As of 10 June 2016

This GOES-R Level 2+ space weather Algorithm Theoretical Basis Document (ATBD) is preliminary and subject to change as the GOES-R Program prepares for a launch in late 2016.

The NOAA National Centers for Environmental Information (NCEI) is currently developing a demonstration version of the Satellite Product Analysis and Distribution Enterprise System (SPADES) which will host the L2+ algorithms. The operational SPADES will be instantiated by the National Weather Service in FY2017 to serve the needs of the Space Weather Prediction Center (SWPC).

Updated versions of the L2+ ATBDs will periodically be made available and posted to the SPADES website.

Bill Denig
NCEI

http://www.ngdc.noaa.gov/stp/space-weather/online-publications/stp_sii/spades/

NOAA NESDIS CENTER for SATELLITE APPLICATIONS and RESEARCH

SUVI THEMATIC MAP ALGORITHM THEORETICAL BASIS DOCUMENT
Version 1.0

Hardcopy Uncontrolled

TITLE: SUVI THEMATIC MAP ALGORITHM THEORETICAL BASIS DOCUMENT VERSION 1.0

AUTHORS:
Dr. E. Joshua Rigler
Dr. Steven Hill

APPROVAL SIGNATURES:
_____<Actual Signature Date>
<Name of Technical Approver> Date
Technical Approver
_____<Actual Signature Date>
<Name of Agency Approver> Date
Agency

SUVI THEMATIC MAP ALGORITHM THEORETICAL BASIS DOCUMENT
VERSION HISTORY SUMMARY

Version	**Description**	**Revised Sections**	**Date**
1.0	Created by Dr. E. Joshua Rigler CIRES/SWPC	New Document	July 21, 2011

LIST OF FIGURES 6
LIST OF TABLES 7
LIST OF ACRONYMS 8
ABSTRACT 9
1 INTRODUCTION 10
1.1 Purpose of This Document 10
1.2 Who Should Use This Document 10
1.3 Inside Each Section 11
1.4 Related Documents 11
1.5 Revision History 12
2 OBSERVING SYSTEM OVERVIEW 12
2.1 Product Generated 12
2.2 Instrument Characteristics 13
3 ALGORITHM DESCRIPTION 15
3.1 Algorithm Overview 15
3.2 Processing Outline 16
3.3 Algorithm Input 18
3.3.1 Primary Sensor Data 18
3.3.2 Ancillary Data 19
3.4 Theoretical Description 20
3.4.1 Physics of the Problem 21
3.4.2 Mathematical Description 21
3.4.3 Algorithm Output 27
4 TEST DATA SETS AND OUTPUTS 29
4.1 Simulated/Proxy Input Data Sets 29
4.2 Output from Simulated/Proxy Inputs Data Sets 35
4.2.1 Precisions and Accuracy Estimates 38
4.2.2 Error Budget 42
5 PRACTICAL CONSIDERATIONS 42
5.1 Numerical Computation Considerations 42
5.2 Programming and Procedural Considerations 43
5.3 Quality Assessment and Diagnostics 43
5.4 Exception Handling 44
5.5 Algorithm Validation 45
6 ASSUMPTIONS AND LIMITATIONS 45
6.1 Performance 45
6.2 Assumed Sensor Performance 45
6.3 Possible Product Improvements 46
6.3.1 Covariance Matrix Check 46

6.3.2 Probability Thresholds 46
6.3.3 Improved Exception Handling 47
7 REFERENCES 48

LIST OF FIGURES

Figure 1 SUVI thematic map 12
Figure 2 SUVI proxy data 14
Figure 3 Flow diagram of SUVI thematic map algorithm 17
Figure 4 Pixel neighborhood system 26
Figure 5 Sample of SDO/AIA SUVI proxy “truth” images 30
Figure 6 Simulated SUVI long exposures 31
Figure 7 Simulated SUVI short exposures 32
Figure 8 Training pixel selection 34
Figure 9 Thematic maps (ML & MAP) for “truth” images 35
Figure 10 Thematic maps (ML & MAP) for simulated long exposures. 36
Figure 11 Thematic maps (ML & MAP) for simulated short exposures. 37

LIST OF TABLES

Table 1 SUVI image spectral channels 13
Table 2. SUVI L2 inputs to thematic map algorithm 18
Table 3. Ancillary inputs to SUVI thematic map algorithm 19
Table 4. Level 2 outputs of SUVI thematic map algorithm. 28
Table 5. Confusion matrix for ML solution derived from “truth” images (κ = 0.961) 39
Table 6. Confusion matrix for MAP solution derived from “truth” images (κ = 0.962) 39
Table 7. Confusion matrix for ML solution derived from long exposures (κ = 0.962) 40
Table 8. Confusion matrix for MAP solution derived from long exposures (κ = 0.961) 40
Table 9. Confusion matrix for ML solution derived from short exposures (κ = 0.950) 41
Table 10. Confusion matrix for MAP solution derived from short exposures (κ = 0.955) 41

LIST OF ACRONYMS

AIA	Atmospheric Imaging Array
AIT	Algorithm Integration Team
AT	Application Team
ATBD	Algorithm Theoretical Basis Document
AU	Astronomical Unit (~150,000,000 km)
AWG	Algorithm Working Group
CCD	Charge Coupled Device
CDR	Critical Design Review
CIRES	Cooperative Institute for Research in Environmental Sciences
CME	Coronal Mass Ejection
EIT	EUV Imaging Telescope
EUV	Extreme UltraViolet
GOES-R	Geostationary Orbiting Environmental Satellite – R series
HDR	High Dynamic Range
ICM	Iterated Conditional Modes
L1B	Level 1B (data)
L2	Level 2 (data)
MAP	Maximum *A priori* Probability
ML	Maximum Likelihood
MRD	Mission Requirements Document
NCEP	National Center for Environmental Prediction
NOAA	National Oceanic and Atmospheric Administration
NWS	National Weather Service
POV	Point Of View
SDO	Solar Dynamics Observatory
SEP	Solar Energetic Particle
SOHO	Solar and Heliospheric Observatory
STAR	Center for Satellite Apps and Research
SUVI	Solar Ultra-Violet Imager
SWPC	Space Weather Prediction Center
SWx	Space Weather
SXI	Solar X-ray Imager
TBD	To Be Determined
TBR	To Be Reviewed
XRS	X-ray Sensor

ABSTRACT

The GOES-R Solar Ultraviolet Imager (SUVI) will provide high-cadence full-disk images of the solar atmosphere in six narrow-band extreme ultraviolet (EUV) spectral channels that will support Space Weather forecast activities at NOAA's Space Weather Prediction Center (SWPC).

This document provides a comprehensive description of the SUVI Thematic Map algorithm. A SUVI Thematic Map is a 2D array of integer-valued pixel labels. Each integer value corresponds to one of a finite set of known solar features, or classes (e.g., coronal hole, solar flare, etc.) that can be uniquely identified using one or more nearly simultaneous narrow-band SUVI spectral channels. The 2D integer array may be displayed by itself, providing a simplified view of the solar disk that emphasizes solar features relevant to space weather forecasting, or it may serve as input for higher level solar image product algorithms like coronal hole boundary maps or flare location reports.

The algorithm chosen for SUVI thematic maps is a supervised multichannel Bayesian pixel classifier. In short, class-dependent multichannel pixel value means and covariances are determined from training data selected offline by solar image experts. These statistical parameters are then combined with actual measurements online to generate data-conditioned probabilities that each pixel belongs to one of the predefined classes. The class with the highest probability, or maximum likelihood (ML), is assigned to each pixel. Additionally, prior probabilities that favor a degree of smoothness in the SUVI thematic map are generated from the ML solution, then combined with the conditional probabilities according to Bayes' theorem to generate a maximum *a posteriori* probability (MAP) solution to reduce pixel classification noise.

Information necessary for developers and reviewers to verify that the algorithm meets operational requirements is presented. Also presented, when applicable, is traceability to heritage, and the design details necessary for development and implementation of the algorithm into operational use. Test and validation procedures are also provided along with assumptions and known limitations of the algorithm.

1 INTRODUCTION

The GOES-R Algorithm Working Group (AWG) Space Weather Application Team (SWx AT) is responsible for producing algorithms that generate products using space environment data from the GOES-R series satellites. These algorithms shall meet the operational needs of the NOAA/NWS/NCEP Space Weather Prediction Center (SWPC) to observe and forecast space weather conditions impacting near-earth systems such as satellites, communications, electrical power grids, manned space missions and many others. They shall also meet the requirements of outside operational and research agencies as agreed.

1.1 Purpose of This Document

This Algorithm Theoretical Basis Document (ATBD) details of the GOES-R Solar Ultraviolet Imager (SUVI) Thematic Map algorithm design and processing. The ATBD provides operational requirements for this product and defines how these requirements will be met using this algorithm. The algorithm inputs, processing, and outputs are described in enough detail to design, develop, test and implement the necessary processing software and storage mechanisms.

1.2 Who Should Use This Document

The members of the Space Weather Forecast Office and the Research and Customer Requirements Section of the SWPC shall use this ATBD to verify their operational requirements are being met by the proposed algorithm. They should also use it to understand the strengths and weaknesses of the algorithm as well as its applicability, accuracy, and robustness. The STAR Algorithm Integration Team (AIT) shall use this document to integrate the algorithm into their collaborative framework environment. It shall also be used by the prime development and implementation contractor to design, develop, test, validate and implement the algorithm into the final operational processing system.

1.3 Inside Each Section

Section 2.0 OBSERVING SYSTEM OVERVIEW:

- provides objectives of the SUVI thematic map algorithm including the output thematic maps and how they may be used operationally;
- discusses SUVI instrument characteristics and the fundamental algorithm approach;

Section 3.0 ALGORITHM DESCRIPTION:

- contains a complete SUVI thematic map algorithm description, including an outline of the processing, input data, and a theoretical description;
- provides estimates of the algorithm performance and output quality;
- reviews the numerical computation, programming and procedural issues and a description of how the algorithm has been validated;

Section 4.0 TEST DATA SETS AND OUTPUTS:

- describes the test data sets used to characterize the performance of the algorithm and quality of the data product(s),;
- discusses the results from algorithm processing on simulated input data;

Section 5.0 PRACTICAL CONSIDERATIONS:

- discusses issues involving numerical computation, programming and procedures, quality assessment and diagnostics and exception handling at a level of detail appropriate for the current algorithm maturity;

Section 6.0 ASSUMPTIONS AND LIMITATIONS:

- describes all assumptions concerning the SUVI Composite algorithm theoretical basis and performance;
- discusses planned product improvements for future enhancements;

Section 7.0 REFERENCES:

- Provides references to all sources cited in the ATBD.

