Using Data Mining Decision Tree Method to Identify the Optimal Fire Detection Thresholds

Yingxin Gu[1], Ivan Csiszar[2], Wei Guo[1], and Marina Tsidulko[1]

[1]IMSG Inc. at NOAA NESDIS Center for Satellite Applications and Research (STAR), College Park, MD USA, [2]NOAA NESDIS STAR, College Park, MD USA

Introduction

The real (or near real) time operational fire products generated from the Satellite Products and Services Division of the National Environmental Satellite, Data, and Information Service (NESDIS) at NOAA are derived from both Geostationary and Polar satellite observations. The GOES-R Fire Detection and Characterization product produces fire information from ABI's 5 minute CONUS and 10 minute Full Disk scans at 2 km resolution. The JPSS Visible Infrared Imaging Radiometer Suite (VIIRS) Active Fire (AF) products are derived from the 750-m M-band and 375-m I-band observations. The algorithms and their predecessors were developed separately over the last 25 years to take advantage of their respective platforms and in response to the needs of different users.

The main objective of this study is to apply AI machine learning (ML) technique to identify the optimal fire detection rules and thresholds that performs comparably to or even better than the current process-based fire detection algorithms do. Landsat and GOES-R data was selected in this study to demonstrate the proposed approach. See5/C5 data mining tool (https://www.rulequest.com/see5win.html#TREES) was used to develop the fire detection Decision Tree (DT) models. The main advantages of the See5/C5 algorithm include (1) non-parametric (independent of data distribution), (2) handling both continuous and discrete (categorical) data, (3) less sensitive to errors in the training data, and (4) generating the interpretative classification rules (https://www.rulequest.com/see5-pubs.html).

Data and processing procedures

1. <u>DT training data selection:</u>

- Collected the 30-m high resolution Landsat fire product (2019) which was used as the "truth" fire data
- Upscaled the 30-m Landsat fire pixels to 2-km to match the geolocations of the GOES-R pixels (i.e., generated the Landsat based GOES-R fire pixels)
- Obtained the GOES-R ABI Level 1B data for the Landsat fire dates/times from NOAA CLASS (https://www.bou.class.noaa.gov/saa/products/welcome), which were used for processing the DT training data
- Calculated the brightness temperature (BT) or reflectance for ABI bands 2,3,6,7,14,and 15, calculated the background (3x3 to 19x19 windows) means and mean absolute differences (MADs) for BT07, BT14 and BT07-14
- Randomly selected ~2200 Landsat based GOES-R non-fire pixels across the CONUS in 2019
- Extracted the BT, reflectance, BT band difference, and background values for the selected fire and non-fire pixels (~3000 samples for the DT training data)

2. <u>DT fire detection model development</u>

- Built up the DT training data which include the Landsat based GOES-R fire mask (dependent variable) and 18 independent fire detection variables (e.g., BTs and background BTs for bands 3.9 and 11.2 µm)
- Developed the DT models using See5/C5 software and the DT training data. The model accuracy assessment for the testing data was included in the DT model outputs

