

1 **New Algorithms for Detecting Forest Fires on a Global Scale**

2 **From MODIS Time Series Analysis**

3
4 Anuj Karpatne, Xi Chen, Yashu Chamber, Varun Mithal, Michael Lau,
5 Karsten Steinhaeuser, Shyam Boriah, Michael Steinbach, and Vipin Kumar

6 University of Minnesota, Minneapolis, MN

7 {anuj,chen,chamber,mithal,mwlau,ksteinha,sboriah,steinbac,kumar}@cs.umn.edu

8
9 Christopher Potter * and Steven Klooster

10 NASA Ames Research Center, Moffett Field, CA

11 chris.potter@nasa.gov, sklooster@gaia.arc.nasa.gov

12
13 * Corresponding author

14 Revision Date: March 9, 2012

15
16

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

Abstract. Mapping forest fires globally is an important task for supporting climate and carbon cycle studies. There are two primary approaches to fire mapping: field- and aerial-based surveys, which are costly and limited in their extent; and satellite remote sensing-based approaches, which are more cost-effective but pose several interesting methodological and algorithmic challenges. In this paper, we describe evaluate a new algorithm framework for mapping forest fires based on satellite observations from NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS) instrument. A systematic comparison and validation against ground truth sources with alternate approaches across diverse geographic regions demonstrates that our algorithmic paradigm is able to overcome many of the limitations in both data and methods employed by prior efforts. We quantitatively show that the new framework out-performs the well-known MODIS Burned Area (BA) framework in the states of California (US), Georgia (US), Yukon, (Canada), and Victoria (Australia). Results demonstrate that our new framework is highly robust to noise in one of its primary inputs, MODIS Active Fires (AF), which is known to have low precision.

1. Introduction

Land cover change is a priority issue for policymakers at the local, national, and international scale (Solomon *et al.*, 2007). Land use contributes about 15% of global carbon dioxide emissions to the atmosphere on an annual basis, and uncertainty in land use change emissions is the highest of any flux component of the global carbon budget (Friedlingstein *et al.*, 2010). Policymakers in the United Nations Framework Convention on Climate Change negotiations have thus agreed to address land use change in a framework for Reducing Emissions from Deforestation and Degradation (REDD) (Miles and Kapos, 2008). Consequently, there is a pressing need to identify strategies for monitoring, reporting, and verifying land use change and emissions from forest cover disturbances in a timely and accurate manner (DeFries *et al.*, 2006; Achard *et al.*, 2007).

46 Forest fires are a major land cover change that can be caused by both natural (lightning) or
47 anthropogenic factors. Accurate and low-cost fire mapping methods are important for
48 understanding the frequency and distribution for forest fires (Pan et al., 2011). While monitoring
49 fires in near-real time is critical for operational fire management, mapping historical fires in a
50 spatially explicit fashion is also important for a number of reasons including climate change
51 studies (e.g., examining the relationship between rising temperatures and frequency of fires), fuel
52 load management (e.g., deciding when and where to conduct controlled burns), and carbon cycle
53 studies (e.g., quantifying how much CO₂ is emitted by fires for reduction efforts such as UN
54 REDD).

55

56 There are two primary approaches to mapping forest fires: (1) field surveys combined with
57 aerial observations, which allow detailed mapping of land cover changes, but are limited in their
58 spatial extent and temporal frequency because of their high cost (Liu and Cai, 2011); and (2)
59 satellite remote sensing-based techniques, such as those from NASA's Moderate Resolution
60 Imaging Spectroradiometer (MODIS) instrument, which offer the most cost-effective method for
61 mapping fires. MODIS data are obtained freely with global repeated coverage. While numerous
62 efforts have mapped forest disturbances at local-to-regional scales (including Badarinath et al.,
63 2011; Bergeron et al., 2004; Cochrane, 2009; Niklasson and Granström, 2000; Potter et al., 2003,
64 2005 and 2007; Somashekar, 2009; Stocks et al., 2002; Talon et al., 2005), only two spatially
65 explicit efforts exist that regularly map fires at a global scale: the MODIS Active Fire (AF) and
66 Burned Area (BA) products (Justice et al., 2011).

67

68 Burned area mapping from remote sensing data is essentially a problem of change (or
69 anomaly) detection. Such datasets have both temporal and spatial dimensions, and there are two
70 primary ways to address the problem. There are approaches that focus on the temporal aspect,
71 wherein fires are mapped based on time series analysis (Mithal et al., 2008; Roy et al., 2002).
72 These types of methods usually take into consideration properties such as seasonality, variability
73 and temporal coherence in a given time series. On the other hand, there are approaches that treat
74 the data as a sequence of image snapshots, and image processing-based methods (Brewer et al.,
75 2005; Gitas et al., 2004) are used to detect burned areas. Such methods commonly take
76 advantage of the spatial properties inherent in the data, for instance, the fact that burned pixels

77 tend to cluster together. Recently, techniques have been developed for land cover change
78 detection that utilize both spatial and temporal properties (Giglio et al., 2009; Lhermitte et al.,
79 2008; Lunetta et al., 2006) to take advantage of autocorrelation structures present along *both*
80 dimensions in the remote sensing data.

81

82 Nonetheless, satellite remote sensing data poses several challenges, some unique, for
83 algorithm development, such as (1) the presence of noise and outliers, (2) inaccuracy and
84 incompleteness of signals, (3) high natural variability and seasonality, (4) influence of climatic
85 factors, (5) availability of multiple temporal scales. In the case of forest fire mapping, additional
86 factors include potential obstruction of the signal due to smoke and the similarity of the signal
87 relative to other types of changes and events, such as logging and wind damage.

88

89 In this paper, we introduce a new spatio-temporal detection framework for forest fire mapping
90 that is both robust and computationally scalable to large regions. Specifically, the proposed
91 approach is unsupervised in nature and exploits both the temporal and spatial structure in the data
92 to combine multiple sources of information. Using independent wildfire perimeter data sets, we
93 have comprehensively evaluated our approach, as well as those from alternate methods, across
94 different forest climatic zones. Finally, we examined specific properties of the MODIS BA
95 algorithm and show how our approach is able to overcome some of its limitations.

