

User Guide for the MODIS Land Cover Type Product (MCD12Q1)

Last Updated: Aug 8, 2012

1. Introduction

Land cover plays a major role in the climate and biogeochemistry of the Earth system. An important use of global land cover data is the inference of parameters that influence biogeochemical and energy exchanges between the atmosphere and the land surface for use in models and other global change science applications. Examples of such parameters include leaf area index, roughness length, surface resistance to evapotranspiration, canopy greenness fraction, vegetation density, root distribution, and the fraction of photosynthetically-active radiation absorbed.

The MODIS Land Cover Type Product provides a suite of land cover types that support global change science by mapping global land cover using spectral and temporal information derived from MODIS. The objective of this document is to provide information related to the Collection 5 MODIS Land Cover Type Product (MCD12Q1). It is not designed to be a scientific document. Rather, it provides three main types of information:

1. An overview of the MCD12Q1 algorithm and product, along with references to published literature where more details can be found.
2. Guidance and information related to data access and data formats, to help users access and use the data.
3. Contact information for users with questions that cannot be addressed through information or websites provided in this document.

2. Overview of the MCD12Q1 Land Cover Type Product

The MODIS Land Cover Type Product is produced using a supervised classification algorithm that is estimated using a database of high quality land cover training sites. The training site database was developed using high-resolution imagery in conjunction with ancillary data (Muchoney et al., 1999). The site database is a "living" database that requires on-going augmentation and maintenance to improve the training data and detect mislabeled sites or sites that have changed over time.

MODIS data used in the classification include a full year of composited 8-day MODIS observations. Specific inputs include Normalized BRDF-Adjusted Reflectance (NBAR; Schaaf et al., 2002) and MODIS Land Surface Temperature (LST; Wan et al., 2002) data. These features are provided to the classifier as monthly composites and annual metrics (see Friedl et al., 2002; 2010).

The classification is produced using a decision tree classification algorithm (C4.5; Quinlan 1993) in conjunction with a technique for improving classification accuracies known as boosting (Freund 1995). Boosting improves classification accuracies by iteratively estimating a decision tree while systematically varying the training sample. At each iteration the training sample is modified to focus the classification algorithm on the most difficult examples. The boosted classifier's

prediction is then based upon an accuracy-weighted vote across the estimated classifiers. The implementation used here is Adaboost.M1 (Freund and Schapire, 1997), which is the simplest multi-class boosting method. Boosting has been shown to be a form of additive logistic regression (Friedman et al. 2000). As a result, probabilities of class membership can be obtained from boosting. These probabilities provide a means of assessing the confidence of the classification results as well as a means of incorporating ancillary information in the form of prior probabilities to improved discrimination of cover types that are difficult to separate in the spectral feature space.

Using this approach, the MODIS Land Cover Type algorithm ingests MODIS training data for all sites in the training database, estimates boosted decision trees based on those data, and then classifies the land cover at each MODIS land pixel. Following the classification a set of post-processing steps incorporate prior probability knowledge and adjust specific classes based on ancillary information. For more specific information and complete details related to the MODIS Land Cover Type algorithm, the reader is referred to the following key references:

- Friedl et al. (1997)
- Friedl et al. (1999)
- McIver and Friedl (2001)
- McIver and Friedl (2002)
- Friedl et al. (2002)
- Friedl et al. (2010)

Full citations to each of these papers are provided below.

3. Product Overview and Science Data Sets

The MODIS Land Cover Type Product supplies global maps of land cover at annual time steps and 500-m spatial resolution for 2001-present. The primary land cover scheme is provided by an IGBP land cover classification (Belward et al., 1999; Scean, 1999; Friedl et al., 2002; Friedl et al., 2010). For ease of use by the community, a number of other classification schemes are also provided, including the University of Maryland classification scheme (Hansen et al., 2000), the Biome classification scheme described by Running et al. (1994), the LAI/fPAR Biome scheme described by Myneni et al. (1997), and the plant functional type scheme described by Bonan et al. (2002). In addition, an assessment of the relative classification quality (scaled from 0-100) is provided at each pixel, along with quality assurance information and an embedded land/water mask.

