Department of Interior

U.S. Geological Survey

**Algorithm Description Document**

**Continuous Change Detection and Classification (CCDC)**

**Land Change Monitoring Assessment and Projection (LCMAP)**

**Version 1.0**

**July 2015**



# Executive Summary

The Continuous Change Detection and Classification (CCDC) algorithm developed at the Center for Remote Sensing, Department of Geography and Environment, Boston University is under consideration as a primary application demonstrating the value of Land Change Monitoring Assessment and Projection. The U.S. Geological Survey Earth Resources Observation and Science Center has translated the original algorithm code into an opensource package. This document provides a high level overview of the code characteristics and functions, and will be updated as the CCDC algorithm becomes more clearly understood in the context of its performance as an opensource process.

# Document History

|  |  |  |
| --- | --- | --- |
| **Document Version** | **Publication Date** | **Change Description** |
| 0.1 | July 22, 2015 | First draft |
| 1.0 | July 29, 2015 | Final draft for USGS review |

Contents

[Executive Summary 2](#_Toc425958334)

[Document History 3](#_Toc425958335)

[Introduction 5](#_Toc425958336)

[Dependencies 5](#_Toc425958337)

[Inputs 5](#_Toc425958338)

[Outputs 6](#_Toc425958339)

[Prototype Code 6](#_Toc425958340)

[Verification Methods 7](#_Toc425958341)

[Maturity 7](#_Toc425958342)

[User Services 7](#_Toc425958343)

[Procedure 8](#_Toc425958344)

[References 10](#_Toc425958345)

[Acronyms 10](#_Toc425958346)

[Table 1 CCDC Product Maturity Matrix 7](#_Toc425958380)

[Figure 1 C-CCDC Work Flow 9](#_Toc425958393)

# Introduction

The Continuous Change Detection and Classification (CCDC) algorithm was developed at the Center for Remote Sensing, Department of Geography and Environment, Boston University as a robust methodology for identifying when and how a land surface pixel changes through time. It employs every clear pixel in a time series of Landsat data to model reflectance values such that they can be predicted and compared mathematically to observations to determine whether change has occurred at any given time. The algorithm further classifies the pixel to indicate what land cover types were observed before and after a change has been detected.

The original implementation of the CCDC code is written in Matrix Laboratory (MATLAB), which has since been translated into an opensource package as C code (C-CCDC). The C version is intended as the primary application used to test the development of a real-time system to support the U.S. Geological Survey (USGS) Land Change Monitoring Assessment and Projection (LCMAP) initiative.

This document provides a high level overview of the code characteristics and functions, and will be updated as the CCDC algorithm becomes more clearly understood in the context of its performance as an opensource process.

# Dependencies

Due to the necessity to identify change based solely on spectral responses, C-CCDC requires a certain level of data consistency. Processing is dependent on:

* All scenes projected to a uniform mapping grid
* Scenes spatially clipped to include the maximum coverage within a path/row time series
* Images formatted as Band Interleaved by Pixel (BIP)
* No cloud, cloud shadow, or snow pixels are present
* No Landsat data prior to 1984 (Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI) only)
* Significant storage and compute resources

# Inputs

The current C-CCDC acquires inputs from the USGS EROS Science Processing Architecture (ESPA), which provides the following.

* Landsat Level 2 Surface Reflectance for TM, ETM+, and OLI
* Landsat C version of Function of Mask (CFmask) for TM, ETM+, and OLI
* gzip (tar.gz) package data delivery
* Georeferenced Tagged Image File Format (GeoTIFF)
* Reprojection and spatial subsetting

# Outputs

The raw output from C-CCDC is a comma separated value (CSV) file for each pixel in the data stack, containing the coefficients used to model a time series and detect change. The variables include the following.

* Spectral band thresholds
* Observed and predicted values
* Dates of change
* Root Mean Square Error
* Coefficient for inter-annual change
* Coefficient for intra-annual change
* Coefficient for the overall spectral band value

In order to provide stakeholders with products adequate for the evaluation of C-CCDC, several annual maps can be generated, as itemized below.

* Land Cover
* Land Cover Confidence
* Land Cover Condition
* Timing of Land Cover Change
* Change Magnitude
* Procedure QA

# Prototype Code

The source CCDC algorithm software was delivered by its principal investigators from Boston University, Zhe Zhu and Curtis Woodcock. The MATLAB-based code and associated Wiki are advertised as available on the following Web sites.

[https://code.google.com/p/ccdc](https://code.google.com/p/ccdc/)

<https://github.com/prs021/ccdc>

The algorithm includes four (4) primary processes: preprocessing, time series modeling, change detection and land cover classification. The completed C versions of the first three functions are publicly available for review and comment on the following Web site.

<https://github.com/USGS-EROS/lcmap-change-detection>

The classification process is currently under development in cooperation with Boston University and will be posted to the same site as soon as its development and translation are complete.