1.4 Related Documents

GOESR SUVI thematic map Test Plan and Results
GOESR SUVI thematic map Implementation and User’s Guide

1.5 Revision History

Revision Number	Date	Author	Revision Description	Reason for Revision

2 OBSERVING SYSTEM OVERVIEW

2.1 Product Generated

A SUVI thematic map is a 2D array of integer-valued pixel labels. Each integer value corresponds to one of a predefined finite set of known solar features (e.g., coronal hole, solar flare, etc.), or classes that can be uniquely identified using one or more nearly simultaneous narrow-band SUVI spectral channels. The 2D integer array may be displayed by itself, providing a simplified view of the solar disk that emphasizes solar features relevant to space weather forecasting (see Figure 1), or it may serve as an input for higher level solar image product algorithms like coronal hole boundary maps or flare location reports.

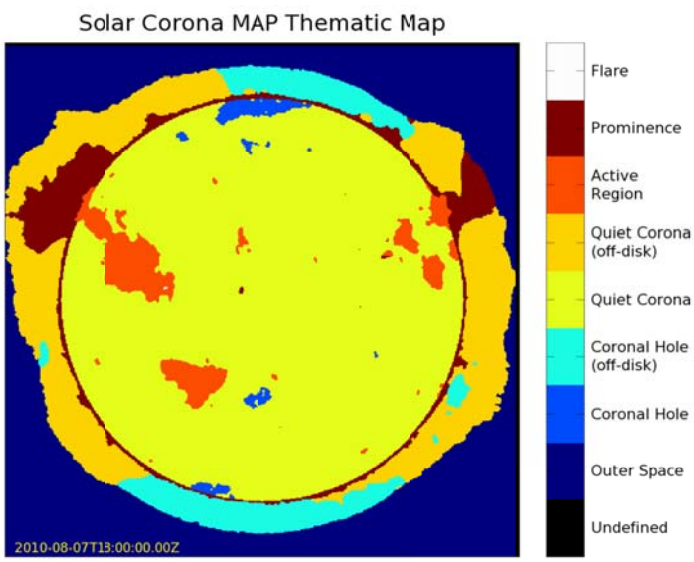


Figure 1 SUVI thematic map
Images from six EUV channels of SUVI proxy data were used to generate a thematic map of the sun during a small flare event on August 7, 2010.

2.2 Instrument Characteristics

The SUVI instrument operational requirements are detailed in section 3.4.2.4 of the GOES-R Series Mission Requirements Document (MRD) Version 3.9, dated November 2009. The SUVI thematic map algorithm assumes inputs that meet these minimum requirements, although in all likelihood, the delivered instrument performance will exceed these requirements. To start, individual SUVI spectral channel specifications are repeated in Table 1 for ease of reference, along with the type of EUV solar feature that motivated each spectral channel choice.

Table 1 SUVI image spectral channels

Channel	Wavelength	Sample Use
SUVI (Fe XVIII)	9.39 nm	Flares (~6x106 °K)
SUVI (Fe VIII)	13.1 nm	Flares (~10x106 °K) and Hot flares (~15x106 °K)
SUVI (Fe IX)	17.1 nm	Active regions (~6x105 °K)
SUVI (Fe XII)	19.5 nm	Active regions (~1x106 °K) and Hot Flares (~20x106 °K)
SUVI (Fe XV)	28.4 nm	Coronal holes (~2x106 °K)
SUVI (He II)	30.4 nm	Filaments (~6x104 °K)

The minimum spatial resolution of a SUVI image is 2.5 arcseconds, which when combined with a field-of-view requirement of 1.3 solar radii, translates to an image array of approximately 1000x1000 pixels. A measurement range of 0.3-10^6 photons/s/cm^2/arcsec^2 is also specified in the MRD, a dynamic range that is not achievable with a single exposure given SUVI's hardware limitations. Multi-exposure, high dynamic range composite images, a separate level 2 SUVI data product, will serve as the primary input to the SUVI thematic map algorithm.

SUVI's minimum refresh rate is one full dynamic range image every 2 minutes. Note that this does not call for all spectral channels to refresh every 2 minutes. If higher level data products (e.g., solar flare location report) require a thematic map input at a relatively high cadence, it is likely, that a single full dynamic range composite image will serve as input to more than one thematic map. The SUVI Thematic Map algorithm assumes all necessary input images are assembled and presented to it by a space weather data processing framework.

Finally, while SUVI has not yet flown aboard any NOAA satellite, the Solar Dynamics Observatory (SDO) was launched in the spring of 2010, and it carries the Atmospheric Imagine Array (AIA), a scientific instrument that has mostly compatible spectral channels to SUVI, and a very similar optical design and hardware selection. Details about the SUVI proxy data generated from SDO/AIA images will be presented later, but Figure 2 offers a preview of the kind of solar image data that will be returned by GOES-R SUVI, and that will serve as input for the SUVI thematic map algorithm.

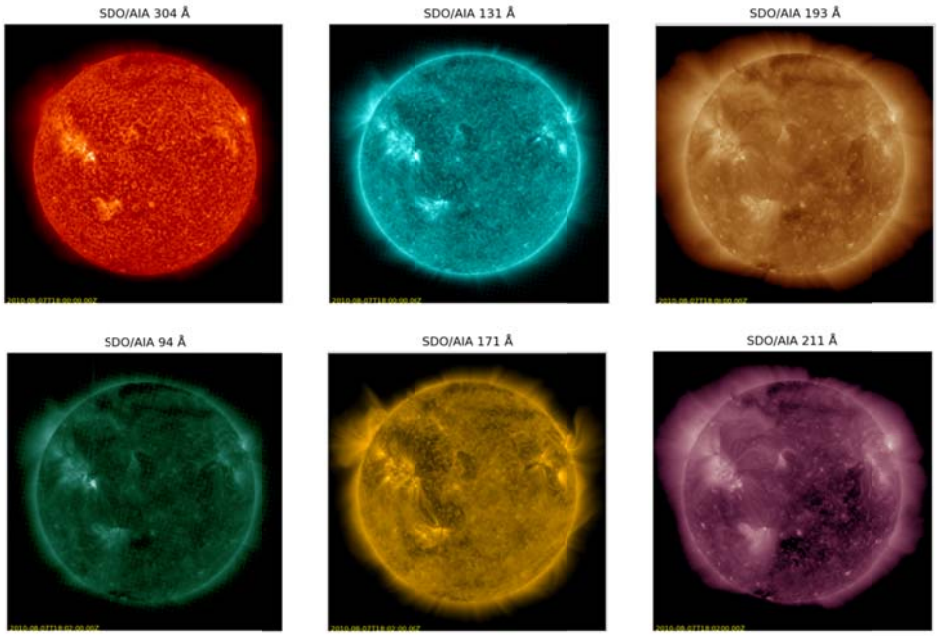


Figure 2 SUVI proxy data
Six select EUV channels from the Solar Dynamics Observatory (SDO) Atmospheric Imaging Array (AIA) are spatially and temporally binned/re-sampled to closely match SUVI's expected characteristics. The spectral band passes also nearly match those expected on SUVI, with the exception of SUVI's 28.4 nm channel, which we have chosen to replace with AIA's 21.1 nm channel for development and testing purposes.

3 ALGORITHM DESCRIPTION

3.1 Algorithm Overview

Class-dependent multichannel pixel means and covariances are determined from training data collected offline. These statistical parameters fully characterize a multivariate normal, or Gaussian, sample distribution of pixel values for each solar pixel class. The mathematics required to generate these input parameters are described in Section 3.4.2, although this is not technically part of the Thematic Map algorithm.

The multichannel pixel means and covariances are combined online with actual measurements to generate data-conditioned probabilities that each pixel belongs to one of the predefined set of classes. The integer corresponding to the class with the highest probability is then assigned to each pixel. Because we assume Gaussian distributions, the pixel with the highest probability is also known as the maximum likelihood (ML) solution.

A ML solution can be quite noisy when viewed in the context of an entire image. By invoking Bayes' theorem, we can incorporate prior probabilities that favor a degree of smoothness in the thematic map. These are initially generated from the whole-image ML solution, then combined with conditional probabilities to give a posterior probability that each pixel belongs to a given class. The class that generates the maximum *a posteriori* probability (MAP) is then assigned to the corresponding pixel.

As will be described in detail in subsection 3.4.2, the smoothness priors are themselves a function of a whole-image thematic map solution. This results in an algorithmic nonlinearity that requires an iterative solution where prior probabilities are recalculated from the MAP whole-image solution repeatedly until a stable solution is found.

3.2 Processing Outline

Steps required to construct a SUVI thematic map are:

1. Process algorithm configuration parameters and input images to construct multichannel image for subsequent classification:
    a. read parameters for generating conditional probabilities
    b. read parameters for generating prior probabilities
    c. Read image metadata, pixel values, pixel weights, and pixel flags for most recent single-channel image
    d. apply image selection criteria and/or replace bad pixels (**TBR**)
    e. save single-channel pixel array to appropriate slice of 3D multichannel pixel array
    f. repeat 1a-1e until 3D multichannel image is complete
2. Generate initial ML thematic map
    a. generate data-conditioned probabilities for each class (Eq. (3.9)) at each pixel, and store for subsequent processing
    b. initialize posterior probabilities with data-conditioned probabilities
    c. determine ML solution from posterior probabilities and use as initial MAP solution
3. Update MAP thematic map:
    a. calculate smoothness prior probability for each class at each MAP TM pixel using class ID of the pixel's nearest neighbors (Eq. (3.11))
    b. combine smoothness priors with class-conditioned probabilities to generate a new set of posterior probabilities (Eq. (3.2))
    c. repeat steps 3a-3b until the maximum number of iterations have been performed
4. Write out thematic map and associated metadata

The process flow is depicted graphically in Figure 3.