96

97 **2. Satellite Remote Sensing Data**

98 Global remote sensing datasets are available from a variety of sources at different
99 resolutions. Our proposed forest fire mapping framework is based on two remotely sensed
100 composite data products from the MODIS instrument aboard NASA's Terra satellite, which are
101 available for public download (U. S. Geological Survey, 2011). Specifically, we use the
102 Enhanced Vegetation Index (EVI) from the MODIS 16-day Level 3 1-km Vegetation Indices
103 (MOD13A2) and the Active Fire (AF) from the MODIS 8-day Level 3 1-km Thermal Anomalies
104 & Fire products (MOD14A2). EVI measures "greenness" (area-averaged canopy photosynthetic
105 capacity) as a proxy for the density of vegetated biomass at a particular location (Figure 1). The
106 AF product is designed to identify thermal anomalies from the middle infrared spectral

107 reflectance bands (Justice et al., 2011) and is used heavily in operational situations by fire-
108 fighting agencies around the world. In order to separate forests from other land cover types, we
109 use the MODIS Vegetation Continuous Fields (VCF) dataset (MOD44B), which provides the
110 percent tree cover for every pixel. MODIS Level 3 products are provided on a global 1km
111 sinusoidal grid in $10^{\circ} \times 10^{\circ}$ tiles. For this study, we focus on subsets of the global MODIS data
112 based on the available wildfire perimeter information for validation.

113

114 Fire-related satellite data products broadly fall into two categories: active fire products,
115 which capture the location and intensity of fires burning at the time of observation (the
116 prototypical example being the AF product); and burned area products, which map areas that
117 were burned by fires based on historical observations. In the following section, we review two
118 existing algorithms for mapping burned areas.

119

120 2.1. The V2DELTA Algorithm

121 Mithal et al. (2011a) presented a time series change detection algorithm that incorporates
122 natural seasonal variation into the change detection framework. The algorithm, called
123 V2DELTA, identifies abrupt forest disturbances using MODIS EVI as an input. More
124 specifically, the V2DELTA compares a drop in EVI the variability in a fixed “training” window,
125 thereby providing a mechanism to ascribe significance to any given drop. This relies on the
126 assumption that EVI values in the initial window were not affected by a land cover change, thus
127 enabling the algorithm to differentiate abrupt changes from naturally occurring vegetation
128 changes. While V2DELTA detected many types of forest disturbances (Mithal et al., 2011b), it
129 failed to distinguish forest fires from other land cover changes, such as those caused by non-fire
130 inducing climatic factors (e.g., droughts).

131

132 2.2. The BA Algorithm

133 The burned area approach (henceforth called BA) presented by Giglio et al. (2009) can be
134 viewed as a semi-supervised Bayesian classification method with two classes: burned and
135 unburned. The technique builds on key concepts and ideas developed over several years by

136 Giglio et al., (2006) and others (Fraser et al., 2000; George et al., 2006; Loboda et al., 2007; Roy
137 et al., 1999). The BA algorithm is run on a regular basis using the latest spectral reflectance and
138 the MODIS AF product. The output is released by the University of Maryland as a product
139 called MODIS Direct Broadcast Monthly Burned Area Product (MCD64A1).

140

141 The key steps of the BA algorithm are outlined below:

- 142 1. Representative sets of samples for the burned and unburned classes are constructed. The
143 sample pixels for each class are discovered using conservative heuristics which label pixels
144 as unburned or burned if they pass a set of conditions.
- 145 2. The burned class is further enriched with closely related pixels from the dataset, while the
146 unburned class is refined by pruning pixels that are geographically close to burned training
147 pixels.
- 148 3. A statistic that estimates the daily loss in vegetation (ΔVI) is computed for all training
149 pixels.
- 150 4. The conditional probability distribution of the vegetation loss statistic is estimated for both
151 *the burned as well as the unburned class*, i.e., $P(\Delta VI|burned)$ and $P(\Delta VI|unburned)$.
- 152 5. Bayes' Rule is applied to obtain the posterior probability of a pixel belonging to the burned
153 class.

154 There are two major limitations in the BA algorithm. First, the burned and unburned training
155 data is very expensive to obtain manually, and is difficult to generate automatically. Second, due
156 to the inherent nature of forest fires (i.e., they happen rarely in both spatial and temporal
157 dimensions), the burned and unburned classes are often highly imbalanced. Since the Bayesian
158 classifier is biased toward the majority class, the recall will suffer when the classes are
159 imbalanced.

160

161 In the BA algorithm framework, estimating two separate distributions for the rare and
162 majority classes is likely to be affected by problems arising from small sample size and over-
163 fitting. In contrast, a one-class approach of building a null distribution over the normal population
164 and detecting rare events as anomalies in the distribution is more robust to noise and outliers as it

165 utilizes a larger sample size. We note that “normal” here denotes the unburned population which
166 may be unchanged or exhibit changes unrelated to fire (logging, drought, etc.).

167

168 **3. Methods**

169 In a new stratified framework for mapping forest fires, we have employed multiple,
170 complementary scoring mechanisms using both MODIS EVI time series and the MODIS AF
171 product. The approach begins by generating the high stratum product, which consists of pixels
172 that exhibit both abrupt change in EVI and an event in the AF signal at the same time. This
173 stratum is then augmented by very similar events in close proximity to generate the middle
174 stratum product. The lowest stratum product is generated by including loosely similar events in a
175 spatial window around the other two strata. Figure 2 shows a flowchart of this framework and
176 each task is described in detail below.

177

178 3.1. Forest Fire Detection Algorithms

179 The algorithm begins operation by looking for fire events using an initial set of candidates
180 associated with the most confident stratum. Since the AF contains information about thermal
181 anomalies, it is arguably a good choice for selecting initial candidates. AF products restrict the
182 search space by pruning out potential false positive examples caused by other land cover changes
183 such as conversion from forest to farm, or changes attributed to climatic factors such as droughts,
184 etc. Once the candidates have been initialized with events having high AF values (≥ 7), we use
185 an array of scoring mechanisms on the EVI time series to determine the significance of the given
186 event as a forest fire. The intent behind using multiple scoring mechanisms is to cover multiple
187 facets of information about the forest fire, where each score captures a distinct characteristic of
188 the EVI change at a forest fire event. Specifically, we introduce the following scoring framework:

189

190 3.1.1. *K-month Delta (KD)*

191 *K-month Delta (KD)* is an extension of the *V2DELTA*, discussed above. Similar to *V2DELTA*,
192 the inter-annual variation (*IAV*) is defined as

$$193 \quad IAV(t) = \mu(EVI(t-sl, K)) - \mu(EVI(t, K))$$

194 where, sl is the number of time steps in one annual EVI segment (23 in our case), and K is the
195 window size of segments being compared.

