

UTILIZAÇÃO DE DADOS SAR NA CLASSIFICAÇÃO DE ESPÉCIES AGRÍCOLAS DE PRIMEIRA E SEGUNDA SAFRA

Victor Hugo Rohden Prudente¹, Denis Corte Vieira¹, Bruno Montibeller², Lucas Volochen Oldoni¹, Ieda Del'Arco Sanches¹ e Marcos Adami¹

¹Instituto Nacional de Pesquisas Espaciais – INPE, Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil {victor.prudente; denis.vieira; lucas.oldoni; ieda.sanches e marcos.adami}@inpe.br

²Department of Geography, University of Tartu, Vanemuise 46, 51003, Tartu, Estonia, bruno.montibeller@ut.ee

RESUMO

Dados dos sensores SAR possuem a vantagem de serem menos influenciados por presença de nuvens, possibilitando maior frequência temporal para monitoramento de alvos agrícolas. Diante disso, o objetivo deste trabalho foi utilizar dados SAR para discriminar diferentes espécies agrícolas e alvos naturais no município de Luiz Eduardo Magalhães – Bahia, em duas safras. Foi utilizada abordagem temporal, com 12 imagens SAR/Sentinel-1 tanto para a segunda safra de 2017 quanto para a primeira safra de 2018. O classificador utilizado foi o *Random Forest* e as amostras de treinamento foram adquiridas em visitas de campo. Entre os resultados, a classificação do algodão obteve as melhores acurácias para ambas as safras. Na segunda safra de 2017 houve confusão entre as classes de milheto, milho e sorgo e entre as classes de eucalipto, café e cerrado, além das classes grama e pastagem. Para a primeira safra de 2018 obteve-se acurácias melhores para separação das espécies agrícolas.

Palavras-chave — Random Forest, microondas, banda C, agricultura.

ABSTRACT

SAR sensors data have the advantage of being less influenced by the presence of clouds, thus allowing higher temporal frequency for the monitoring of agricultural targets. Therefore, the objective of this work was to use SAR data to classify different agricultural species and natural targets in the municipality of Luiz Eduardo Magalhães – Bahia, in two harvests. Temporal approach was used, with 12 SAR / Sentinel-1 images for the second harvest of 2017 and for the first harvest of 2018. The Random Forest algorithm was used, and the training samples were acquired at field visits. The classification of Cotton obtained the best results for both harvests. In the second crop of 2017 there was confusion among the classes of millet, corn and sorghum and among the classes of eucalyptus, coffee and cerrado, besides the grass and pasture classes. Better accuracy for the separation of agricultural species was obtained for the first harvest of 2018.

Key words — Random Forest, microwave, C-band, agriculture.

1. INTRODUÇÃO

As culturas agrícolas são alvos dinâmicos espacialmente e

temporalmente, assim a frequência temporal de dados é um fator importante para seu monitoramento [1]. Entretanto, devido a influência de nuvens, somente o uso de dados de sensores ópticos é desafiador [2]. Neste sentido, os sensores que operam na faixa de microondas, como sensores SAR (*Synthetic Aperture Radar*), se destacam por terem a capacidade de imagear durante dia e noite, além de serem menos influenciados pela presença de nuvens [3]. Isso possibilita maior disponibilidade de dados, quando comparado com os sensores ópticos, para o monitoramento de culturas agrícolas.

Contudo, a forma de interação da radiação de microondas com os objetos de superfície terrestre é complexa. Esta interação é influenciada por diferentes parâmetros do alvo (e.g. geometria (estrutura) e constante dielétrica do dossel, sistema de manejo, densidade de plantio e orientação da linha) e do sensor (e.g. ângulo de incidência e polarização da onda incidente) [4].

Esta complexidade de interação e o número reduzido de sensores com disponibilidade gratuita de dados, resulta em poucos estudos com dados SAR voltados para a agricultura [5, 6], principalmente para o Brasil. Portanto, mais estudos nessa área são necessários.