3. <u>DT fire product evaluation and DT model re-adjustment</u>

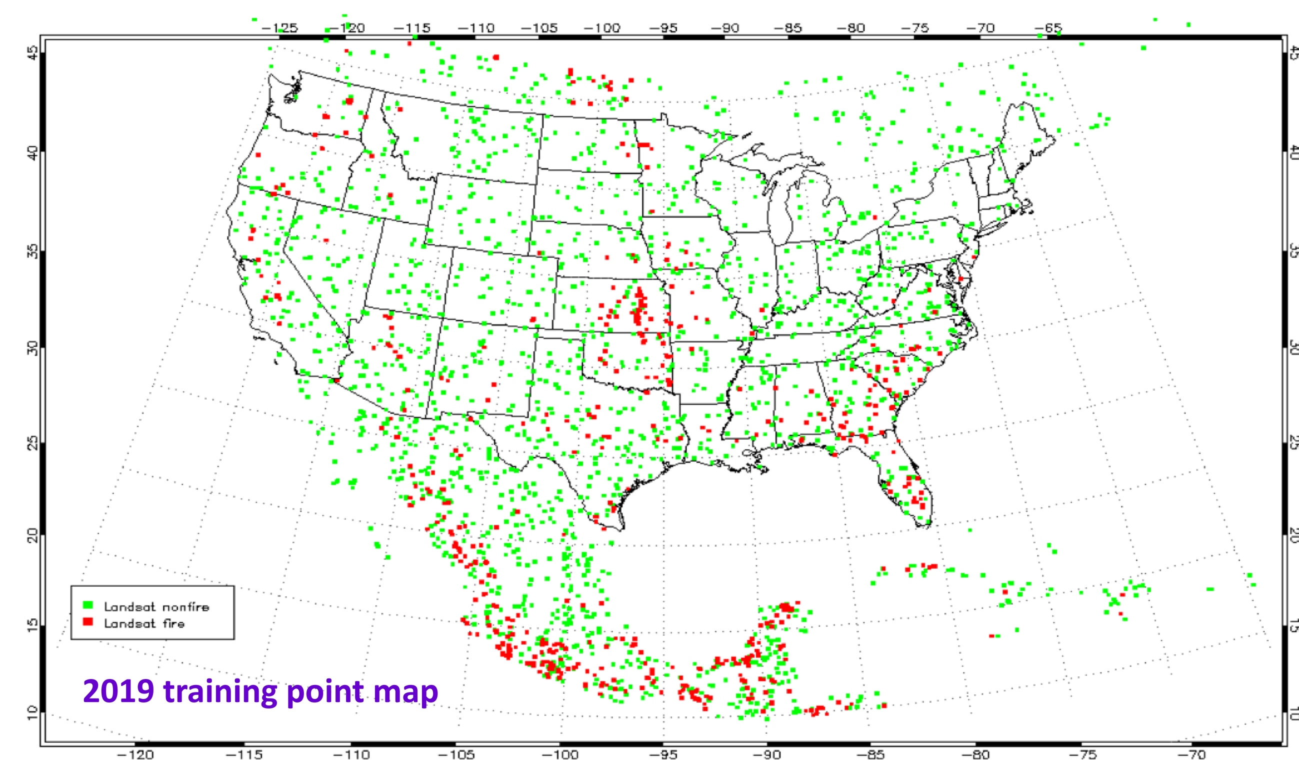
- Run the DT models for the 2019 GOES-R data and evaluated the derived DT fire maps by multi-sensor (Landsat, VIIRS, GOES DT, and GOES ABBA) fire product comparisons
- Re-adjusted the DT rules and thresholds based on the multi-sensor fire product comparison maps
- Identified the optimal DT rules/thresholds and performed accuracy assessments on the final DT models

Flow chart on processing procedure and future plan

<u>Initial test</u>
2019 Landsat fire based GOES pixels → Training data selection strategy (random samples, balance, rare cases) → Training data derivation (BT, ΔT, latitude, background) → Decision tree model development / Accuracy assessment → Optimal fire test thresholds selection → IDL code for the DT model, evaluation, and validation

<u>Future test</u>
2015-2020 historic (>5 year) data: Satellite data (BT, ΔT, etc.); Environmental variables (e.g., Ecosystem type, DEM data) → Training data selection and derivation → Developing a more <u>robust</u> DT model → Optimal fire test thresholds selection → IDL code for the DT model and validation → Apply DT method to the other satellites (e.g., VIIRS) if the results are good

DT model training data and accuracy assessment results



<u>Satellite derived variables used in the DT training data:</u>

- Latitude
- Longitude
- Cloud mask (0, 1)
- BT07
- BT14
- BT07-14
- BT07m
- BT14m
- BT07-14m
- BT07mad
- BT14mad
- BT07-14mad
- BkgFireMeanT07
- BkgFireMadT07
- Refl02 (0.64 µm)
- Refl03 (0.86 µm)
- Refl06 (2.2 µm)
- BT15 (12.3 µm)