# Verification Methods

The MATLAB code and its results have been published successfully by the principal investigators. Their methodologies are described in the references listed in this document.

The USGS C code was verified line by line against its MATLAB source to ensure process translations produced the same results as the original code. C-CCDC was also peer-reviewed by Landsat and EROS Science developers, and consequent revision is included in the current version of code. No scientific validation of the C-CCDC has yet been performed.

# Maturity

In the context of the USGS adaptation of the National Oceanic and Atmospheric Administration (NOAA) Climate Data Record (CDR) Maturity Matrix (<http://www1.ncdc.noaa.gov/pub/data/sds/cdr/Guidelines/Maturity_Matrix_Template.xlsx>), the C-CCDC falls within the early research stage. Its maturity definition is listed below, and is described without regard to the potentially higher maturity levels in the heritage MATLAB CCDC code.

|  |  |  |
| --- | --- | --- |
| **Criteria** | **Level** | **Definition** |
| Software Readiness | 3 | Moderate code changes expected. |
| Metadata | 1 | Little or none. |
| Documentation | 1 | Draft Algorithm Description Document (ADD) in progress. |
| Product Validation | 2 | Minimal. |
| Public Access | 2 | Limited data availability to develop familiarity. |
| Utility | 2 | Limited or ongoing. |

Table CCDC Product Maturity Matrix

# User Services

All higher level Landsat products, services and interfaces are supported by User Services staff at USGS EROS. Any questions, comments, or concerns are welcomed through the Landsat “Contact Us” on-line correspondence form. Please indicate “Surface Reflectance Data Request” as the topic of regard. Electronic mail can also be sent to the customer service address included below, with the same indication of topic.

USGS User Services

<http://landsat.usgs.gov/contactus.php>

[mailto:custserv@usgs.gov?subject=Surface Reflectance Data](mailto:custserv@usgs.gov?subject=Surface%20Reflectance%20Data)

User support is available Monday through Friday from 8:00 a.m. – 4:00 p.m. Central Time. Inquiries received outside of these hours will be addressed during the next business day.

# Procedure

This section will be updated as CCDC algorithm processes become more clearly understood in the context of their performance as C functions.

The general workflow of the C-CCDC begins with pre-processing. USGS EarthExplorer (<https://earthexplorer.usgs.gov>) is used to generate a list of scenes in the path/row and temporal range of interest. The scene list is submitted to the ESPA on-demand interface (<https://espa.cr.usgs.gov>) to order processing of C-CCDC input data. The delivered files are unzipped and untarred, and then fed into a format converter to transform them into BIP files.

A script reads through the CFmask product delivered from ESPA to determine whether a scene consists of at least 20% clear pixels, and if not, all input files for that scene are discarded.

The remaining files are tested a second time for clouds, cloud shadows, and snow using a robust iteratively reweighted least squares (RIRLS) model, which generates time series plots of the green and shortwave infrared bands for the pixels CFmask determined are clear. Model results are compared to observations to remove outliers CFmask may have missed.

The next step is the time series modeling, which is driven by least absolute shrinkage and selection operator (LASSO) regression to enhance distinction between inter-annual, intra-annual, and abrupt change.

LASSO is also used in the change detection portion of the C-CCDC algorithm code. The software functions by plotting pixel values per spectral band and fit them as sines and cosines to determine trend trajectories. Root mean square error (RMSE) thresholds identify deviations between modeled and observed values, and if a significant number of consecutive deviant values is found, a change is detected.

The final classification model ingests the coefficients generated during time series and change detection processes to compare with spectral information from reference land cover data with a random forest classifier (RFC) decision tree.

Figure 1 illustrates some major steps executed by the C-CCDC code, and is embedded in the following step by step narrative.

1. Only Landsat scenes with > 20% clear-sky pixels are used in CCDC analysis.
2. Only run CCDC for pixels where more than 50% of image have valid data.
3. If > 75% of valid data pixels are snow, then the pixel is considered permanent snow.
4. For permanent snow pixels, get model coefficients and RMSEs either from LASSO regression (if > 12 valid data points) or from median values (if < 12 points).
5. At least 12 data points and 1 year of data required for clear-sky land/water analysis.