196
197 By modeling EVI as the combination of yearly trend and Gaussian noise which indicates the
198 normal variations attributed to seasonal changes and sensor noise, we assume that IAV follows a
199 Gaussian distribution. Specifically, we expect the distribution to be $N(0, \sigma^2)$ when there is no
200 land cover change. Therefore, the KD score at time step t is given by

$$201 \quad KD(t) = \frac{IAV(t)}{\sigma}$$

202 which is the z -score of IAV . Here, σ is estimated based on the data in a four year window
203 preceding t using bootstrapping (this makes the algorithm more robust than V2DELTA). K is set
204 to 12 months in the proposed approach.

205 206 3.1.2. Local Instant Drop (LID)

207 Forest fires are commonly observed as instantaneous drops in EVI time series, but sometimes
208 these drops may not persist for a long time period (more than two years), i.e. the observed
209 greenness recovers rapidly. Generally, these sudden drops in the EVI are much higher than the
210 normal variations attributed to climatic seasonality and sensor noise. The Local Instant Drop
211 (LID) algorithm scores the instantaneous drop at time step t in EVI by comparing the amount of
212 the drop around time step t with the normal variations ($NVar$). The algorithm accounts for noise
213 and outliers occurring in the temporal locality of a candidate fire, as well as the seasonal context
214 in EVI to improve the robustness of the scoring algorithm. The LID score is given by the
215 following equation:

$$216 \quad LID(t) = \frac{EVI(t-1,1) - EVI(t+1,1)}{NVar}$$

217 where, $NVar$ is the largest drop that occurs in the temporal neighborhood (of size 3) and in the
218 previous two year history. In order to account for the seasonality of EVI, only the time steps
219 within a small window (of size 1) in the previous years around (for a given time step) are
220 considered.

221

222 3.1.3. *Near Drop (ND)*

223 The Near Drop (ND) algorithm measures the change in the average EVI, before and after
224 a candidate forest fire event at time step t . The ND score is given by the following equation:

225
$$ND = \mu(EVI(t-k, k)) - \mu(EVI(t+1, k))$$

226 where $k=3$. ND captures the instantaneous nature of the drop in EVI. As the only score which
227 reflects the real amount of drop in EVI, ND is well-suited as a filter in our framework to be
228 satisfied by a candidate fire event detected by other scoring mechanisms, even though it is
229 affected by noise and outliers.

230

231 3.1.4. *Spatial Growing Algorithms*

232 The AF product can fail to detect forest fire events which do not register a thermal anomaly
233 because of smoke or satellite overpass timing. Thus, the initial candidate pixels might suffer
234 from low coverage. To overcome this limitation, we exploited the inherent spatio-temporal
235 autocorrelation of forest fire events to increase coverage. Since events corresponding to the same
236 forest fire occur in close proximity of space and time, we exploit this property by searching for
237 candidate fire events around the initial candidates classified as forest fires by the scoring
238 mechanism above. In the current framework, we consider the 24 spatial neighbors in a 5×5
239 spatial grid around the initial candidates, with a temporal constraint of being within one time step
240 from the change time of the initial candidate fire event. We then apply our scoring mechanism on
241 the new pool of candidate events with exactly the same scoring criteria as we used for detecting
242 initial candidate forest fire events. We iteratively grow in a spatial neighborhood to exhaustively
243 detect candidate forest fire events. They represent candidate forest fire events (middle stratum)
244 which have fire characteristics in the EVI but were not initial candidates because of the absence
245 of AF detection.

246

247 The new framework creates a pool of candidate forest fire events with a relaxed scoring
248 criteria indicating a lower confidence. We accept events to be part of this pool (lowest stratum) if
249 they exhibit a positive ND and either a moderately large LID (≥ 1.5), or a moderately large KD
250 (≥ 1.5) in conjunction with a moderate LID (≥ 1). Thus, we iteratively grow in a spatial
251 neighborhood (5×5 grid) to exhaustively include any probable candidate fire event pixels.

252

253 3.2. Algorithm Performance Evaluation

254 We examined the performance of AF, BA, the new forest fire detection algorithms detailed
255 above in several regions around the world, including the states of California (United States),
256 Georgia (United States), Yukon (Canada) and Victoria (Australia). These geographic areas
257 represented diverse regions with differing forest cover types, topography, hydrology, and MODIS
258 data quality characteristics. The following describes the independent evaluation data used in this
259 study and of the evaluation methodology.

260 *3.2.1. Forest Fire Evaluation Data Sources*

261 For each region considered in our evaluation, we obtained spatially explicit evaluation data
262 sets from government agencies responsible for monitoring and managing forests and wildfires.
263 The evaluation data were in the form of wildfire perimeter polygons, each of which is associated
264 with the date of burning. The regions studied in this paper and the respective sources of fire
265 polygon data are listed in Table 1.

266

267 Although government agencies make their best effort in documenting historical fires, wildfire
268 perimeter datasets are neither complete nor without error, due to finite resources available to any
269 mapping agency. However, inaccuracies and incompleteness are represented only in a small
270 portion of the evaluation data, and these datasets are still useful for quantitatively comparing
271 remote sensing detection methods across large spatial regions. The AF, BA, and MODIS EVI
272 products are each precisely geo-referenced by the latitude and longitude values of their pixel
273 centers. We considered a location to be positive for a forest fire event if the corresponding pixel
274 center fell inside a wildfire perimeter polygon (Table 1). Similarly, a pixel was considered to be
275 unburned (forming the negative class) only if the entire pixel fell outside a wildfire perimeter
276 polygon. The remaining pixels (which partially overlap polygon boundaries) were discarded from
277 the evaluation framework to avoid ambiguity.