A frequência temporal de dados exigida pela agricultura, aliada a complexidade dos dados SAR, demanda classificadores específicos, com capacidade de lidar com a dimensionalidade dos dados. Dentre estes, o classificador *Random Forest* (RF) se destaca devido a sua robustez e capacidade de lidar com alto número de variáveis [7]. Além do mais, resulta em classificações acuradas de uso e cobertura da terra [8].

Outro fator preponderante é a capacidade computacional necessária para estas abordagens. Neste sentido, destacam-se as plataformas em nuvens, que permitem que o usuário tenha acesso ao dado, sem a necessidade de armazenamento local, e conferem uma melhor configuração para processamento. Destaca-se nesse sentido a máquina virtual *RUS Machine* da Agência Espacial Europeia (ESA/Copernicus) [9].

Diante disto, o objetivo deste trabalho foi utilizar dados SAR/Sentinel-1 para discriminar diferentes espécies agrícolas e alvos naturais, em Luiz Eduardo Magalhães, Bahia, na segunda safra de 2017 e na primeira safra de 2018.

2. MATERIAL E MÉTODOS

Á área de estudo compreende município de Luiz Eduardo Magalhães, no oeste do estado da Bahia (12°05'31"S e 45°48'18"O). Neste município foram levantados dados a campo nos meses de junho de 2017 (segunda safra) e no mês de março de

2018 (primeira safra) conforme apresentado por [10]. Identificou-se 18 classes entre espécies agrícolas e naturais, sendo elas feijão, cerrado, café, áreas de conversão, algodão, crotalaria, eucaliptos, gramínea, feno, milho, milho e crotalaria, milheto, outras culturas não comerciais, pastagem, sorgo, soja, trigo e solo não cultivado.

Para ambas as safras, foram consideradas as classes agrícolas de algodão, café, eucaliptos, feijão, feno, gramínea, milho, pastagem e solo não cultivado; e as classes naturais de cerrado e áreas de conversão. Devido ao menor número de amostras, para o ano de 2017, criou-se a classe “outras culturas” corresponde ao agrupamento das áreas com crotalaria, milho e crotalaria, trigo, soja e culturas não comerciais. Pelo mesmo motivo, em 2018, a classe milheto foi agrupada com a classe milho e a classe de outras culturas não comerciais foi agrupada com a classe de pastagem. Totalizaram-se 14 classes para a segunda safra de 2017 e 12 classes para a primeira safra de 2018.

Foram utilizados dados SAR do Sentinel-1A, banda C, modo de imageamento IW (*Interferometric Wide swath*), produto GRD (*Ground Range Detect*). As polarizações utilizadas foram VV e VH [11, 12]. Um cena foi necessária para garantir a cobertura quase que total de LEM. Foram utilizadas 12 imagens de cada polarização, tanto para a segunda safra de 2017, entre as datas de 01 de maio e 31 de setembro, quanto para a primeira safra de 2018, de 01 de dezembro de 2017 e 31 de abril de 2018.

Considerando o grande volume de dados a serem processados e analisados, utilizou-se a máquina virtual *RUS Machine* [9], para o pré-processamento e classificação dos dados. Nesta máquina virtual, foram utilizados os softwares *SNAP* [13] e *R* [14], para pré-processamento e classificação.

As imagens SAR foram submetidas aos pré-processos de calibração radiométrica, filtragem do ruído *speckle*, correção do terreno e transformação da escala linear para a escala logarítmica (dB). Com o processo de calibração radiométrica se torna possível comparar imagens SAR adquiridas em diferentes datas e/ou diferentes modos, ou processadas por diferentes processos [15]. Nesta etapa as imagens foram transformadas para sigma zero (σ^0) [16].

Posteriormente foi utilizado o filtro *Gamma Map 5 x 5* para a filtragem do ruído *speckle*. Estes ruídos são inerentes ao sistema SAR e causam efeito de salpicamento nas imagens [15]. Por último, foi realizada a correção das distorções topográficas de geometria presentes nas imagens SAR, utilizando dados auxiliares do modelo digital de elevação, *SRTM (Shuttle Radar Topography Mission)* de 30 metros [15, 16].