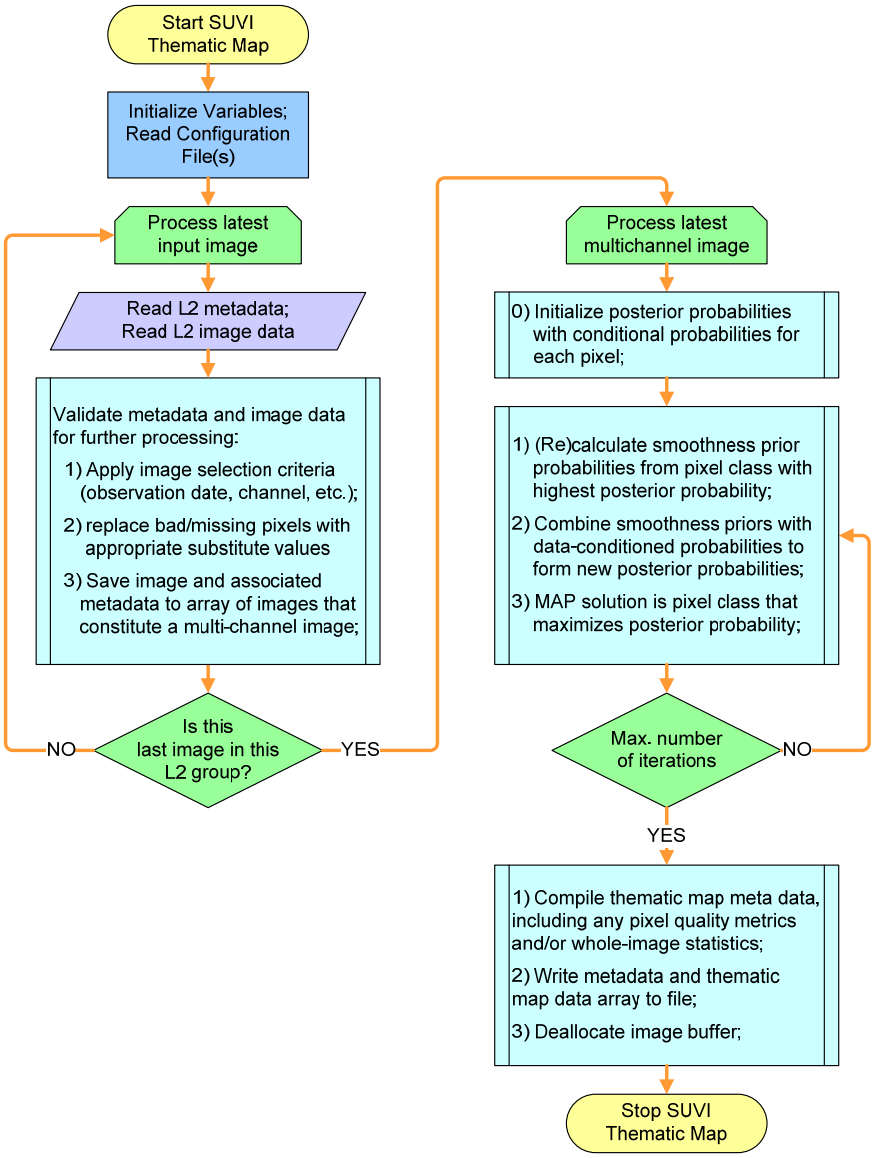


Figure 3 Flow diagram of SUVI thematic map algorithm
This flow diagram depicts at a high level steps required to generate a SUVI thematic map from SUVI composite images.

3.3 Algorithm Input

3.3.1 Primary Sensor Data

The SUVI thematic map algorithm requires as its primary input high dynamic range composite images, a separate level 2 SUVI data product, or something compatible at the metadata and pixel value levels. It is assumed that a Space Weather L2 data processing framework will make all necessary images available. A listing of the information expected in any input image is provided in Table 2.

Table 2. SUVI L2 inputs to thematic map algorithm

Name	Description	TYPE (Dimension)
Date/Time	Image date and time tags (e.g. YYYY,MM,DD,hh,mm,ss.sss)	CHAR(1) or Integer(7)
Channel/Filter	Identifier for channel/filter combo	CHAR (1)
Exposure Time	Accumulated exposure duration	REAL (1)
Pixels	Composite image pixel value	REAL (n_x, n_y)
Pixel Weights	Composite image pixel weights	REAL (n_x, n_y)
Pixel Flags	Composite image pixel flags	INTEGER (n_x, n_y)
DC_{pix}	X & Y pixel coordinates of solar disk center	REAL (2)
Δ_{pix}	Pseudo-angle describing a pixel's plate scale in the X & Y directions	REAL (2)
P_0	Position angle between solar north and image Y-axis	REAL (1)
L_0	Carrington longitude of Earth	REAL (1)
ϕ_S	Stonyhurst longitude of SUVI (0 if SUVI and Earth are co-located)	REAL (1)
B_0	Heliographic latitude of SUVI	REAL (1)
D_S	Distance from sun center to SUVI	REAL (1)

3.3.2 Ancillary Data

Ancillary data are assumed to be data not generated by SUVI or the GOES-R spacecraft. We include multichannel class-conditioned statistical parameters and user configurable inputs in this category. It may be desirable to ignore a channel or solar pixel class in the processing, but to demonstrate that this was deliberate, associated metadata should be retained. This is allowed via Boolean arrays that indicate which channel/class will be processed. These are all listed in Table 3.

Table 3. Ancillary inputs to SUVI thematic map algorithm

Name	Description	TYPE (Dimension)
Channel-Class Means	Multichannel mean vector for each pixel class	REAL (1, n_{chan}, n_{class})
Channel-Class Covariances	Multichannel covariance matrix for each pixel class	REAL (n_{chan}, n_{chan}, n_{class})
Channel Strings	Strings to associate with the column indices of channel-class arrays	CHAR(n_{chan})
Processed Channels	Flags to specify which channels should be processed	BOOLEAN(n_{chan})
Max Bad Pixels	Maximum number of bad pixels in a given channel before whole channel is automatically discarded.	INTEGER (1)
Class Strings	Strings to associate with integer-valued pixel classes	CHAR(n_{class})
Processed Classes	Flags to specify which classes should be processed	BOOLEAN(n_{class})
α	Class weights parameter vector	REAL (n_{class})
β	Smoothness prior parameter	REAL (1)
N_{iter}	Number of iterations to run ICM solver	INTEGER (1)

Optional “logical channels” may also be considered ancillary data. These may be non-SUVI solar images (e.g., hydrogen-alpha images from ground-based observatories), or even artificially generated pseudo-channels designed to apply some *a priori* knowledge of the solar scene (e.g., a mask of on-disk vs. off-disk pixels). These must be compatible with SUVI high dynamic range composite images in terms of the coordinates of the solar disk center and the solar disk radius, along with all appropriate metadata. It is assumed that the operational framework will make these images available to the thematic map algorithm as required.

3.4 Theoretical Description

Thematic maps derived from multispectral imagery have long been used to classify landcover types in Earth-directed satellite remote sensing. Assuming that ratios of reflected sunlight measured at a single pixel across multiple spectral channels are representative of a single “thin slab”, they tend to exhibit unique profiles for different dominant landcover types.

The classification of solar image pixels into scientifically and operationally useful categories has long been based on labor-intensive expert analysis (e.g., NOAA’s synoptic solar disk drawings), or theoretically-derived thresholds or bins in brightness (e.g., NOAA’s current solar flare location reports). More recently, solar imaging research has focused on statistically robust pixel classification techniques, but these have been mostly based on a single spectral channel.

There are good reasons for this, not the least of which is that, until recently, there has been neither the quantity nor quality of solar imagery as has existed in the Earth remote sensing world for many decades. Experts in solar image analysis have had the time and ability to manually tease information out of lower cadence, lower resolution, mono-spectral images because humans are inherently good at recognizing spatial and temporal patterns in images, albeit slow. They are not nearly so good at simultaneously considering multiple spectral channels.

Also, it is only recently that the technological ability to measure multiple spectral bands that exhibit meaningful differential variation in spectral intensity across different solar features of interest has matured to the point of operational utility. This is mostly because many solar spectral features of interest to space weather lie in the EUV and X-ray portions of the electromagnetic spectrum, which is difficult to design optics and electronic hardware to measure, and actually

impossible to measure from Earth's surface due to almost complete atmospheric absorption at these wavelengths. Extremely high-flying balloons and satellites are the only platforms on which such instrumentation would be able to function.

3.4.1 Physics of the Problem

In contrast to the reflectances measured by most Earth-pointing remote sensing instruments, the spectral bands that correspond to interesting solar features are usually emission lines falling in the EUV and soft X-ray portions of the solar electromagnetic spectrum. These lines are typically generated by fairly localized dynamical processes, which are often organized by height above the solar "surface". The thin slab assumptions noted for terrestrial landcover pixel classification may not be perfectly justified.

On the other hand, information in a magnetized plasma like the solar corona is transmitted more readily along field lines than via more isotropic fluid mechanical processes (i.e., compression waves), and field lines nearer to the sun are generally radial across a range of altitudes. A change to any particular emission line tends be quickly communicated to other altitudes at the same heliographic latitude and longitude. Therefore, thin slab assumptions may be reasonable for many events, especially if some sort of geometric correction is applied.

Using this logic, we justify applying techniques developed to analyze one regime (Earth's surface reflectance) to analyze and classify images in a very different regime (solar corona emission). Clearly there is inherent uncertainty to this approach, but it is no more than what is already associated with any analysis, manual or automated, of solar corona images.

3.4.2 Mathematical Description

Instrument and algorithmic noise, atmospheric distortion, and sub-pixel variations in actual landcover all contribute to a degree of uncertainty in the multichannel ratios described above. Therefore, there is an implied probability density function describing the likelihood that any given combination of single-channel pixel values corresponds to a particular thematic map pixel class. If this probability density function is known, or can be parameterized in some fashion from a limited training data set, it becomes possible to generate a probability that a given pixel belongs to a certain class. A sensible decision criterion is then to assign the pixel to the class with the highest probability.

Bayes' Theorem:

We choose to use Bayes' Theorem as a framework for calculating probabilities that a given multichannel pixel belongs to a particular solar pixel class. This is shown in its most general form via Equation (3.1).

$$P(w_j \mid \mathbf{x}_i) = \frac{P(\mathbf{x}_i \mid w_j)P(w_j)}{P(\mathbf{x}_i)} \tag{3.1}$$

In words, this states that the posterior probability that the ith multichannel pixel vector $\mathbf{x}_i$ belongs to the jth class w_j is equal to the data-conditioned probability of pixel vector $\mathbf{x}_i$ when class w_j is assumed, scaled by the ratio of a prior probability of class w_j to a prior probability of pixel vector $\mathbf{x}_i$.

The prior probability of pixel vector $\mathbf{x}_i$ can be difficult to determine if there are more than a few possible pixel values for any given channel, so it is common practice to just assume a uniform probability distribution. This implies that the posterior probability is proportional numerator of the RHS in Equation (3.1).

$$P(w_j \mid \mathbf{x}_i) \propto P(\mathbf{x}_i \mid w_j)P(w_j) \tag{3.2}$$

Finally, if there is no *a priori* knowledge of the probability that a pixel belongs to a certain class, the probability that a given multichannel pixel belongs to a class is simply proportional to the data-conditioned probability.

$$P(w_j \mid \mathbf{x}_i) \propto P(\mathbf{x}_i \mid w_j) \tag{3.3}$$

Calculating the Maximum Likelihood (ML) Solution:

Class-dependent mean vectors ($\boldsymbol{\mu}_j$) and covariance matrices ($\mathbf{C}_j$) can be readily calculated from carefully selected training data, and used to fully characterize the pixel value distributions if all pixel value distribution functions are assumed to be Gaussian. These statistical parameters will be provided as inputs to the

algorithm, but we demonstrate how they are constructed here to assist with explanations due later in this document.