278

279 Since our primary focus is on detecting forest fires, we utilized the MODIS Vegetation
280 Continuous Fields (VCF) dataset (MOD44B) which contains the percentage tree cover
281 information. We only consider pixels with high percentage tree cover (i.e., $\geq 20\%$) in our
282 evaluation scheme, a threshold commonly cited to separate forest from non-forest covers
283 (Bandyopadhyay et al., 2011; Giglio et al., 2009).

284

285 3.2.2 Evaluation Methodology

286 In this paper, we use *precision* and *recall* as evaluation metrics for quantitatively
287 comparing the performance of AF, BA, the new forest fire detection algorithms. These two well-
288 known metrics are used to evaluate the performance of algorithms in information retrieval,
289 machine learning and data mining (Tan et al., 2006). Each algorithm provides a set of positive
290 and negative events that it detects, which is evaluated using fire perimeter polygons to obtain the
291 number of true positives (*TP*), false positives (*FP*), false negatives (*FN*) and true negatives (*TN*)
292 for each algorithm (as shown in Table 2). We note that detection of a *TP* event means that the
293 pixel fell entirely inside a polygon and the time of change agrees with the fire polygon date.

294

295 The precision (*p*) and recall (*r*) values for each algorithm were determined by the
296 equations:

$$p = \frac{TP}{TP + FP}$$

297

$$r = \frac{TP}{TP + FN}$$

298 Values of *p* and *r* closer to 1.0 were indicate a higher fraction of retrieved fire events that
299 were relevant (*TP*, i.e., pixel areas detected as burned inside the fire polygon perimeters), and the
300 fraction of relevant (*TP*) instances that are retrieved, respectively (Olson and Delen, 2008). Stated
301 in another way, precision results will be highest where the number of *FP* results (i.e., pixel areas
302 detected as burned outside the fire polygon perimeters) are lowest, and recall results will be

303 highest where the number of FN results (i.e., pixel areas not detected as burned inside the fire
304 polygon perimeters) are lowest.

305 **4. Results**

306 The new forest fire detection algorithms introduced in this study consistently outperformed
307 the V2DELTA and AF algorithms in all forest regions evaluated (Table 3). The new algorithms
308 (finalized with the result labelled Level 3 in Table 3) initially achieved high precision by using
309 strict criteria (Level 1 and Level 2), then increased recall results through Spatial Growing using
310 lower thresholds (Level 3), and finally recovered the majority of the precision by discarding FP
311 results using its anomaly detection approach. Precision for the Level 3 results were above 0.92 in
312 all regions; recall results were above 0.93 in California and the Yukon and above 0.65 in all
313 regions. Although the AF and V2DELTA algorithms produced comparable recall results to our
314 Level 3 product in some cases, they both tended to generate a large number of FP results (i.e.,
315 pixel areas detected as burned outside the fire polygon perimeters), which reduced their precision
316 results markedly. This was not unexpected, particularly given that V2DELTA was developed as
317 a more general land cover change detection method and the AF product may detect fires in non-
318 forest areas.

319

320 The new forest fire detection algorithms generally performed as well, and in some cases
321 notably better, than the BA algorithm. A close examination of both FP and FN results revealed
322 that the two approaches sometime produced different errors in the same geographical region.
323 Two particular limitations of the BA algorithms were illustrated by the results for Georgia (Table
324 3c). The AF algorithm also showed low precision in Georgia, and since BA and AF both
325 generate training data, we hypothesize that BA produces error in cases where the precision of AF
326 was also poor.

327

328 We observed a peculiar case where both precision and recall are higher for the “low-quality”
329 BA, and postulate that this counterintuitive observation can be explained by the fact that MODIS
330 pixel values in the immediate (temporal) neighborhood of a forest fire tend to be of lower quality,
331 due largely to obstruction from smoke. We hypothesize that, since BA uses high quality

332 observations within a limited window surrounding the fire date, the algorithm does not have
333 sufficient information and is thus unable to correctly classify such pixel areas. To test this
334 hypothesis, we generated two histograms of the pixel reliability extracted from the MODIS
335 Quality Assurance (QA) fields (Figure 4), one for the TPs of the high-quality product, and
336 another for the FNs of the high-quality product which are true positives in the low-quality
337 version. Pixel areas that the BA correctly identified were, by and large, of high quality in the QA
338 fields, although BA was able to take advantage of the additional information from lower-quality
339 inputs and thus detect high proportion of burned pixels.

340

341 Mapped examples for the new framework results show that large fire perimeters delineated by
342 national agencies in the U. S. and Canada were replicated closely in California and the Yukon
343 with TP pixels (Figure 5). The extremely steep topography (500 m elevation changes over less
344 than 2 km inland from the ocean shore) in the coastal Big Sur, California wildfire example may
345 explain the small number of FN results within the fire perimeters from the new framework
346 algorithms.

347

348 Results of forest fire detection algorithms introduced in this study showed that a number of
349 pixel areas exhibiting strong fire characteristics were not included in the fire perimeter polygons
350 from Victoria and California (Figure 6a). Such pixels were incorrectly counted as FP results in
351 our evaluation scheme. In addition, we encountered a small number of FP results which
352 exhibited other types of land cover changes (Figure 6b), and were incorrectly included in the
353 highest stratum of detected points. This limitation can be overcome by applying context-based
354 anomaly detection on the entire spectrum of detected events, not just the lowest strata.

355

356 Most FN results of the forest fire detection algorithms introduced in this study consist of
357 vegetation types with high natural variability and noisy EVI, which got poorly scored by the
358 proposed algorithms and went undetected (Figure 7a). Other FN results can be attributed to weak
359 characteristics of fire in the EVI time series, with examples shown from California (Figure 8a)
360 and Georgia (Figure 8b) Other FN results in Victoria exhibited atypical characteristics in their
361 EVI time series, which appeared spurious and were not detected by the proposed approach
362 (Figure 7b). In addition, some of the FN results were pruned during the anomaly detection phase

363 due to lack of adequate number of similar (unburned) neighbors required to construct the normal
364 distribution. This poses a challenge in performing context-based anomaly detection when the
365 number of similar objects is small.

366

367

368 **5. Discussion**

369 In this paper, we described and evaluated a framework for forest fire mapping that is
370 unsupervised in nature and has the potential to be used globally, providing spatially explicit wall-
371 to-wall coverage. We quantitatively showed that the algorithm framework performs better than
372 well-known BA framework in the states of California (US), Yukon, (Canada), Victoria
373 (Australia) and much better in Georgia (US). There were also complementarities between the
374 two frameworks. We also showed that the new framework is highly robust to noise in one of its
375 primary inputs, AF, which is known to have low precision.