Após o pré-processamento, as imagens SAR, com ambas polarizações, VV e VH, foram agregadas em dois *stacks*, um para cada safra. Estes *stacks* foram utilizados para os processos de classificação. Foi utilizado o classificador não paramétrico *Random Forest* [17], no ambiente *R* [14], pacotes *caret* e *raster*. O número de árvores (*ntrees*) utilizadas foi de 500.

Também visando o processo de classificação, foram separados aleatoriamente 70% dos talhões de cada classe para o treinamento do classificador em cada ano. Os 30% restantes foram utilizados para validação das classificações. Os talhões de treinamento de cada ano foram utilizados para extrair as informações de retrospalhamento de todos os pixels contidos em seu interior,

formando dois conjuntos de treinamento. Dentro destes conjuntos, foram sorteados aleatoriamente 7000 pixels para cada classe para o treinamento do classificador RF. A acurácia do classificador é realizada por meio da técnica *Bootstrap*.

As amostras de validação, 30% dos talhões, foram comparadas com as classificações geradas pelo RF para cada ano, gerando a matriz de confusão para cada ano. Posteriormente foram calculadas as métricas de Exatidão Global (EG) e Acurácia do Produtor (AP) e Acurácia do Usuário (AU) [18].

3. RESULTADOS E DISCUSSÃO

Na etapa de treinamento do classificador RF, os valores de EG foram de 90,50% e 88,07%, para 2017 e 2018, respectivamente.

Os mapas das classificações para ambos os anos se encontram na Figura 1. Observa-se que há efeito de salpicamento nas classificações, principalmente em áreas de Cerrado e solo não cultivado. Na segunda safra de 2017 (Figura 1a) há maior presença de salpicamento, quando comparada com a classificação da primeira safra de 2018 (Figura 1b). Estas classificações tiveram EG de 60,14% em 2017 e de 86,06% em 2018. Além disso, no ano de 2017 há uma maior fragmentação das classes, enquanto que em 2018 predomina a cultura da soja.

As matrizes de confusão, AP e AU de 2017 e 2018 são apresentadas nas Tabela 1 e Tabela 2, respectivamente. De maneira geral, maiores valores de AP e AU foram obtidas para as classes mapeadas no ano de 2018 quando comparados com 2017, ou seja, os erros de exclusão e inclusão foram menores. Em 2017, os erros de inclusão foram maiores quando comparados com 2018, devido a inclusão de áreas das classes cerrado e solo não cultivado nas áreas de todas as demais classes.

Os maiores valores de AP para 2017 foram obtidos para a classe algodão, seguida do sorgo e milheto. Das áreas de algodão, apenas 2% não foram classificadas, sendo essas atribuídas à classe feno. O sorgo teve um erro de exclusão de aproximadamente 24%, pois houve muitos pixels dessa classe classificadas como milheto. O erro de exclusão do milheto foi de 26%, confundindo principalmente com as classes de solo, sorgo, milho e feijão.

Ainda em 2017, os maiores valores de AU foram obtidos para as classes de solo não cultivado e milheto. A principal classe incluída como solo não cultivado foi o milheto, e como milheto a classe solo não cultivado.

Quase que a totalidade das amostras de referências das classes outras culturas e eucaliptos em 2017 foram classificadas nas classes feijão e cerrado, respectivamente. Ainda em 2017, 84% da classe gramíneas não foi classificada como tal, sendo em sua maioria classificada como cerrado. A pastagem também teve valores baixos de AP, sendo suas amostras classificadas principalmente como solo, algodão e milheto.

Os maiores erros de inclusão da safra de 2017 correspondem às classes de outras culturas e eucaliptos. Fenos e solos não cultivados foram as principais classes classificadas como outras culturas, enquanto que cerrado foi a principal fonte de inclusão na classe eucaliptos. Além disso, a maior porcentagem das áreas classificadas como pastagem e gramíneas, pertenciam a classe de solo não cultivado.