Start by defining a multichannel pixel as a vector of pixel values, one for each channel being considered:

$$\mathbf{x} \equiv \begin{bmatrix} x_1 & x_2 & \cdots & x_{n_{\text{chan}}} \end{bmatrix}^{\text{T}} \tag{3.4}$$

Here "T" is the transpose operator, so $\mathbf{x}$ is a column vector with a length equal to the number of channels, n_{chan}. This orientation becomes important later. The first statistical moment of $\mathbf{x}$ is its mean vector.

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^{n_j} \mathbf{x}_{j,i}}{n_j} \tag{3.5}$$

Here, n_j is the number of multichannel pixels that have been assigned to the jth class in the training data set. The mean vector for each class will look much like the multichannel pixel vector:

$$\boldsymbol{\mu}_j \equiv \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_{n_{\text{chan}}} \end{bmatrix}_j^{\text{T}} \tag{3.6}$$

A covariance matrix is an N-dimensional generalization of the variance, or the 2nd statistical moment of a distribution of multichannel pixel values. Using training pixels assigned to a particular class, we get:

$$\mathbf{C}_j = \frac{\sum_{i=1}^{n_j} \left[\mathbf{x}_{j,i} - \boldsymbol{\mu}_j\right] \times \left[\mathbf{x}_{j,i} - \boldsymbol{\mu}_j\right]^{\text{T}}}{n_j} \tag{3.7}$$ [1]

[1] Equation (3.7) is a biased estimate of the sample covariance. The denominator is n_j-1 for an unbiased estimate, but if n_j is sufficiently large, the difference is negligible. Furthermore, this form allows means and covariances from different samples to be merged using standard mixture reduction (SMR) techniques.

Given that $\mathbf{x}_{j,i}$ and $\boldsymbol{\mu}_j$ are column vectors of length n_{chan}, matrix multiplication rules turn the term inside the summation of Equation 3.7 into a symmetric square matrix of dimension n_{chan}. The diagonal elements of $\mathbf{C}_j$ hold the class-dependent variance of each channel, while the off-diagonal elements of $\mathbf{C}_j$ hold the class-dependent covariance between each channel. A covariance matrix for the jth class will look like:

$$\mathbf{C}_j \equiv \begin{bmatrix} V_1 & C_{1,2} & \cdots & C_{1,n_{\text{chan}}} \\ C_{2,1} & V_2 & \cdots & C_{2,n_{\text{chan}}} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n_{\text{chan}},1} & C_{n_{\text{chan}},2} & \cdots & V_{n_{\text{chan}}} \end{bmatrix}_j \tag{3.8}$$

Equation (3.7) guarantees that variances will always be positive. However, a valid covariance matrix must also be "positive-definite". This is guaranteed if n_j is at least equal to n_{chan}, but to avoid issues related to machine precision, n_j should be much larger than n_{chan}. There are more and less efficient ways to test for positive definiteness, but in effect, they all check to see if all eigen values for $\mathbf{C}_j$ are positive. Invalid covariance matrices should never be used by the algorithm.

Given these definitions and constraints, and assuming that valid class-dependent mean vectors and covariance matrices were provided to the algorithm, the data-conditioned probability that the ith multichannel pixel belongs to the jth class is:

$$P\left(\mathbf{x}_i \mid w_j\right) = \frac{1}{\sqrt{2\pi}^{\rho}\sqrt{\left|\mathbf{C}_j\right|}} \exp\left(-\tfrac{1}{2} \times \left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)^{\mathsf{T}} \times \mathbf{C}_j^{-1} \times \left(\mathbf{x}_i - \boldsymbol{\mu}_j\right)\right) \tag{3.9}$$

Here, "ρ" is the number of spectral channels, "$\left|\mathbf{C}_j\right|$" is the scalar-valued determinant of the covariance matrix, "$\mathbf{C}_j^{-1}$" is the inverse of the covariance matrix (all other terms and operators have been previously defined). Expanding this to fully represent the necessary matrix operations, we get:

$$P\left(\mathbf{x}_i \mid w_j\right) = \frac{1}{\sqrt{2\pi}^{\rho}\sqrt{\left|\mathbf{C}_j\right|}} \exp\left(\begin{array}{l} -\frac{1}{2}\times \left[\left(x_{i,1}-\mu_{j,1}\right) \;\; \left(x_{i,2}-\mu_{j,2}\right) \;\; \cdots \;\; \left(x_{i,n_{\text{chan}}}-\mu_{j,n_{\text{chan}}}\right)\right]_i \times \\ \begin{bmatrix} V_1 & C_{1,2} & \cdots & C_{1,n_{\text{chan}}} \\ C_{2,1} & V_2 & \cdots & C_{2,n_{\text{chan}}} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n_{\text{chan}},1} & C_{n_{\text{chan}},2} & \cdots & V_{n_{\text{chan}}} \end{bmatrix}_j^{-1} \times \\ \left[\left(x_{i,1}-\mu_{j,1}\right) \;\; \left(x_{i,2}-\mu_{j,2}\right) \;\; \cdots \;\; \left(x_{i,n_{\text{chan}}}-\mu_{j,n_{\text{chan}}}\right)\right]_i^{\mathrm{T}} \end{array} \right) \tag{3.10}$$

Applying standard matrix multiplication rules, the large matrix operation inside the exponent reduces to a simple scalar value, so $P\left(\mathbf{x}_i \mid w_j\right)$ is just a scalar value that equals the probability that pixel *i* belongs to class *j*. The class with the highest probability is assigned to a thematic map pixel *i*, and is referred to as the maximum likelihood (ML) solution.

Smoothness Prior Probabilities:

ML thematic maps can be prone to noise in the form of spatially isolated pixels that are mistakenly classified as one solar pixel type, even though they are surrounded by a fairly homogenous field of another solar pixel type. While this is sometimes a perfectly legitimate phenomenon, more likely than not, solar pixels will be surrounded by pixels of a similar class.

To formalize this prior assumption of smoothness, we turn to Markov Random Field theory. A pixel *r* resides in a 2D “neighborhood” system. A neighborhood system can be ordered according to the Euclidian distance between *r* and its neighbors. A zeroth-order system includes only *r*. A first-order system includes those pixels directly above/below/left/right of *r*. A second-order system includes these and those pixels off of pixel *r*’s diagonals (see Figure 4). This could be continued *ad infinitum*, but a second-order system will suffice for us.

2	**1**	**2**
1	***r***	**1**
2	**1**	**2**

Figure 4 Pixel neighborhood system
The pixel r and its 2nd order neighborhood system in a discrete 2D coordinate system. The rank or the Euclidian distance from *r* defines the order of the system.

The probability that a pixel belongs to a particular class is a function of the number of neighboring pixels that belong to that class, and the prior probability of belonging to that pixel class.

$$P(w_j \mid N_r) = \frac{\exp(\alpha_{w_j} + \beta n_r)}{\sum_{j=1}^{n_w} \exp(\alpha_{w_j} + \beta n_r)} \tag{3.11}$$

where…

w_j – pixel class j of n_w
N_r – pixels in r's neighborhood
n_r – number of pixels in N_r belonging to class w_j
α_{wj} – relative weight for class w_j
β – smoothness parameter

The parameters α_{wj} and β are defined by the user, and should be determined in such a manner as to best meet the operational requirements of SWPC and the space weather community in general. Roughly speaking however, if α_{wj} is bigger than other weights, class w_j is more likely to be assigned to a given pixel. If β is bigger, spatially isolated pixels of any class are less likely.

Maximum A Posteriori Probability (MAP) Solution:

You will recall that Bayes' Theorem states that the posterior probability that a pixel belongs to a particular class is proportional to the product of the data-conditioned probability when assuming a given class, and the prior probability of that given class. The class corresponding to the highest posterior probability is then assigned to each pixel. This is referred to as the Maximum *A Posteriori* Probability (MAP) solution.

If smoothness priors are to be used to construct a MAP solution, it is clear that an iterative solver is necessary because each smoothness prior is itself a function of the class of its neighboring pixels, which may change once Equation (3.11) is applied to every pixel in an image. A variety of techniques could be applied, some faster, and some more accurate. A balance of speed and accuracy is obtained using the Iterated Conditional Modes (ICM) algorithm described in Section 3.2.

3.4.3 Algorithm Output

The SUVI thematic map algorithm generates a 2D array of integer-valued pixel classes. It also generates metadata necessary to register the image in space and time. These will, in all likelihood, simply be coordinates passed through from the most recent composite input image. An exhaustive list of channel and pixel class labels to assist in final interpretation of the product should also be provided. Also, because it may be desirable for the algorithm to function with one or more missing spectral channels, or with a particular class being ignored, Boolean arrays of lengths equal to the number of channels and classes respectively should be provided, and defined such that the channels and classes used are obvious. Finally, the parameters used to configure the ICM algorithm itself should be listed. All outputs are listed in Table 4.

Table 4. Level 2 outputs of SUVI thematic map algorithm.

Name	Description	TYPE (Dimension)
Date/Time	Date and time tags of latest composite image	CHAR(1) or Integer(7)
Channel-Class Version #	Version number of the channel-class input parameters	CHAR (1)
Channel Strings	Strings to associate with the column indices of channel-class arrays	CHAR (n_{chan})
Processed Channels	Flags indicating processed channels	BOOLEAN(n_{chan})
Max Bad Pixels	Maximum number of bad pixels in a given channel before whole channel is automatically discarded.	INTEGER (1)
Class Strings	Strings to associate with integer-valued pixel classes	CHAR (n_{class})
Processed Classes	Flags indicating processed classes	BOOLEAN(n_{class})
Pixels	Integer-valued pixel classes	INTEGER (n_x, n_y)
P_0	Position angle between solar north and image Y-axis	REAL (1)
L_0	Carrington heliographic longitude of Earth	REAL (1)
ϕ_S	Stonyhurst heliographic longitude of SUVI (0 if SUVI and Earth are co-located)	REAL (1)
B_0	Heliographic latitude of SUVI	REAL (1)
D_S	Distance from sun center to SUVI	REAL (1)
α	Class weights parameter vector	REAL (n_{class})
β	Smoothness prior parameter	REAL (1)
N_{iter}	Number of iterations to run ICM solver	INTEGER (1)

4 TEST DATA SETS AND OUTPUTS

4.1 Simulated/Proxy Input Data Sets

GOES satellites 12-15 carry a Solar X-ray Imager (SXI) designed to provide forecasters with a global picture of the solar corona using fairly broad spectral bands in the soft X-ray portion of the solar spectrum. The SXIs served, and continue to serve, this purpose well, but for a variety of reasons, the GOES-R program chose to switch to a narrower band, EUV imager that is now known as SUVI. SXI images cannot serve as an adequate proxy for SUVI for the purposes of testing and demonstrating the SUVI Thematic Map algorithm.