376

377 The forest fire mapping framework proposed in this paper faces limitations in a number of
378 scenarios, leading to both FP and FN results. These include situations where (1) the vegetation
379 rapidly recovers after a fire or if there are multiple fires in short succession (FN), (2) the loss in
380 vegetation is insignificant (FN), and (3) the vegetation has high natural variability (FP and FN).
381 Each of these scenarios poses distinct challenges for our current fire detection framework.
382 Additionally, if a fire polygon does not contain any pixel with an AF, it will not be detected by
383 our framework (FN). However, we observed that such instances happen only in small polygons,
384 and hence its effect on the performance of the proposed framework is insignificant.

385

386 We are in effect considering less information than the BA framework, yet we are able to
387 achieve comparable performance. Thus, there is reason to believe that some of the limitations
388 above will be addressed by increasing the spatial and temporal resolution to 250-m and daily,
389 respectively. In particular, we expect that smaller fires (250-m data has sixteen times more detail
390 than 1-km resolution data) and pixels that exhibit rapid recovery (because of compositing,
391 neighboring time steps can be up to a month apart in the current 16-day data) can be detected

392 with higher resolution data. We have not used higher resolution data in this paper, since these are
393 not standard MODIS products and hence require extensive processing to generate.

394

395

396 While the current framework already performs relatively well in a variety of geographies,
397 there are a number of interesting directions for future work. The data inputs used in this paper,
398 EVI time series and AF, have temporal resolutions of 16 and 8 days, respectively. Both of these
399 inputs can be computed on a daily basis, which is likely to provide much more information in
400 many cases. This information can be exploited to identify the precise day of the fire and to ensure
401 temporal coherence between neighboring pixels. Challenges with daily data include increase in
402 the noise level and additional effort required to generate a daily EVI (since this is not a standard
403 MODIS product). Another potential extension is to use BA (and similar products) as an *input*,
404 taking advantage of complementarities that exist between the frameworks. Finally, the anomaly
405 scoring that is currently applied to the lowest stratum events can also be extended to the middle
406 and highest stratum to further increase the precision in these strata.

407

408

409 **Acknowledgements.** The authors thank the Planetary Skin Institute and the NASA Carbon
410 Monitoring System program for funding support.

411

412

412
413

414 **References**

415

416 Achard, F., DeFries, H Eva, M Hansen, P Mayaux, and H-J Stibig, 2007, Pan-tropical monitoring of
417 deforestation, *Environ. Res. Lett.* 2 045022, doi: 10.1088/1748-9326/2/4/045022.

418 Badarinath, K. V. S. , A. R. Sharma, and S. K. Kharol. Forest fire monitoring and burnt area
419 mapping using satellite data: a study over the forest region of Kerala state, India.

420 *International Journal of Remote Sensing*, 32(1):85–102, 2011.

421 Bandyopadhyay, S., P. Shyamsundar, and A. Baccini. Forests, biomass use and poverty in
422 Malawi. *Ecological Economics*, 70(12):2461–2471, 2011.

423 Bergeron, Y., S. Gauthier, M. Flannigan, and V. Kafka. Fire regimes at the transition between
424 mixedwood and coniferous boreal forest in northwestern Quebec. *Ecology*, 85(7):1916–
425 1932, 2004.

426 Brewer, C. K., J. C. Winne, R. L. Redmond, D. W. Opitz, and M. V. Mangrich. Classifying and
427 mapping wildfire severity: A comparison of methods. *Photogrammetric Engineering &*
428 *Remote Sensing*, 71(11):1311—1320, 2005.

429 Canadian National Fire Database (CNFDB), 2011, Natural Resources Canada,
430 cwfis.cfs.nrcan.gc.ca/en_CA/nfdb/poly

431 Colditz, R.R., Conrad, C., Wehrmann, T., Schmidt, M. and Dech, S.W., 2007. Analysis of the quality of
432 collection 4 and 5 vegetation index time series from MODIS. ISPRS Spatial Data Quality
433 Symposium, Enschede, The Netherlands, CRC press.

434 DeFries R, Achard F, Brown S, Herold M, Murdiyarso D, Schlamadinger B and de Souza C 2006
435 *Reducing Greenhouse Gas Emissions from Deforestation in Developing Countries: Considerations*
436 *for Monitoring and Measuring* (Rome: Global Terrestrial Observing System) p 23 (GTOS report
437 46).

438 DeFries R, Achard F, Brown S, Herold M, Murdiyarso D, Schlamadinger B and de Souza C 2007 Earth
439 observations for estimating greenhouse gas emissions from deforestation in developing countries
440 *Environ. Sci. Policy* **10** 385–94.

441 Department of Sustainability and Environment, Victoria, Australia, 2011, Fire History Records of
442 Fires on Public Land. <http://www.dse.vic.gov.au>.

443 FRAP (Fire and Resource Assessment Program), 2011, California Department of Forestry and
444 Fire Protection, <http://frap.fire.ca.gov>.

445 Fraser, R. H., Z. Li, and J. Cihlar. Hotspot and NDVI Differencing Synergy (HANDS): a new
446 technique for burned area mapping over boreal forest. *Remote Sensing of Environment*,
447 74(3):362–376, 2000.

448 Friedlingstein, P., R. A. Houghton, G. Marland, J. Hackler, T. A. Boden, T. J. Conway, J. G. Canadell,
449 M. R. Raupach, P. Ciais & C. Le Quéré, 2010, Update on CO2 emissions. *Nature Geoscience*, doi:
450 10.1038/ngeo_1022.

451 Friend, G. R. , M. Leonard, S. Troy, K. G. Tolhurst, and M. Wouters. Fire and biodiversity
452 management in Victoria—integrating the science, planning and implementation process. In
453 *Proceedings of the 3rd International Wildland Fire Conference*, pages 3–6, 2003.

454 George, C., C. Rowland, F. Gerard, and H. Balzter. Retrospective mapping of burnt areas in
455 central Siberia using a modification of the normalised difference water index. *Remote*
456 *Sensing of Environment*, 104(3):346–359, 2006.