The most recent version of the MODIS Land Cover Type Product is Collection 5.1, which includes adjustments for significant errors that were detected in Collection 5 of the MCD12Q1 product. This version is available on the Land Processes DAAC and is the recommended version for users. Essential information required for accessing and using these data include the following:

- Overview of data set characteristics (temporal coverage, spatial resolution, image size, data types, etc.).

- Science data sets included in the MODIS Land Cover Type Product, and their associated definitions.
- Information and specifications related to the MODIS Land Cover Type QA Science data set.

Up-to-date information related to each of these topics including science data sets, data formats, and quality information are available from the Land Processes DAAC at the following URL:

https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1

3.1. Data Formats and Projection

MODIS data are provided as tiles that are approximately 10° x 10° at the Equator using a sinusoidal grid in HDF4 file format. Information related to the MODIS sinusoidal projection and the HDF4 file format can be found at:

- MODIS tile grid: http://modis-land.gsfc.nasa.gov/MODLAND_grid.html
- MODIS HDF4: <http://www.hdfgroup.org/products/hdf4/>

3.2. Accessing and Acquiring Data

MCD12Q1 data can be acquired from the Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov/get_data). There are multiple portals for downloading the data. Reverb is the easiest to use and does not require a user account, but you only have the option to download the data in its original projection and HDF format. The MRTWeb portal enables more advanced options such as reprojection, subsetting, and reformatting but does require a user account.

4. Contact Information

Product PI: Mark Friedl (friedl@bu.edu)

Associate team member and contact for users: Damien Sulla-Menashe (dsm@bu.edu)

5. References Cited

1. Belward, A. S., Estes, J. E., & Kline, K. D. (1999). The IGBP-DIS Global 1-km Land-Cover Data Set DISCover: A Project Overview. *Photogrammetric Engineering and Remote Sensing*, 65, 1013-1020.
2. Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X. B., & Dai, Y. (2002). The land surface climatology of the community land model coupled to the NCAR community land model. *Journal of Climate*, 15, 3123-3149.
3. Freund, Y. (1995). Boosting a weak learning algorithm by majority. *Information and Computation*, 121(2), 256-285.
4. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 5(1), 119-139.

5. Friedl, M.A., & Brodley, C.E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399-409.
6. Friedl, M.A., Brodley, C.E., & Strahler, A.H. (1999). Maximizing land cover classification accuracies at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 969-977.
7. Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H., Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., & Schaaf, C. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83, 287-302.
8. Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168-182.
9. Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28(2), 337-374.
10. Hansen, M. C., DeFries, R. S., Townshend, J. R. G., & Sohlberg, R. (2000). Global land cover classification at the 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, 21, 1331-1364.
11. Muchoney, D., Strahler, A., Hodges, J., & LoCastro, J. (1999). The IGBP DISCover Confidence Sites and the System for Terrestrial Ecosystem Parameterization: Tools for Validating Global Land Cover Data. *Photogrammetric Engineering and Remote Sensing*, 65(9), 1061-1067.
12. McIver, D. K., & Friedl, M. A. (2001). Estimating pixel-scale land cover classification confidence using non-parametric machine learning methods. *IEEE Transactions on Geoscience and Remote Sensing*, 39(9), 1959-1968.
13. Mciver, D. K., & Friedl, M. A. (2002). Using prior probabilities in decision-tree classification of remotely sensed data. *Remote Sensing of Environment*, 81, 253-261.
14. Myneni, R. B., Nemani, R. R., & Running, S. W. (1997). Estimation of global leaf area index and absorbed PAR using radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*, 35, 1380-1393.
15. Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
16. Running, S. W., Loveland, T. R., & Pierce, L. L. (1994). A vegetation classification logic based on remote sensing for use in global scale biogeochemical models, *Ambio*, 23, 77-81.
17. Scepan, J. 1999. Thematic Validation of High-Resolution Global Land-Cover Data Sets, *Photogrammetric Engineering and Remote Sensing*, 65, 1051-1060.
18. Schaaf, C.B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., Strugnell, N. C., Zhang, X., Jin, Y., Muller, J. P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., d'Entremont, R. P., Hu, B., Liang, S., Privette, J. L., & Roy, D. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83, 135-148.

19. Wan, Z. M., Zhang, Y. L., Zhang, Q. C., and Li, Z. L. (2002). Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data. *Remote Sensing of Environment*, 83, 163-180.