The Solar Dynamics Observatory (SDO) Atmospheric Imaging Array (AIA) was launched in the spring of 2010, and has a very similar design, both optical and electronic, to SUVI. The important differences are:

- SDO-AIA is comprised of four separate telescopes, compared to SUVI's single telescope, quadrupling the rate at which images can be acquired;
- each SDO-AIA primary mirror has ~4x the photon collecting area of the SUVI primary mirror for each spectral channel;
- the SDO-AIA CCD is 4096x4096 pixels, compared to the SUVI CCD's ~1000x1000 pixels;

Given SDO-AIA's higher cadence, sensitivity, and spatial resolution, one can bin and resample to time and spatial grids comparable to SUVI and reduce the noise attributable to Poissonian photon counting statistics by over a factor of 4[2]. We choose to consider these "truth" images, which we then systematically corrupt by applying random Poisson noise to simulate photon counting statistics. To do this properly, digital numbers (DN) provided in the SDO-AIA numbers must be converted to photons, which requires assumptions about the electronics gain (electrons per DN), physical gain (electrons per absorbed photon), and CCD quantum efficiency (% incident photons actually absorbed). These were not provided with the images presented here, so best guesses were made.

Figures 5, 6, and 7 show the "truth", simulated long exposures, and simulated short exposures used for all subsequent tests and validation.

[2] This rough estimate assumes that signal-to-noise (SNR) scales as $\sqrt{N}$, and that *N*, the number of collected photons, scales linearly with accumulated exposure time and photon collection area.

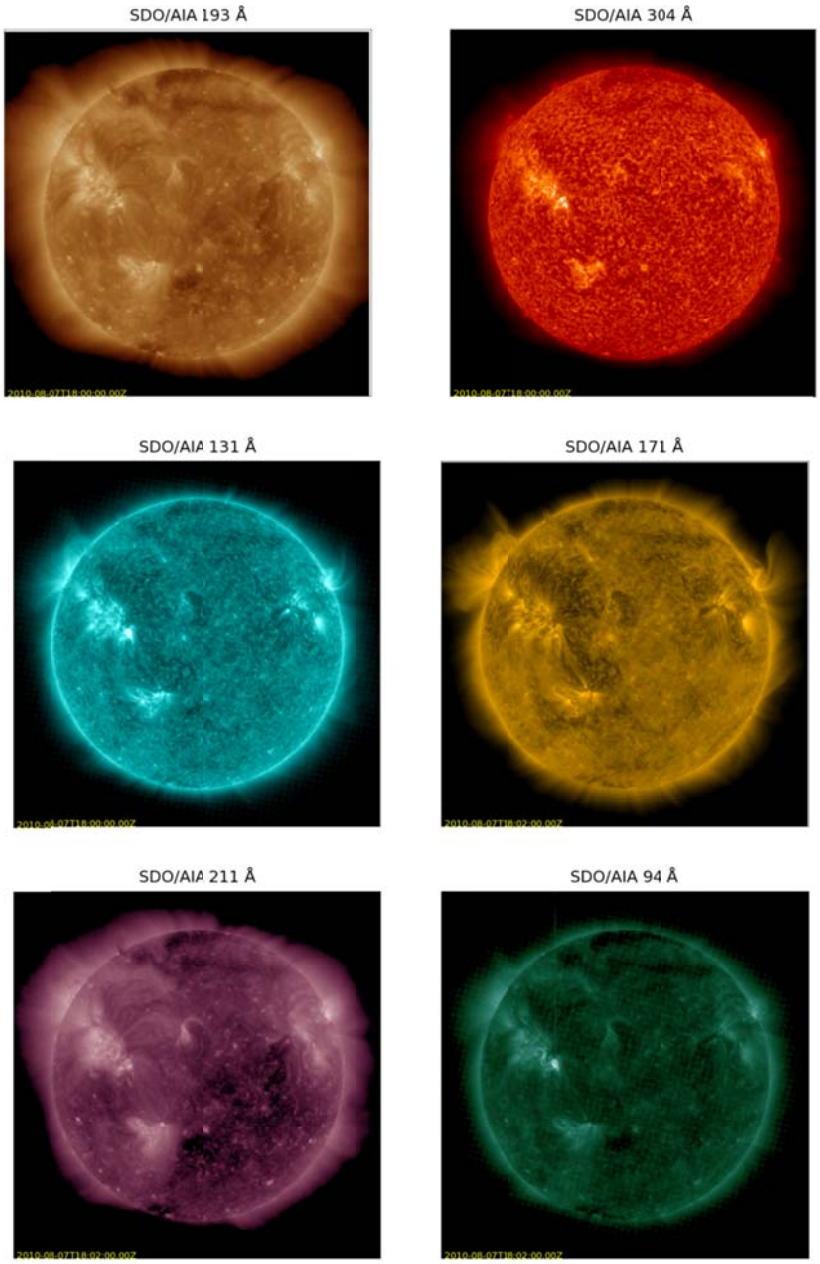


Figure 5 Sample of SDO/AIA SUVI proxy "truth" images
Six low-noise EUV solar images constructed from SDO-AIA data that nearly match SUVI's expected cadence, spatial resolution, and spectral channels.

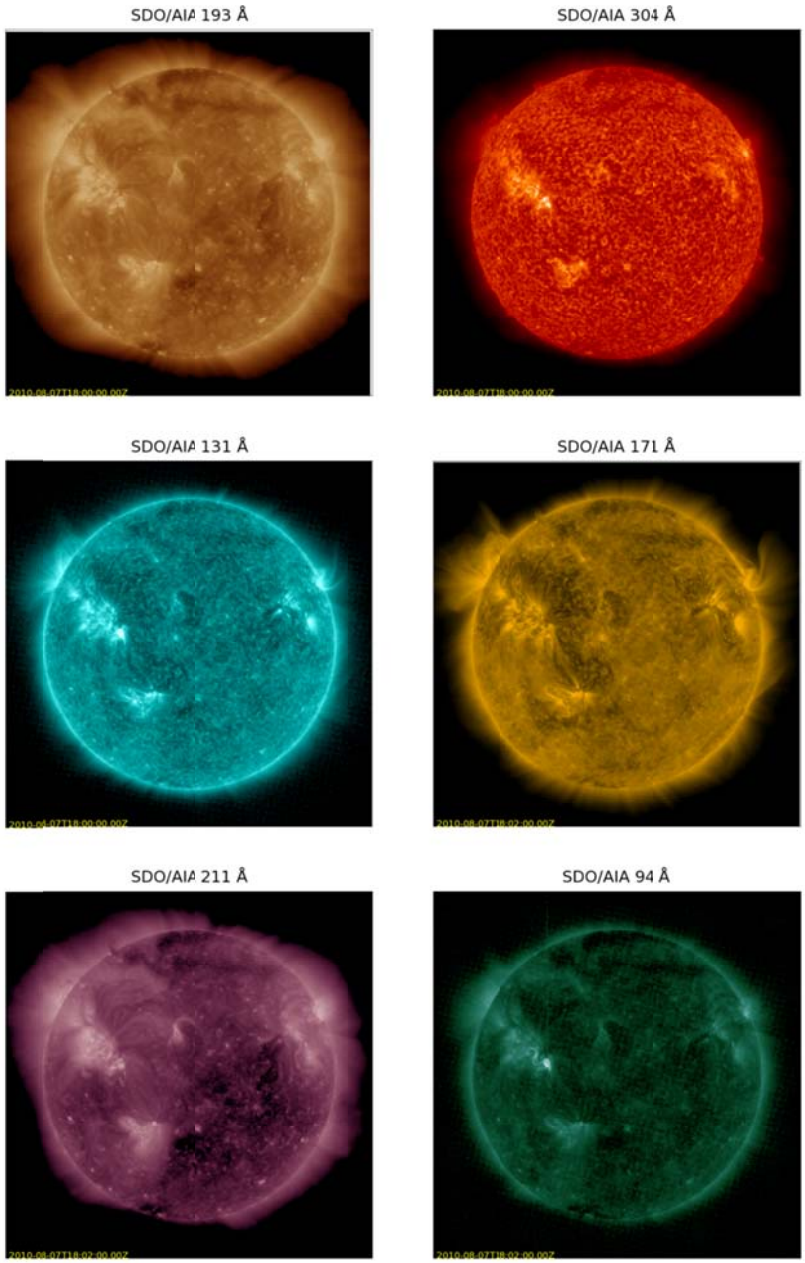


Figure 6 Simulated SUVI long exposures
These are the "truth" exposures with random Poisson noise added to simulate a 1 second exposure.

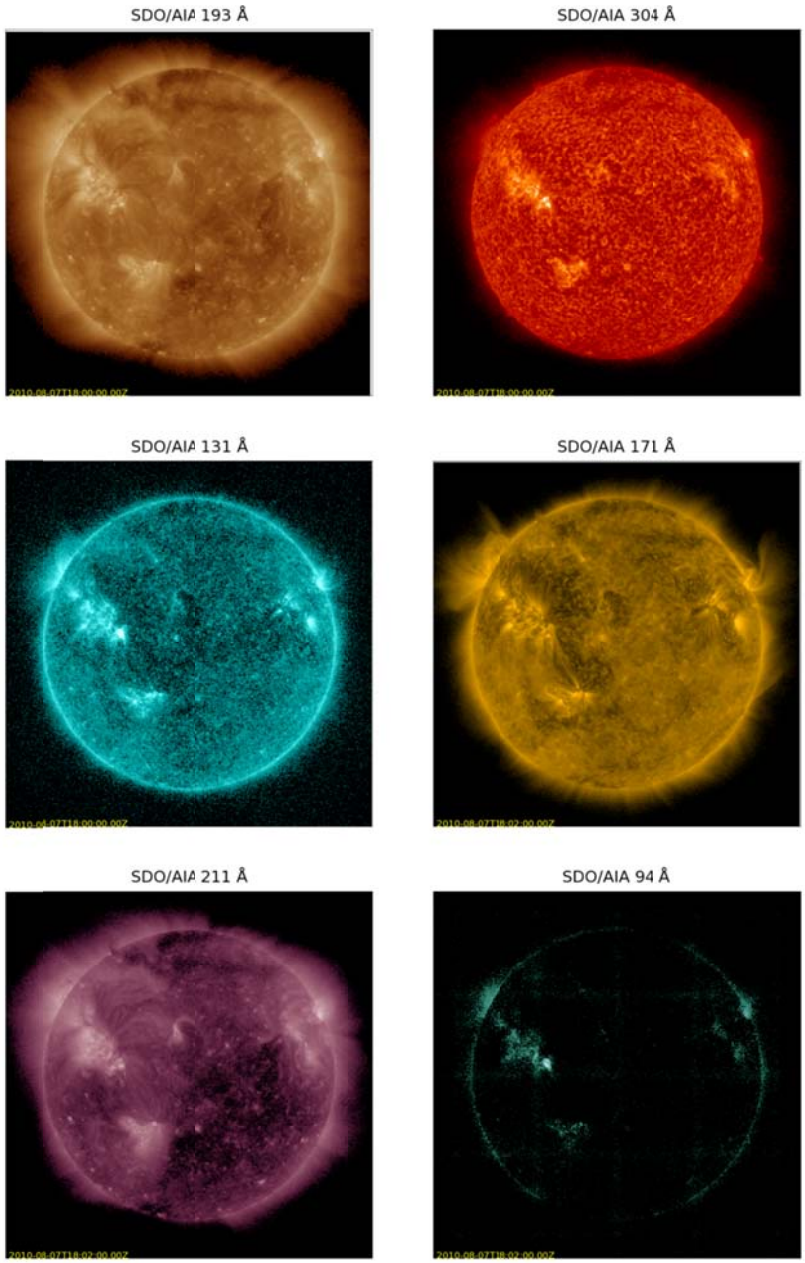
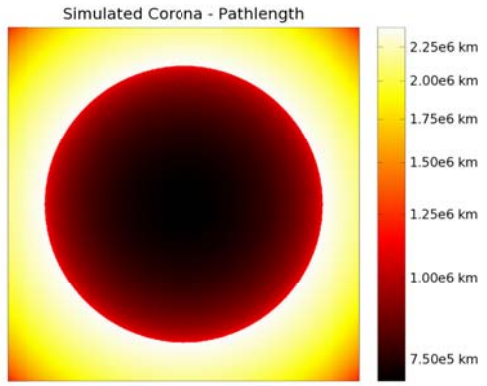