457 Giglio, L., G. R. van der Werf, J. T. Randerson, G. J. Collatz, and P. Kasibhatla. Global
458 estimation of burned area using MODIS active fire observations. *Atmospheric Chemistry*
459 *and Physics*, 6(4):957–974, 2006.

460 Giglio, L., T. Loboda, D. Roy, B. Quayle, and C. Justice. An active-fire based burned area
461 mapping algorithm for the MODIS sensor. *Remote Sensing of Environment*, 113(2):408–
462 420, 2009.

463 Gitas, I. Z., G. H. Mitri, and G. Ventura. Object-based image classification for burned area
464 mapping of Creus Cape, Spain, using NOAA-AVHRR imagery. *Remote Sensing of*
465 *Environment*, 92(3):409–413, 2004.

466 Hansen, M.C., Roy, D., Lindquist, E., Justice, C.O., and Altstaad, A., 2008, A method for integrating
467 MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin.
468 *Remote Sens. Environ* 112:2495–2513.

469 Huete, A. R., K. Didan, Y. E. Shimabukuro, P. Ratana, S.R. Saleska, L.R. Hutya, D. Fitzjarrald, W.
470 Yang, R.R. Nemani, and R. Myneni, 2006, Amazon rainforests green-up with sunlight in dry
471 season, *Geophys. Res. Lett.*, 33, L 06405, doi:10.1029/2005GL025583

472 Huete, A., Didan, K., Miura, T., & Rodriguez, E., 2002. Overview of the radiometric and biophysical

473 performance of the MODIS vegetation indices. *Remote Sens. Environ.*, 83, 195 – 213.

474 Justice, C. O., L. Giglio, D. Roy, L. Boschetti, I. Csiszar, D. Davies, S. Korontzi, W. Schroeder,
475 K. O’Neal, and J. Morisette. MODIS-derived global fire products. In B. Ramachandran,
476 C. O. Justice, and M. J. Abrams, editors, *Land Remote Sensing and Global Environmental*
477 *Change*, pages 661–679. Springer, 2011.

478 Lhermitte, S., J. Verbesselt, I. Jonckheere, K. Nackaerts, J. A. van Aardt, W. W. Verstraeten, and
479 P. Coppin. Hierarchical image segmentation based on similarity of NDVI time series.
480 *Remote Sensing of Environment*, 112(2):506–521, 2008.

481 Liu, D. and S. Cai. A spatial-temporal modeling approach to reconstructing land-cover change
482 trajectories from multi-temporal satellite imagery. *Annals of the Association of American*
483 *Geographers*, 2011.

484 Loboda, T., K. O’Neal, and I. Csiszar. Regionally adaptable dNBR-based algorithm for burned
485 area mapping from MODIS data. *Remote Sensing of Environment*, 109(4):429–442, 2007.

486 Lunetta, R. S., J. F. Knight, J. Ediriwickrema, J. G. Lyon, and L. D. Worthy. Land-cover change
487 detection using multi-temporal MODIS NDVI data. *Remote Sensing of Environment*,
488 105(2):142–154, 2006.

489 M. Cochrane. M., *Tropical fire ecology: climate change, land use and ecosystem dynamics*.
490 Springer, 2009.

491 Miles, L. and V. Kapos, 2008, Reducing greenhouse gas emissions from deforestation and forest
492 degradation: Global land-use implications, *Science*, 320 (5882), 1454-1455.

493 Mithal, V. A. Garg, I. Brugere, S. Boriah, V. Kumar, M. Steinbach, C. Potter, and S. A. Klooster.
494 Incorporating natural variation into time series-based land cover change detection. In
495 *Proceedings of the 2011 NASA Conference on Intelligent Data Understanding (CIDU)*,
496 pages 45–59, 2011b.

497 Mithal, V. A. Garg, S. Boriah, M. Steinbach, V. Kumar, C. Potter, S. Klooster, and J. C. Castilla-
498 Rubio. Monitoring global forest cover using data mining. *ACM Transactions on Intelligent*
499 *Systems and Technology*, 2:36:1–36:24, 2011a.

500 NFDSC (National Fire Decision Support Center), 2011, Wildland Fire Decision Support System.
501 National Fire Perimeters. <http://wfdss.usgs.gov>.

- 502 Niklasson M. and A. Granström. Numbers and sizes of fires: Long-term spatially explicit fire
503 history in a Swedish boreal landscape. *Ecology*, 81(6):1484–1499, 2000.
- 504 Olson, D. L. and D. Delen, 2008, *Advanced Data Mining Techniques*, Springer, 1st edition pp.138, [ISBN](#)
505 [3540769161](#)
- 506 Pan, Y., R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips,
507 A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D.
508 McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes. A large and persistent carbon sink
509 in the world’s forests. *Science*, 333(6045):988–993, 2011.
- 510 Potter, C., P. Tan, M. Steinbach, S. Klooster, V. Kumar, R. Myneni , V. Genovese, 2003, Major
511 disturbance events in terrestrial ecosystems detected using global satellite data sets. *Global Change*
512 *Biology*, 9 (7), 1005-1021.
- 513 Potter, C., P. Tan, V. Kumar, C. Kucharik, S. Klooster, V. Genovese, W. Cohen, S. Healey, 2005.
514 Recent history of large-scale ecosystem disturbances in North America derived from the AVHRR
515 satellite record, *Ecosystems*, 8(7), 808.
- 516 Potter, C., V. Kumar, S. Klooster, and R. Nemani, 2007, Recent history of trends in vegetation greenness
517 and large-scale ecosystem disturbances in Eurasia, *Tellus B*, 59, 260-272.
- 518 Roy, D. P. , L. Giglio, J. D. Kendall, and C. O. Justice. Multi-temporal active-fire based burn scar
519 detection algorithm. *International Journal of Remote Sensing*, 20(5):1031–1038, 1999.
- 520 Roy, D. P., P. E. Lewis, and C. O. Justice. Burned area mapping using multi-temporal moderate
521 spatial resolution data—a bi-directional reflectance model-based expectation approach.
522 *Remote Sensing of Environment*, 83(1-2):263–286, 2002.
- 523 Solomon, S. *et al.*, 2007, Technical summary *Climate Change 2007: The Physical Science Basis.*
524 *Contribution of Working Group I the Fourth Assessment Report of the Intergovernmental Panel, on*
525 *Climate Change* ed S Solomon, D Qin, M Manning, Z Chen, M Marquis, K B Averyt, M Tignor and
526 H L Miller (Cambridge, Cambridge University Press).
- 527 Somashekar, R., P. Ravikumar, C. Mohan Kumar, K. Prakash, and B. Nagaraja. Burnt area
528 mapping of Bandipur National Park, India using IRS 1C/1D LISS III data. *Journal of the*
529 *Indian Society of Remote Sensing*, 37:37–50, 2009.

