 < 1550

Rule 8 Band 7 > 0.40 AND Band 1 < 1550 AND Band 3 < 1.5 AND (Band 9 > 0.45 OR Band 8 > 0.80 OR Band 11 > 0.40 OR Band 12 > 0.50)

Rule 9 (Band 1 > 2000 AND Band 5 < 0.7 AND Band 10 < 0.23) OR (Band 1 > 1700 AND Band 1 < 2000 AND Band 10 < 0.17 AND Band 7 > 0.40) OR (Band 1 > 1300 AND Band 1 < 1700 AND Band 2 > 30 AND Band 10 > -0.10 AND Band 7 >0.40)

Rule 10 Band 10 < 0.04 AND Band 5 < 0.47

5. References and Related Publications:

References

Yilmaz, M. Tugrul, E. Raymond Hunt, Jr., Lyssa D. Goins, Susan L. Ustin, Vern C. Vanderbilt, and Thomas J. Jackson. 2007. Vegetation Water Content During SMEX04 from Ground Data and Landsat 5 Thematic Mapper Imagery. *Remote Sensing of Environment* 112: 350–362, doi:10.1016/j.rse.2007.03.029.

Related Data Collections:

Refer to the USDA SMEX04 Web site for in-depth information on the science mission and goal of the SMEX project: http://hydrolab.arsusda.gov/smex04/.

6. Document Information:

Document Creation Date:

18 February 2005