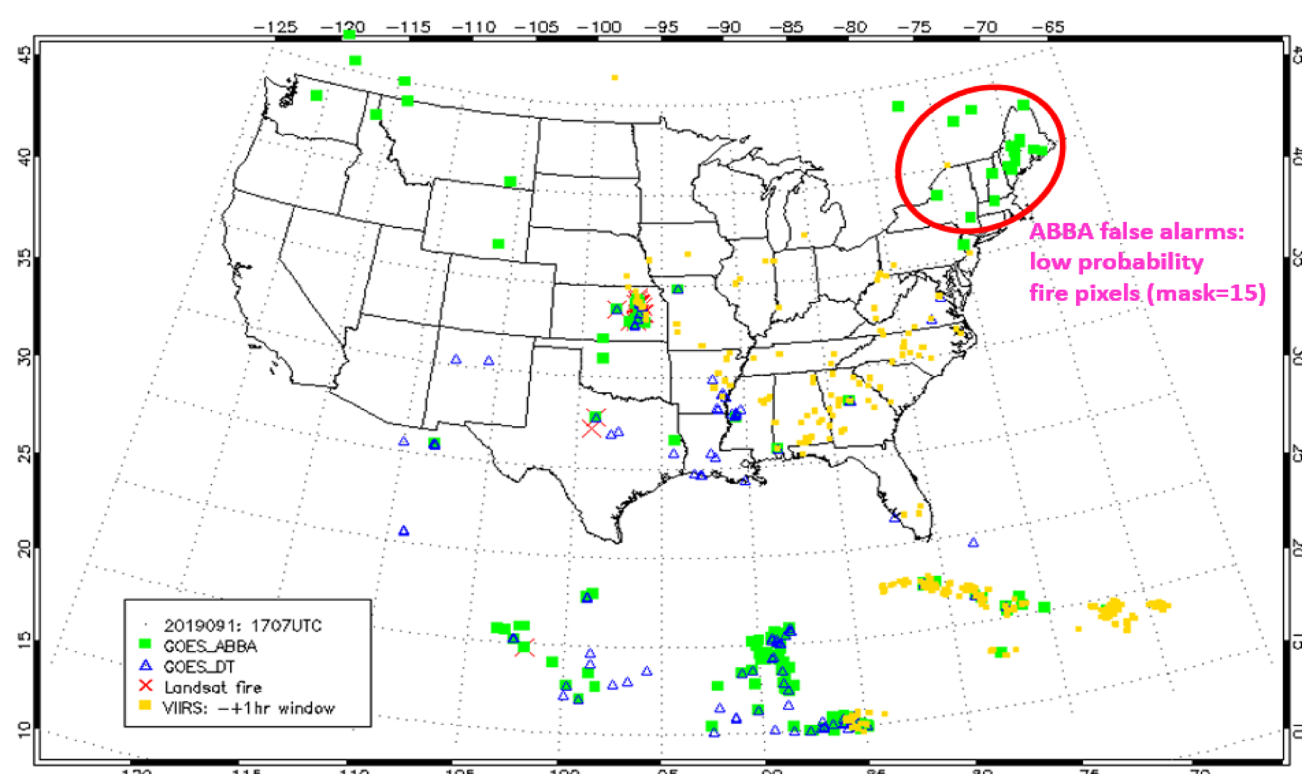
<u>DT model evaluation based on the training data (2980 cases):</u>