530 Stocks, B. J., J. A. Mason, J. B. Todd, E. M. Bosch, B. M. Wotton, B. D. Amiro, M. D. Flannigan, K. G.
531 Hirsch, K. A. Logan, D. L. Martell, and W. R. Skinner. 2002, Large forest fires in Canada, 1959–
532 1997. *J. Geophys. Res.*, 107

533 Talon, B., S. Payette, L. Filion, and A. Delwaide. Reconstruction of the long-term fire history of
534 an old-growth deciduous forest in Southern Québec, Canada, from charred wood in mineral
535 soils. *Quaternary Research*, 64(1):36–43, 2005.

536 Tan, P.-N., M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison-Wesley, Boston,
537 MA, 2006.

538 United Nations. Collaborative Programme on Reducing Emissions from Deforestation and Forest
539 Degradation in Developing Countries. <http://www.un-redd.org/>.

540 US Geological Survey and NASA. Land Processes Distributed Active Archive Center (LP
541 DAAC). 2011, <https://lpdaac.usgs.gov/>.

542 USDA Forest Service and US Geological Survey, Monitoring Trends in Burn Severity Project,
543 Burned Area Boundaries Dataset. <http://mtbs.gov>.

544
545
546

546

547 Table 1. Regions studied in this paper and their respective sources of historical wildfire
548 evaluation data. The total number of 1-km MODIS pixels center locations falling entirely inside
549 (positives) or entirely outside (negatives) wildfire polygon perimeters are included.

550

Region	References	Positives	Negatives
California (US)	FRAP (2011)	4597	597208
Georgia (US)	NFDSC (2011) and USDA (2011)	2003	425528
Yukon (Canada)	Canadian National Fire Database (2011)	5597	697208
Victoria (Australia)	Friend et al. (2003) and Department of Sustainability and Environment, Victoria, (2011)	17190	604391

551

552

553 Table 2. Fire Polygon Evaluation and Confusion Matrix.

554

Evaluation Data	Predicted	
	Fire	No Fire
	Fire	<i>TP</i> <i>FN</i>
No Fire	<i>FP</i> <i>TN</i>	

555

556

556

557 Table 3. Algorithm evaluation results for (a) California (US), (b) Yukon (Canada), (c) Georgia
558 (US), and (d) Victoria (Australia).

559

(a) California (US)

Algorithm	Precision	Recall	TP	FP	FN
Level 1	0.989	0.797	1765	20	449
Level 2	0.982	0.842	1864	34	350
Level 3	0.909	0.936	2073	208	141
BA highQ	0.982	0.925	2047	38	167
BA lowQ	0.978	0.925	2047	46	167
AF	0.463	0.890	1970	2286	244
V2DELTA	0.605	0.717	1588	1035	626

560

561

(b) Yukon (Canada)

Algorithm	Precision	Recall	TP	FP	FN
Level 1	0.999	0.289	975	1	2401
Level 2	0.988	0.421	1422	17	1954
Level 3	0.983	0.975	3293	58	83
BA highQ	0.897	0.778	2626	303	750
BA lowQ	0.912	0.957	3231	313	145
AF	0.801	0.663	2237	556	1139
V2DELTA	0.522	0.868	2930	2679	446

562

563

(c) Georgia (US)

Algorithm	Precision	Recall	TP	FP	FN
Level 1	0.982	0.475	951	17	1052
Level 2	0.978	0.600	1201	27	802
Level 3	0.965	0.706	1414	51	589
BA highQ	0.468	0.193	386	438	1617
BA lowQ	0.699	0.674	1350	581	654
AF	0.152	0.645	1291	7197	712
V2DELTA	0.349	0.464	930	1738	1073

564

565

(d) Victoria (Australia)

Algorithm	Precision	Recall	TP	FP	FN
Level 1	0.997	0.420	7192	21	9939
Level 2	0.995	0.564	9654	51	7477
Level 3	0.950	0.657	11257	594	5874
BA highQ	0.993	0.692	11850	83	5282
BA lowQ	0.993	0.696	11919	88	5212
AF	0.854	0.532	9118	1554	8013
V2DELTA	0.907	0.555	9510	975	7621