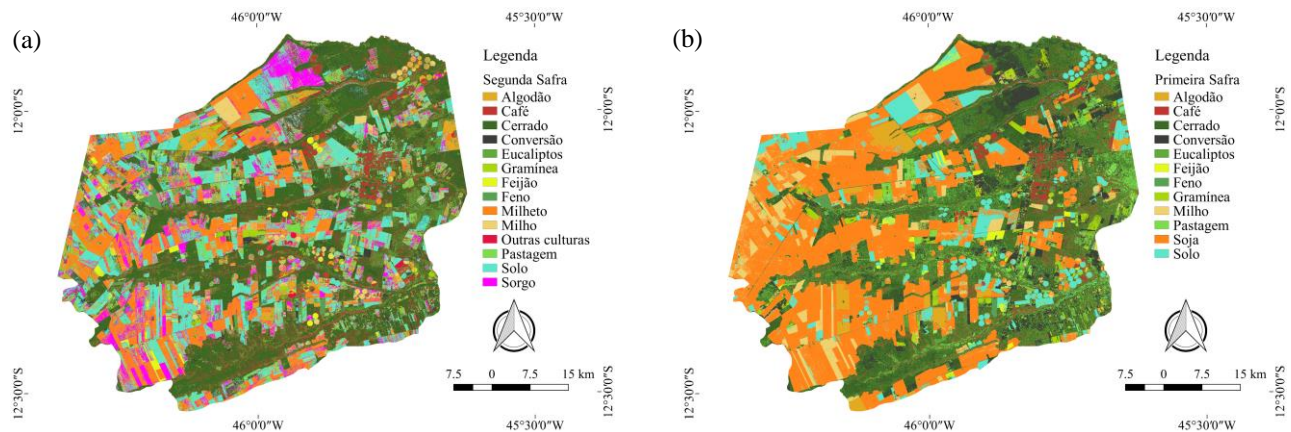


Figura 1. Mapas das classificações referentes à segunda safra de 2017 (a) e à primeira safra de 2018 (b).

Tabela 1. Matriz de confusão, AP e AU para a validação da classificação da segunda safra de 2017.

Classificação	Referência													AU	
	Algodão	Café	Cerrado	Conversão	Eucaliptos	Feijão	Feno	Gramínea	Milheto	Milho	Outras	Pastagem	Solo		Sorgo
Algodão	21771	0	0	0	0	107	306	0	1613	7722	0	2788	2443	107	0,591
Café	4	8860	36501	0	1123	1	0	1463	2	1966	0	337	3636	23	0,164
Cerrado	0	2939	164942	1022	3616	0	0	49022	2026	4680	0	255	18607	35	0,667
Conversão	0	2	12482	4098	863	0	54	1516	1187	608	0	618	14267	68	0,115
Eucaliptos	0	579	30311	0	333	0	0	253	0	0	0	652	427	11	0,010
Feijão	41	0	3	1	0	6106	1910	31	12967	15	5896	421	18164	185	0,133
Feno	376	387	0	0	0	46	10184	0	3	1848	1	3	3052	11	0,640
Gramínea	0	9	3473	742	5	59	3	10114	1433	1413	0	158	25972	150	0,232
Milheto	27	0	125	65	0	170	41	35	278785	2737	0	2339	79099	12290	0,742
Milho	8	0	15	0	0	295	1	0	14126	29323	0	385	22584	145	0,438
Outras	0	0	150	0	12	0	3263	25	14	558	52	737	1105	10	0,009
Pastagem	1	1	1636	25	4	0	22	323	6594	6584	0	3900	29788	3683	0,074
Solo	1	0	4862	4614	2	670	2478	1137	35307	2218	0	2875	365879	1607	0,868
Sorgo	0	0	1458	69	0	3472	5	185	23661	1124	0	395	78151	58952	0,352
AP	0,979	0,693	0,644	0,385	0,056	0,559	0,558	0,158	0,738	0,482	0,009	0,246	0,552	0,763	

Tabela 2. Matriz de confusão, AP e AU para a validação da classificação da primeira safra de 2018.