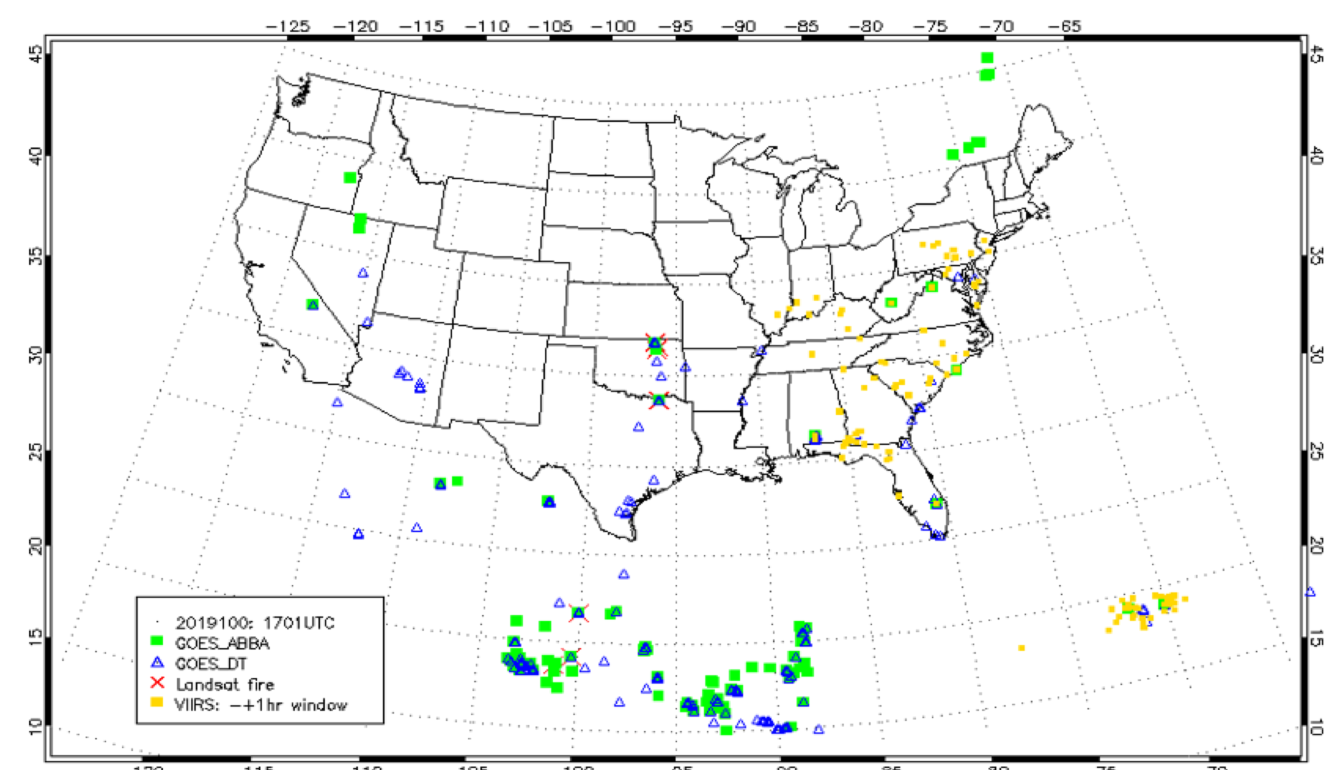
Total number of DT rules: **27**
Detection rate: **88%**
False alarm rate: **3%**

Attribute usage:

100% BkgFireMad07
56% BT07
43% Refl03
35% BT07mad
24% BT14mad
21% Latitude
17% BT07-14mad
16% BT07-14
16% BT07-14m
14% Refl06
13% Refl02
9% BT07m
2% BT14m

Multi-sensor fire product comparisons and DT model accuracy assessments





Detection Method	20 example dates in 2019			
	Matched Landsat fire events	Matched Landsat rate (Detection rate)	No. GOES fire events unmatched VIIRS fire events (within ±1 hour time window, in VIIRS covered region)	Unmatched rate
GOES DT model	53	60%	68	9%
GOES ABBA	57	64%	61	8%
Total Landsat fire events	89		Total VIIRS fire events: 695	

Conclusions and future recommendations

- ❑ The initial fire detection DT model has a comparable fire detection and false alarms rates compared to the GOES ABBA algorithm (60% vs 64% fire detection rates, 9% vs 8% false alarm rates).
- ❑ One GOES ABBA false fire cluster (20190911707, red circle area) with low probability fire mask code (15) was found. The DT fire product can be used as a reference for identifying the false fire alarms caused by the low probability GOES ABBA fires.
- ❑ The current DT model was only based on 1-year (2019) data, which is not enough to represent the entire climate and environment conditions of the COUNS. Collecting more training data (>5 year data) to build a **more robust DT model** is recommended in the future work.
- ❑ Adding more independent variables (e.g., ecosystem type, elevation, and detailed land cover type) into the model training data to further improve the DT model is also recommended in the future research.