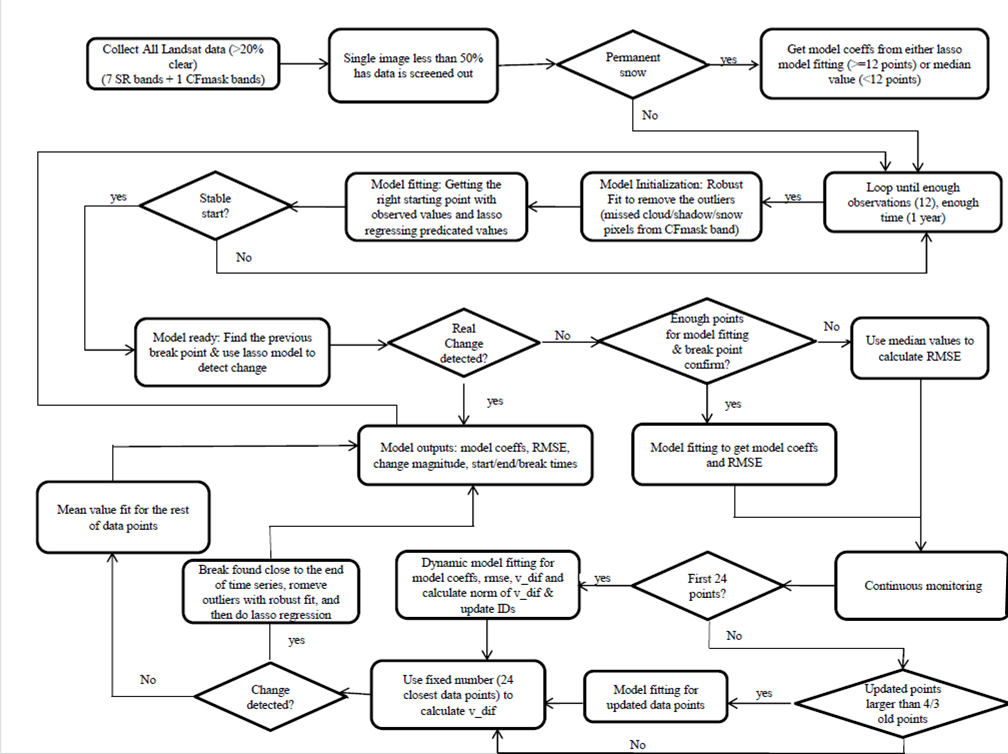


Figure C-CCDC Work Flow

1. Model initialization: Further remove outliers (cloud/cloud shadow/snow) using Robust Fit regression.
2. Model fitting: Do LASSO regression and get the right starting point with observed values and LASSO regressing predicted values.
3. Stable start: 12 data points and 1 years of data? If no, increases data points, otherwise, next step.
4. Model ready: Find the previous break point & use LASSO model to detect change if there are enough points (6) or use median values if less than (6) points.
5. Real change detected? (minimum magnitude of change is larger than a threshold and the first point of change is less than another threshold).
6. If yes, then model coefficients, RMSEs, magnitude of change are generated through LASSO regression.
7. If no, the falsely detected point is removed.
8. Check to see if enough points for model fitting and break point confirmation? If yes, do LASSO regression, otherwise, use median values to calculate RMSEs.
9. Continuous monitoring: If first 24 points, dynamic model fitting for model coefficients, RMSE, v\_dif and calculate norm of v\_dif & update IDs, otherwise, next step.
10. Update data points as needed to use the closest 24 data points for model fitting.
11. Detected change based on RMSEs and normalized value v\_dif
12. If break happens close to the end of time series, Robust Fit outlier removing process is first applied, then if the data points after the change detected (break point) are less than (6) points, median values are used, otherwise, LASSO model fitting is used.
13. Model outputs: model coefficients, RMSEs, change of magnitude, times of start, end, break, change probability, change category.

# References

Zhu, Z. & C.E. Woodcock, Continuous change detection and classification of land cover using all available Landsat data, *Remote Sensing of Environment*, 144, 152-171, 2014, [doi:10.1016/j.rse.2014.01.011](http://dx.doi.org/10.1016/j.rse.2014.01.011)

Zhu, Z., C.E. Woodcock, & P. Olofsson, Continuous monitoring of forest disturbance using all available Landsat imagery, *Remote Sensing of Environment*, 122, 75-91, 2012, <doi:10.1016/j.rse.2011.10.030>

# Acronyms

|  |  |
| --- | --- |
| **Acronym** | **Description** |
| ADD | Algorithm Description Document |
| BIP | Band Interleaved by Pixel |
| CCDC | Continuous Change Detection and Classification |
| CDR | Climate Data Record |
| CFmask | C version of Function of Mask |
| CSV | Comma Separated Values |
| DOI | Department of Interior |
| EROS | Earth Resources Observation and Science Center |
| ESPA | EROS Science Processing Architecture |
| ETM+ | Enhanced Thematic Mapper Plus |
| GEOTiff | Georeferenced Tagged Image File Format |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| LCMAP | Land Change Monitoring Assessment and Projection |
| MATLAB | Matrix Laboratory |
| NOAA | National Oceanic and Atmospheric Administration |
| OLI | Operational Land Imager |
| RFC | Random Forest Classifier |
| RIRLS | Robust Iteratively Reweighted Least Squares |
| RMSE | Root mean square error |
| TM | Thematic Mapper |
| USGS | U.S. Geological Survey |