Classificação	Referência												AU
	Algodão	Café	Cerrado	Conversão	Eucaliptos	Feijão	Feno	Gramínea	Milho	Pastagem	Soja	Solo	
Algodão	47213	0	0	0	0	0	0	2377	0	0	2375	0	0,909
Café	0	11023	18661	0	381	22	882	232	33	47	22	117	0,351
Cerrado	0	1890	172669	2	213	0	313	414	884	2237	272	1517	0,957
Conversão	0	1	26643	4298	0	0	157	8843	214	10491	2357	9591	0,069
Eucaliptos	0	1016	16951	0	620	0	1110	120	166	68	561	408	0,029
Feijão	0	0	5	0	3	15045	0	792	2163	0	0	932	0,794
Feno	3	115	1199	0	97	0	11820	2149	10	785	19	217	0,720
Gramínea	0	0	106	313	0	0	63	23069	353	1654	3119	477	0,791
Milho	0	5	187	0	3	0	0	5153	130739	138	35021	809	0,760
Pastagem	0	323	20172	0	9	0	222	8408	1696	10676	1255	913	0,244
Soja	0	0	0	0	0	0	0	10	924	1	920859	4401	0,994
Solo	0	0	21	0	0	3	0	710	368	10	32606	130655	0,795
AP	1,000	0,767	0,673	0,932	0,468	0,998	0,811	0,441	0,950	0,409	0,922	0,871	

Em 2018, os maiores valores de AP (superiores a 90%) foram obtidos para as classes de algodão, feijão, milho, áreas de conversão e soja. As classes algodão e feijão tiveram erros de exclusão abaixo de 0,02%, logo praticamente todas as áreas de referências destas classes foram corretamente classificadas. Os erros de exclusão do milho foram devido a classificação deste como pastagem e feijão. Já para a área de conversão, o maior erro de exclusão ocorreu com a classe gramínea. Na soja, predominante nesta safra, os maiores erros de exclusão ocorreram com o milho e com o solo não cultivado.

Os maiores valores de AU (acima de 90%) para 2018 foram obtidos para a soja, cerrado e algodão. O solo não cultivado foi a classe com maior inclusão na classe soja. Pastagem, café e solo não cultivado foram as classes com maior inclusão na classe cerrado. Por sua vez, na classe de algodão, as maiores fontes de erro de inclusão vieram das classes gramínea e soja.

Os baixos valores de AP em 2018 foram registrados para as mesmas classes de 2017, com exceção da classe outras culturas que não ocorreu em 2018. A maior parte das amostras de pastagem foram classificadas como áreas de conversão. Neste ano, a classe

gramínea confundiu principalmente com as classes de áreas em conversão e de pastagem.

As classes de eucaliptos e áreas de conversão tiveram os menores valores de AU para 2018. A maior parte das áreas classificadas como eucaliptos e áreas de conversão, eram cerrado. Além disto, a classe cerrado também foi a principal fonte de inclusão na classe de pastagem.

Algodão em ambos os anos, soja, milho e feijão em 2018 obtiveram bons resultados, destacando o potencial dos dados SAR em identificar estas espécies agrícolas.

Em ambos os anos, as classes de cerrado e solo não cultivado foram confundidas com a maioria das demais classes. Para o solo não cultivado era esperada esta confusão, uma vez que essa classe está presente antes e depois nos talhões de todas as culturas agrícolas de ciclo anual.

A estrutura do cerrado, com vegetação mais esparça, favorece a confusão desta classe com solo não cultivado e pastagem nas regiões de campo e com as áreas de café e eucaliptos nas regiões arbustivas. Além disto, para a segunda safra de 2017, a menor acurácia reflete a maior confusão entre classes milho, sorgo e milheto, as quais são da família das gramíneas e apresentam estruturas semelhantes.

4. CONCLUSÕES

A partir de dados SAR foi possível classificar espécies agrícolas em ambas as safras, porém com melhores resultados para a primeira safra de 2018, onde houve menor número de classes, menor agrupamento de classes distintas e predomínio da classe soja. O algodão e a soja apresentaram as maiores acurácias, demonstrando a potencialidade do uso de dados SAR para identificar estas espécies agrícolas. Por outro lado, as classes eucaliptous e café (perenes) não tiveram bons resultados para esta abordagem, por possuírem estruturas semelhantes ao cerrado.

A utilização do ambiente virtual da *RUS machine* teve um bom desempenho, reduzindo o tempo de download dos dados e melhorando o processamento.