Figure 7 Simulated SUVI short exposures
These are the "truth" exposures with random Poisson noise added to simulate a .025 second exposure.

For the purpose of developing, testing, and demonstrating a prototype of the SUVI Thematic Map algorithm, a small M-class flare event was chosen that occurred on August 7, 2010. The images of proxy data shown in Figures 5, 6, and 7 correspond to the immediately pre-flare phase of this event, as determined from NOAA's so-called Edited Events List, available online from the Space Weather Prediction Center. In addition to these pre-flare scenes, a peak-flare, and a declining phase set of images were used to train and test the algorithm. These are not shown here.



In addition, a pseudo-channel was created that simulated the geometric effects of limb brightening. This phenomenon occurs in optically thin plasmas, and leads to a brightening near and immediately above the solar limb because the look vector of the observer passes through more photon-emitting solar atmosphere. The pseudo-channel used here is the path length in kilometers of the look vector passing through a solar atmosphere that was modeled as a spherical shell around the photosphere, with a thickness of 1 solar radii. This is by no means intended to perfectly resemble reality, and was provided merely as a demonstration of how pseudo-channels can help the SUVI Thematic Map algorithm generate better output.

Finally, an "expert" in solar image analysis used a semi-GUI manual pixel selection tool to assign one of a predefined set of 8 solar feature classes to subsets of pixel coordinates. All channels represented by the proxy inputs were examined, even though only a single channel at a single time-step is presented in Figure 8. These training data were assembled into vectors of multichannel pixels (Eq. (3.4)), and processed into class-dependent multi-channel means and covariances (Eqs. (3.5)). These were in turn combined with multichannel proxy images to generate a sequence of SUVI Thematic Maps for this event. A representative ML and MAP thematic map for each of the "truth", long exposure, and short exposure images is shown in the next subsection.

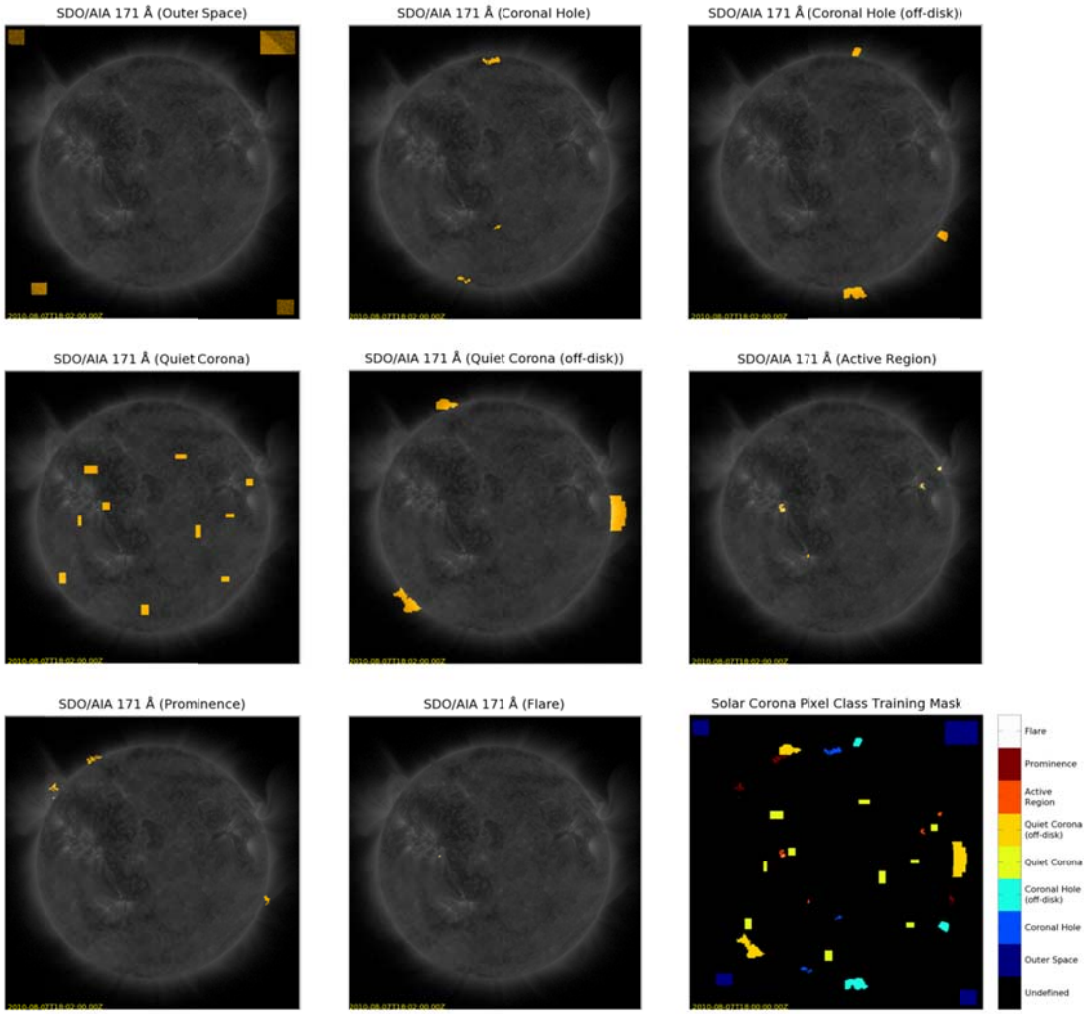


Figure 8 Training pixel selection
Colorized training pixel masks overlay a 171 A grayscale image (the same training pixels apply to all channels). A more compact representation of color-coded training pixels is also shown.

4.2 Output from Simulated/Proxy Inputs Data Sets

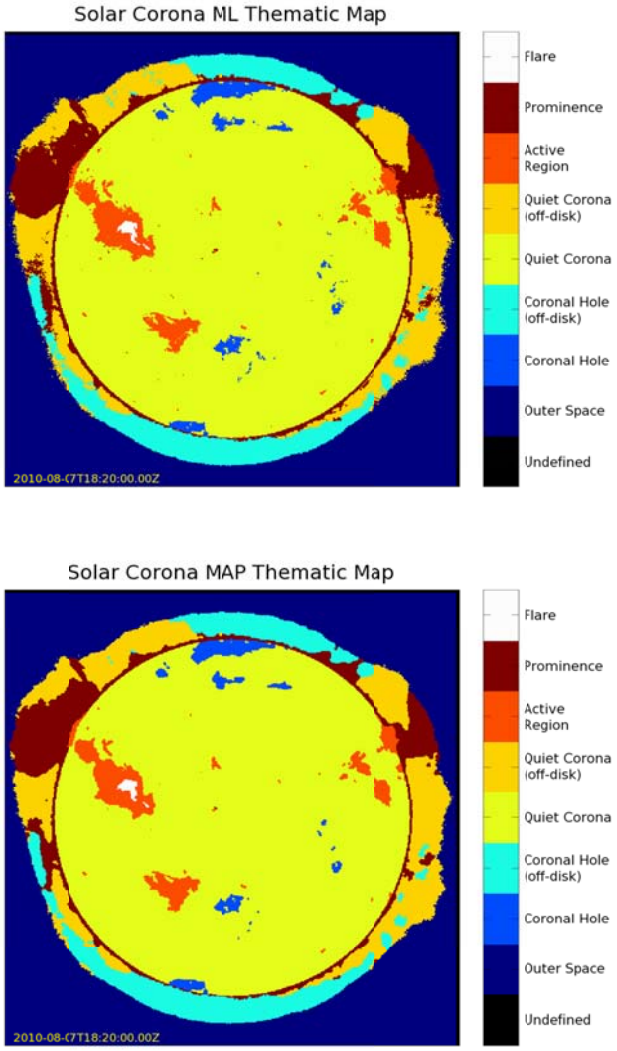


Figure 9 Thematic maps (ML & MAP) for "truth" images
Sample thematic maps generated from images taken at onset of M class solar flare.

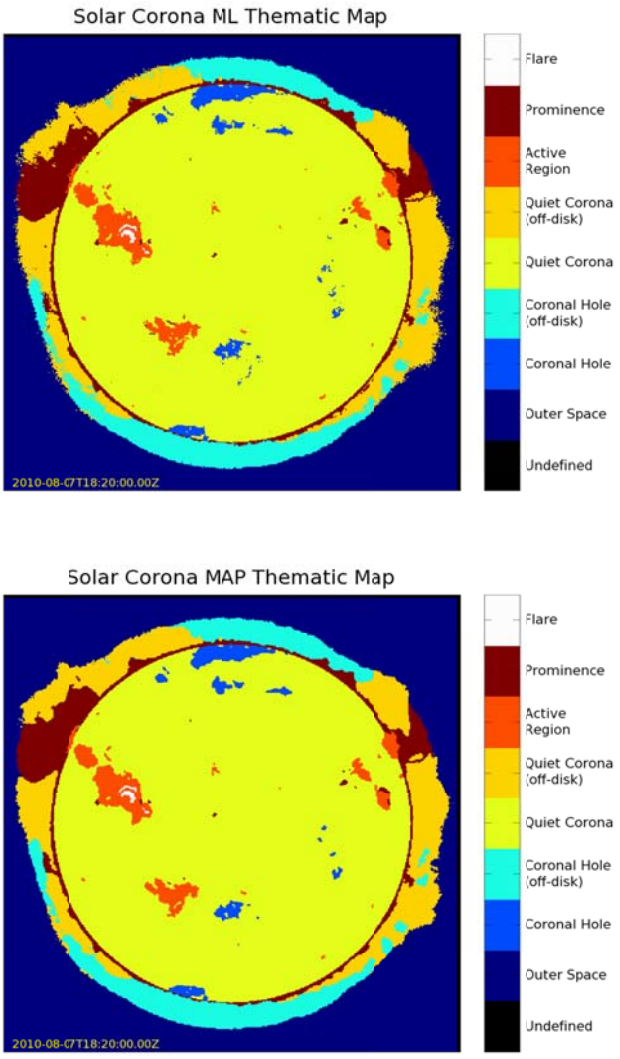


Figure 10 Thematic maps (ML & MAP) for simulated long exposures. Sample thematic maps generated from images taken at onset of M class solar flare.