Como trabalhos futuros, sugerimos testar abordagens para melhorar a classificação das classes perenes. As abordagens sugeridas são uma série temporal maior de dados SAR, métricas como desvio padrão e amplitude da referida série e outros classificadores.

5. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ). Agradecimentos a Copernicus/ESA pelo acesso a RUS Machine.

6. REFERÊNCIAS

[1] Atzberger, C., "Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs," *Remote Sens.*, vol. 5, no. 2, pp. 949–981,

Feb. 2013.

[2] Eberhardt, I. D. R. *et al.*, "Cloud cover assessment for operational crop monitoring systems in tropical areas," *Remote Sens.*, vol. 8, no. 3, pp. 1–14, 2016.

[3] Moreira, A.; Prats, P.; Younis, M.; Krieger, G.; Hajnsek, I.; and Papathanassiou, K., "A Tutorial on Synthetic Aperture Radar," *IEEE Geosci. Remote Sens. Mag.*, no. march, pp. 1–43, 2013.

[4] Liu, C.; Shang, J.; Vachon, P. W.; and McNairn, H., "Multiyear Crop Monitoring Using Polarimetric RADARSAT-2 Data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2227–2240, Apr. 2013.

[5] Zhou, T.; Pan, J.; Zhang, P.; Wei, S.; and Han, T., "Mapping winter wheat with multi-temporal SAR and optical images in an urban agricultural region," *Sensors (Switzerland)*, vol. 17, no. 6, pp. 1–16, 2017.

[6] Kussul, N.; Lemoine, G.; Gallego, F. J.; Skakun, S. V.; Lavreniuk, M.; and Shelestov, A. Y., "Parcel-Based Crop Classification in Ukraine Using Landsat-8 Data and Sentinel-1A Data," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 6, pp. 2500–2508, 2016.

[7] Jhonnerie, R.; Siregar, V. P.; Nababan, B.; Prasetyo, L. B.; and Wouthuyzen, S., "Random Forest Classification for Mangrove Land Cover Mapping Using Landsat 5 TM and Alos Palsar Imageries," *Procedia Environ. Sci.*, vol. 24, pp. 215–221, 2015.

[8] Rodriguez-Galiano, V. F.; Ghimire, B.; Rogan, J.; Chica-Olmo, M.; and Rigol-Sanchez, J. P., "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, no. 1, pp. 93–104, Jan. 2012.

[9] Palazzo, F. *et al.*, "RUS: A New Expert Service for Sentinel Users," *Proceedings*, vol. 2, no. 7, p. 369, Mar. 2018.

[10] Sanches, I. D. *et al.*, "LEM benchmark database for tropical agricultural remote sensing application," in *ISPRS TC I Midterm Symposium Innovative Sensing - From Sensors to Methods and Applications*, 2018, pp. 1–6.

[11] ESA, E. S. A., "Sentinel-1 SAR User Guide Introduction," *Sentinel Online*, 2018. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar>.

[Accessed: 20-Jul-2018].

[12] Torres, R. *et al.*, "GMES Sentinel-1 mission," *Remote Sens. Environ.*, vol. 120, pp. 9–24, May 2012.

[13] ESA, E. S. A., "SAR Basics with the Sentinel-1 Toolbox," *Tutorials*, 2018. [Online]. Available: http://step.esa.int/docs/tutorials/S1TBX_SAR_Basics_Tutorial.pdf. [Accessed: 20-Aug-2018].

[14] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Viena, Austria, 2017.

[15] Gómez, M. G. C., "Joint use of Sentinel-1 and Sentinel-2 for land cover classification: A machine learning approach," Lund University, 2017.

[16] Zhang, X.; Wu, B.; Ponce-Campos, G.; Zhang, M.; Chang, S.; and Tian, F., "Mapping up-to-Date Paddy Rice Extent at 10 M Resolution in China through the Integration of Optical and Synthetic Aperture Radar Images," *Remote Sens.*, vol. 10, no. 8, p. 1200, Jul. 2018.

[17] Breiman, L., "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[18] Olofsson, P.; Foody, G. M.; Stehman, S. V.; and Woodcock, C. E., "Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation," *Remote Sens. Environ.*, vol. 129, pp. 122–131, Feb. 2013.