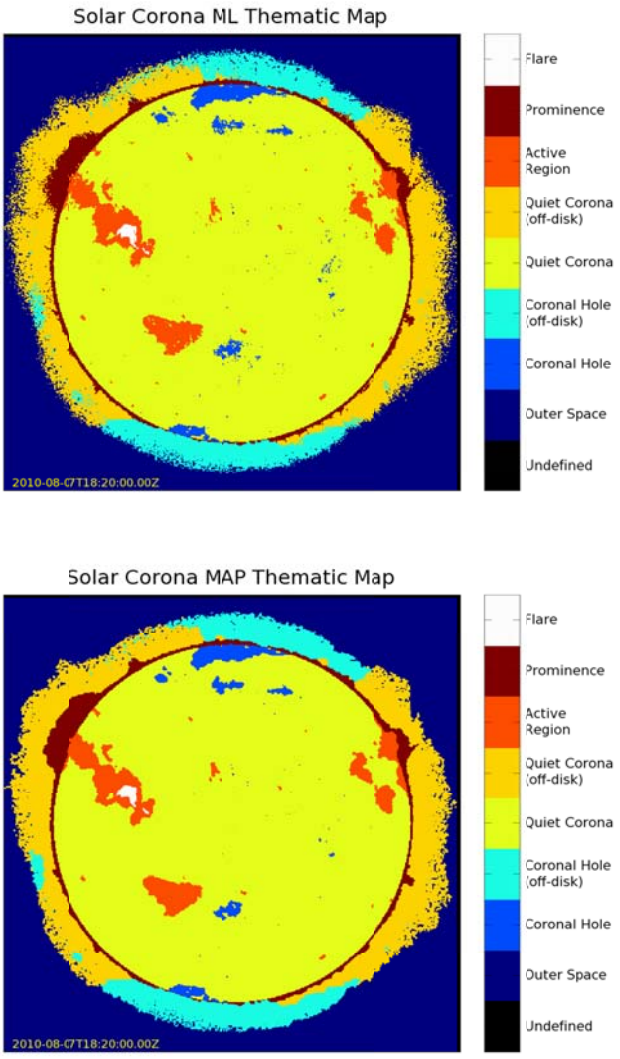


Figure 11 Thematic maps (ML & MAP) for simulated short exposures. Sample thematic maps generated from images taken at onset of M class solar flare.

4.2.1 Precisions and Accuracy Estimates

The accuracy of a pixel classification scheme can be characterized using a so-called "confusion" matrix. This contingency table for more-than-binary output is a square matrix with diagonal members that hold the number of correctly classified pixels, and off-diagonal members that hold the number of incorrectly classified pixels. The off-diagonal values down a column are label counts from the test data for each class, and correspond to type II or omission error (false negatives). The values across a row are label counts from the pixel classification, and correspond to type I, or commission error (false positives).

The sum of a column from a confusion matrix represents the number of pixels that should have been identified as the class corresponding to that column. The sum of a row represents the number of pixels that were identified as the class corresponding to that row. The ratio of the diagonal to these totals provides class-specific "producer's" and "user's" accuracy, respectively. The overall accuracy is determined by the ratio of the sum of the diagonal members of the confusion matrix to the total number of classified pixels.

To obtain a measure of the overall accuracy in a manner that considers type I and type II errors (i.e., the off-diagonal members of the confusion matrix), the κ (read "kappa") coefficient offers a reasonable normalized metric, where 1 is a perfect classification, and 0 implies a classification as good as random chance.

$$\kappa = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}\left(x_{i+} \cdot x_{+i}\right)}{N^2 - \sum_{i=1}^{r}\left(x_{i+} \cdot x_{+i}\right)} \tag{4.1}$$

Here, r is the number of columns (and rows) in the confusion matrix, x_{ii} is the number of correct classifications for class i (i.e., the diagonal members of the confusion matrix), and x_{i+} and x_{+i} are the row and column totals for each class, respectively. N is the total number of observations, and should equal the sum of either the row or column totals if the confusion matrix was constructed properly.

Table 6-Table 10 present confusion matrices for ML and MAP thematic maps derived from "truth", simulated long exposures, and simulated short exposures. Their captions include the κ coefficient calculated from the matrix values.

Table 5. Confusion matrix for ML solution derived from "truth" images (κ = 0.961)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	0	0	0	0	0	0	0	**29,243**
Coronal Hole	0	3233	0	0	8	0	0	0	**3241**
Coronal Hole (off-disk)	0	0	5806	0	22	0	20	0	**5848**
Quiet Corona	0	30	0	20,281	23	5	236	0	**20,575**
Quiet Corona (off-disk)	0	0	802	0	14,904	0	430	0	**16,136**
Active Region	0	0	0	66	7	2418	3	1	**2495**
Prominence	0	1	2	0	696	20	3156	0	**3875**
Flare	0	0	0	0	0	57	0	764	**821**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

Table 6. Confusion matrix for MAP solution derived from "truth" images (κ = 0.962)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	0	0	0	0	0	0	0	**29,243**
Coronal Hole	0	3240	0	0	8	0	0	0	**3248**
Coronal Hole (off-disk)	0	0	5795	0	24	0	16	0	**5835**
Quiet Corona	0	24	0	20,303	30	5	236	0	**20,598**
Quiet Corona (off-disk)	0	0	815	0	14,910	0	417	0	**16,142**
Active Region	0	0	0	44	0	2443	0	1	**2488**
Prominence	0	0	0	1	688	15	3176	0	**3879**
Flare	0	0	0	0	0	37	0	764	**801**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

Table 7. Confusion matrix for ML solution derived from long exposures (κ = 0.962)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	0	10	0	2	0	0	0	**29,255**
Coronal Hole	0	3233	0	0	8	0	0	0	**3241**
Coronal Hole (off-disk)	0	0	5774	0	60	0	17	0	**5851**
Quiet Corona	0	30	0	20,305	30	15	232	0	**20,612**
Quiet Corona (off-disk)	0	0	825	0	14,961	0	464	0	**22,423**
Active Region	0	0	0	40	0	2454	2	21	**2517**
Prominence	0	1	1	2	599	12	3130	0	**3745**
Flare	0	0	0	0	0	19	0	744	**763**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

Table 8. Confusion matrix for MAP solution derived from long exposures (κ = 0.961)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	0	9	0	2	0	0	0	**29,254**
Coronal Hole	0	3239	0	0	7	0	0	0	**3246**
Coronal Hole (off-disk)	0	0	5701	0	72	0	17	0	**5790**
Quiet Corona	0	25	0	20,312	31	16	235	0	**20,619**
Quiet Corona (off-disk)	0	0	899	0	14,964	0	446	0	**16,309**
Active Region	0	0	0	35	0	2467	0	21	**2523**
Prominence	0	0	1	0	584	9	3147	0	**3741**
Flare	0	0	0	0	0	8	0	744	**752**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

Table 9. Confusion matrix for ML solution derived from short exposures (κ = 0.950)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	7	26	3	62	0	0	0	**29341**
Coronal Hole	0	3220	0	15	7	0	0	0	**3242**
Coronal Hole (off-disk)	0	0	5668	0	240	0	34	0	**5942**
Quiet Corona	0	37	0	20,247	30	4	205	0	**20,523**
Quiet Corona (off-disk)	0	0	916	0	14,482	0	557	0	**15,955**
Active Region	0	0	0	82	1	2397	0	2	**2482**
Prominence	0	0	0	0	838	18	3049	0	**3905**
Flare	0	0	0	0	0	81	0	763	**844**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

Table 10. Confusion matrix for MAP solution derived from short exposures (κ = 0.955)

	Outer Space	Coronal Hole	Coronal Hole (off-disk)	Quiet Corona	Quiet Corona (off-disk)	Active Region	Prominence	Flare	Sum automatic pixel labels
Outer Space	29,243	7	14	3	63	0	0	0	**29,330**
Coronal Hole	0	3242	0	3	7	0	0	0	**3252**
Coronal Hole (off-disk)	0	0	5645	0	74	0	19	0	**5738**
Quiet Corona	0	15	0	20,306	31	3	205	0	**20,560**
Quiet Corona (off-disk)	0	0	951	0	14,693	0	543	0	**16,187**
Active Region	0	0	0	35	0	2421	0	2	**2458**
Prominence	0	0	0	0	792	12	3078	0	**3882**
Flare	0	0	0	0	0	64	0	763	**827**
Sum expert-chosen pixel labels	**29,243**	**3264**	**6610**	**20,347**	**15,660**	**2500**	**3845**	**765**	**82,234**

The accuracies of these classifications are almost certainly overstated. This is not surprising because the same data was used for training and testing the thematic map, and because the training data was selected in a highly non-random manner (usually as clusters of similar pixels). Ideally, training and test data would be generated independently, and the test data at least should be selected in a more random fashion.

One important point to note here, however, is that as input images become more corrupted by noise, the MAP thematic map solution tends to perform increasingly better, as measured by the κ coefficient, relative to the ML solution. Other case studies in which the proxy images were corrupted by significantly more noise bear this point out even more. This result justifies the decision to add a level of complexity to the thematic map algorithm in the form of smoothness priors.

4.2.2 Error Budget

TBD

5 PRACTICAL CONSIDERATIONS

5.1 Numerical Computation Considerations

Test runs of a prototype Thematic Map algorithm suggest that ML classification, plus ~10 iterations of the ICM statistical smoother, should complete in well under a minute on a modern PC workstation with no substantial optimization. However, if this is not sufficient, there are several options for speeding up the algorithm:

- there are several runtime variables that are static, and can be defined outside any loop over the million or so image pixels (e.g., the covariance matrix determinant and inverse);
- Equation (3.9) can be simplified if one takes its natural logarithm, thus causing the exponential function to disappear; all remaining terms with a natural log are constant (again, allowing them to be removed from the big loop), and because the natural log function is monotonic, maximizing log probabilities is equivalent to maximizing probabilities to obtain the ML solution for a pixel (it is more complicated to obtain a MAP solution using smoothness assumptions, but Tso & Mather (2009) explain how);
- the SUVI Thematic Map algorithm operates on a per-pixel basis, or on a pixel and its immediate neighbors, so it should be easy to parallelize the code to exploit a multi-CPU system, and increase speed as needed.

5.2 Programming and Procedural Considerations

Multichannel pixel statistics (i.e., the class-dependent means and covariances) are to be determined prior to runtime from an exhaustive set of training data (i.e., the multichannel input images are not missing any expected channels, and the solar scenes contain all solar features expected to be classified during daily operations). They should be updated periodically using training data collected by experts in solar imagery, preferably using a more randomized procedure than that used to generate the examples presented in Section 4.2.1. These updates should occur at least once per year, if not more frequently, and are necessary to track changes in instrument performance as well as variations in solar activity. All pixel statistics updates must be made available to the space weather data processing framework well before their first use.

The Thematic Map algorithm should work whether a subset of SUVI channels is used, all six SUVI channels are used, or even additional non-SUVI channels are used. All input images/channels are expected to be made available to the SUVI Thematic Map algorithm via the Space Weather data processing framework, although the algorithm itself should be responsible for assembling multichannel images so that they correspond to the multichannel pixel statistics provided as runtime configuration parameters.

All individual input images should be translated, scaled, and rotated to align with a common point of view (POV). This POV should remain constant over time so Thematic Maps for different epochs can be compared. One useful POV is to center the sun in the pixel array, rotate it so that solar north points to the top of the image array, and rescale the solar disk so that its pixel radius matches what is expected for SUVI at a distance of 1AU from the sun.

5.3 Quality Assessment and Diagnostics

The quality of the SUVI Thematic Map algorithm should be assessed and tracked over time using “confusion matrices”, and associated κ coefficients, as described in Section 4.2.1. These in turn should be determined using test data collected periodically by experts in solar imagery, preferably using a more randomized procedure than that used to generate the examples presented in Section 4.2.1. These (re)assessments should occur at least once a year, if not more frequently. They may coincide with updates to the multichannel pixel statistics, although care should be taken to keep the training and test data statistically independent.

If possible, the quality of the algorithm output should be compared to what would be obtained using alternative but similar input data. For example, if the SDO-AIA

instrument is still operational, the training data (pixel coordinates and labels) should be used to generate SDO-AIA-specific statistic (because SDO-AIA's calibration factors will likely vary substantially from SUVI's), and these statistics should then be used as input along with the SDO-AIA images to generate alternative thematic maps for comparison with SUVI-derived thematic maps.

These comparisons should be made using confusion matrices and κ coefficients generated from identical pixel coordinates and pixel labels.

Finally, as there are at least two downstream L2 SUVI products that are expected to use SUVI Thematic Maps as their primary input (coronal hole boundaries and flare location reports), the quality of these downstream products should be assessed and tracked with an eye toward the influence of their respective inputs.

5.4 Exception Handling

The following is a list of exceptions and required actions for the SUVI Thematic Maps algorithm:

- Class-dependent covariance matrix fails positive-definiteness test
  - **action:** assign "undefined" class (i.e., 0) to all pixels in thematic map
  - **action:** check for other invalid covariance matrices
  - **action:** adjust metadata to reflect which classes had invalid covariance matrices
- Input pixel value is NaN, OR input pixel is flagged as "missing" or "bad", OR input pixel weight is equal to zero:
  - **action:** stop processing current pixel, and assign "undefined" class
  - **action:** proceed to next pixel
- One or more entire input channels are missing, OR there is bad or incomplete metadata for one or more input channels, OR a configurable threshold number or percentage of bad pixels (e.g., MAXBADPIX) is exceeded for one or more input channels:
  - **action:** assign "undefined" class to all pixels in thematic map
  - **action:** check for other invalid channels
  - **action:** adjust metadata to reflect which channels were bad when generating the thematic map
  - **TBR:** it may make sense for MAXBADPIX to be channel-dependent

5.5 Algorithm Validation

The SUVI Thematic Map algorithm will be validated by ascertaining its operational value to the SWPC forecast office, both as an independent data product, and as input to other more derived data products. Ideally this will be done as part of a GOES Program Proving Ground demonstration project.

6 ASSUMPTIONS AND LIMITATIONS

6.1 Performance

The SUVI Thematic Map algorithm is a supervised pixel classification scheme. As such, its accuracy can only be as good as the quality of training and/or test data used to configure it. This is by design, and places a higher premium on matching forecaster and other experts' judgment than on any kind of physics.

It should be noted that a deliberate choice was made to use a statistical classifier that assumes all class-dependent multichannel pixel distributions are Gaussian. These distributions can be fully characterized by their means and covariances, allowing actual probabilities of membership to a specific pixel class to be easily calculated. This in turn offers an intuitive explanation for why one pixel class might have been selected over another. However, depending on how training data is selected, this assumption is not always valid, and can occasionally lead to end results that, while statistically consistent, are not realistic or intuitive.

6.2 Assumed Sensor Performance

The SUVI instrument is expected to perform to its operational requirements described in the GOES-R Mission Requirements Document (MRD), and related documentation. As of the time of this writing, the SUVI vendor has informally agreed to provide metadata with each image that informs the user or algorithm if/when these level 1b performance requirements may not be getting met. This includes, for example, channel-specific noise floors and saturation thresholds in units comparable to the pixel values.

6.3 Possible Product Improvements

6.3.1 Covariance Matrix Check

It is expected that the multichannel pixel statistics for each solar feature class can be assembled into a valid positive definite covariance matrix. The algorithm should test for this condition at runtime, and exit immediately on failure. A positive definite matrix is one where all eigen values are positive. Round-off error can lead to eigen values that are not quite zero, even though the matrix is NOT positive definite. To ensure that the covariance matrix is positive definite, a test should invoke some kind of numerical tolerance that returns a “fail” result for positive definiteness if any eigen value is less than the tolerance. This tolerance should scale with the so-called Frobenious norm of the matrix:

$$\text{tol.} \propto f = \sqrt{\sum_k \left[\mathbf{C}^{\mathrm{T}}\mathbf{C}\right]_{kk}} \tag{6.1}$$

Typically, *f* is multiplied by the machine epsilon (*eps*; i.e., the smallest floating point number that can be added to 1.0 and still change its value) corresponding to binary precision of the data type used to hold $\mathbf{C}$.

6.3.2 Probability Thresholds

It is not uncommon for the maximum conditional probability that a pixel belongs to a specific class (Eq. (3.9)) to be exceedingly low, especially in low-brightness regions. However, such results are often erroneous, resulting in increased classification noise. Perhaps it makes sense then to prescribe a minimum probability below which pixels will be considered unclassifiable.

It turns out that the scalar product of the matrix multiplication in Eq. (3.9) (i.e., the Mahalanobis distance) follows a chi-square distribution for ρ degrees of freedom. This should allow one to determine a threshold Mahalanobis distance for each class based on a specified critical value (0.95, 0.99, 0.999, etc.). This can be easily transformed into a minimum significant probability. If this threshold is not met, the probability should simply be set equal to zero, and if all class-specific probabilities equal zero, the thematic map pixel in question should be labeled zero, or “undefined”. If the ambiguity that arises between these pixels and those left undefined due to missing data is not acceptable, the label might be multiplied by -1 instead.

6.3.3 Improved Exception Handling

The initial version of the SUVI Thematic Map algorithm does not handle the case of a missing channel at the pixel level, except to set the multichannel pixel class as undefined (i.e., zero). However, it may be desirable to generate class-dependent probabilities with an incomplete set of channels, even for a specific pixel. This requires adjusting class-dependent means and covariance matrices accordingly (i.e., removing the channel-specific elements), and re-calculating the normalization constant. This could be algorithmically onerous if performed for individual pixels, but is none-the-less possible.

It is not always appropriate to proceed directly in this manner. Degradation may be too severe if too few "good" channels are available, either per-image, or per multichannel pixel. The following lists possible exceptions, and the required actions in response to these exceptions:

- one or more input channels are missing, OR there is bad or incomplete metadata for one or more input channels, OR a configurable threshold number or percentage of bad pixels (e.g., MAXBADPIX) is exceeded for one or more input channels
    - **action:** adjust multichannel pixel statistics to reflect bad or missing channel, then proceed normally
    - **action:** adjust metadata to reflect which channels were skipped when generating the thematic map
    - **TBR:** it may make sense for MAXBADPIX to be channel-dependent
- configurable threshold number of bad channels (e.g., MAXBADCHAN) exceeded
    - **action:** assign "undefined" class (i.e., 0) to each pixel
    - **action:** adjust metadata to reflect bad channels
- configurable threshold number of bad channels (e.g., MAXBADCHAN) is exceeded for a given multichannel pixel
    - **action:** skip probability calculations and assign "undefined" class (i.e., 0) to pixel; note that this action does not allow one to specify which channel(s) may have been bad, and is therefore **TBR**.

7 REFERENCES

Debevec, P. E. and M. Jitendra (1997), Recovering high dynamic range radiance maps from photographs, *SIGGRAPH Proceedings*

Li, S.Z. (2009), *Markov Random Field Modeling in Image Analysis, 3rd Ed.*, Advances in Pattern Recognition, Springer-Verlag

Mather, P.M. (2004), *Computer Processing of Remotely Sensed Data: An Introduction, 3rd Ed.*, John Wiley & Sons Ltd.

Reece, S., and S. Roberts (2010), Generalised Covariance Union: A Unified Approach to Hypothesis Merging in Tracking, *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 46, No.1

Rigler, E.J., S. Hill (2011), *GOES R SUVI Coronal Hole Image Implementation and User's Guide, Version 1.0*.

Rigler, E.J., S. Hill (2011), *GOES R SUVI Coronal Hole Image Algorithm Test Plan and Results, Version 1.0*.

Rigler, E.J., S. Hill (2011), *GOES R SUVI Thematic Maps Implementation and User's Guide, Version 1.0*.

Rigler, E.J., S. Hill (2011), *GOES R SUVI Thematic Maps Algorithm Test Plan and Results, Version 1.0*.

Stern, R.A., L. Shing, and M.M. Bouke (1994), Quantum efficiency measurements and modeling of ion-implanted, laser-annealed charge-coupled devices: x-ray, extreme-ultraviolet, ultraviolet, and optical data, *Applied Optics*, Vol. 33, No. 13

Thompson, W. T. (2006), Coordinate systems for solar image data, *Astronomy & Astrophysics*

Tso, B., and P.M. Mather (2009), *Classification Methods for Remotely Sensed Data, 2nd Edition*, CRC Press

Need reference for SDO-AIA images