566

567

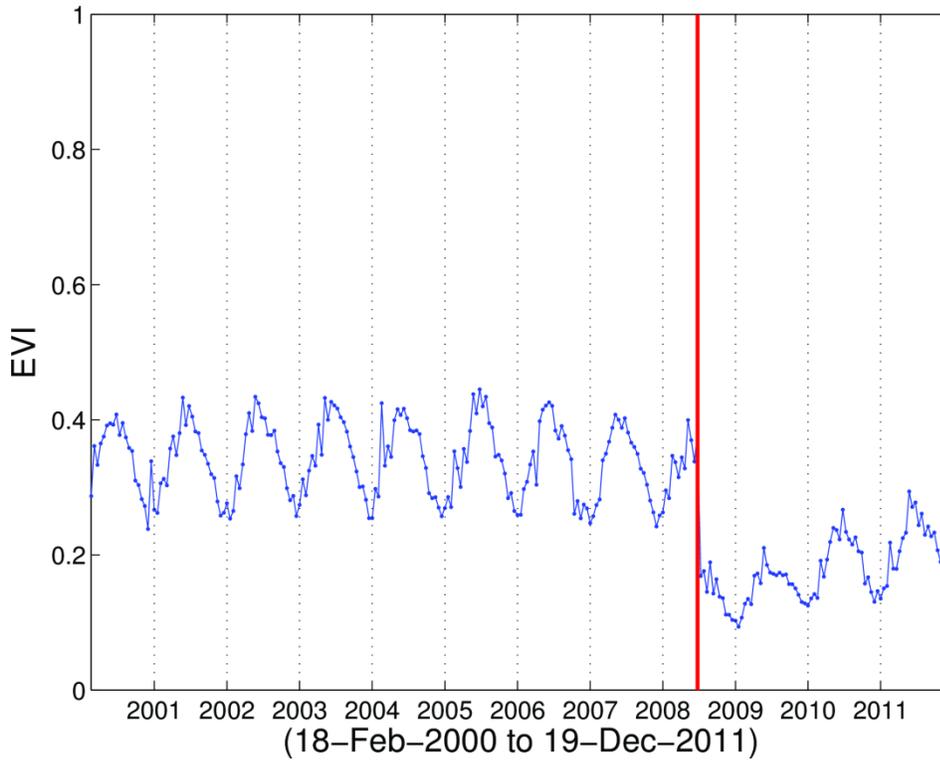
568

569

570

571 Figure 1. MODIS EVI time series at the location of the Basin Complex fire, which was started
572 by lightning near Big Sur, CA in June 2008.

573

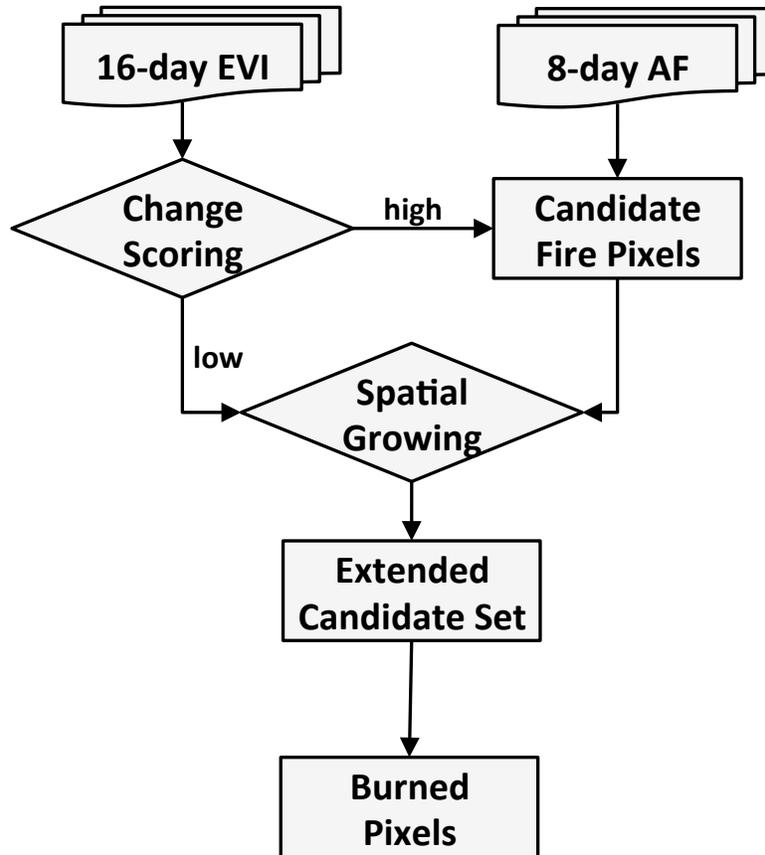


574
575

575

576 Figure 2. Flowchart illustrating the new framework for mapping forest fires.

577



578

579

580

581

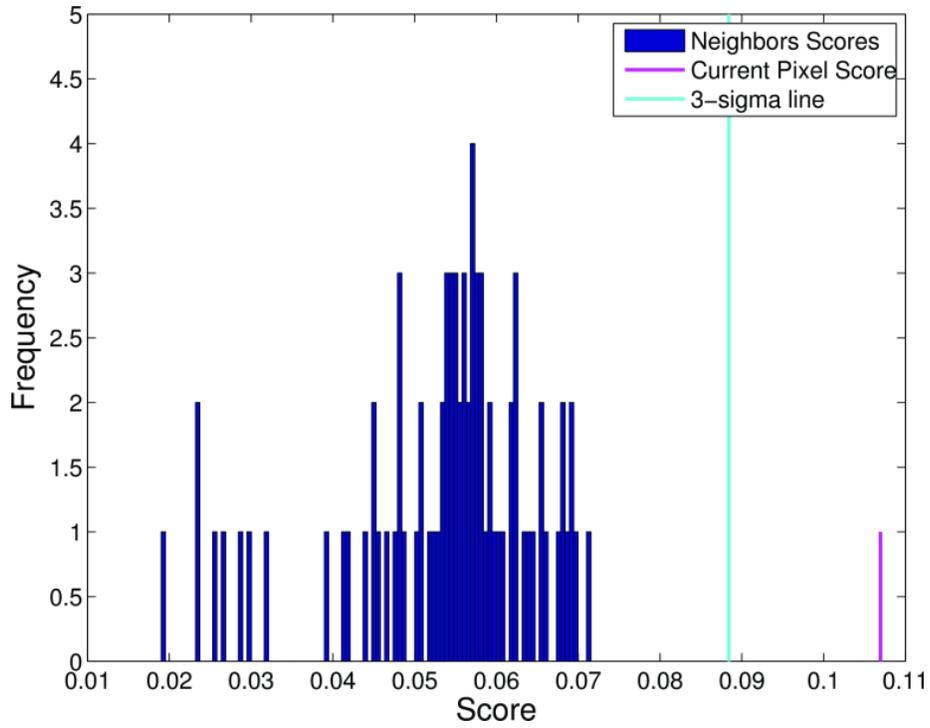
582

582

583

584 Figure 3. Example of context-based anomaly detection, based on the distribution of the ND
585 scores of a fire candidate location and its neighbors.

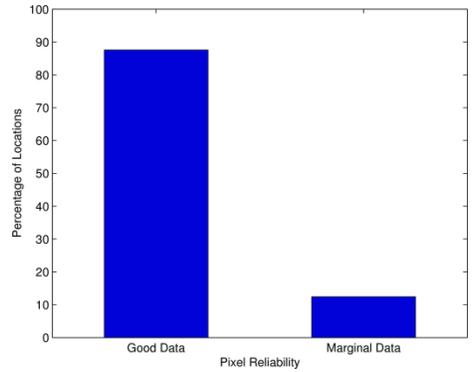
586



587

588

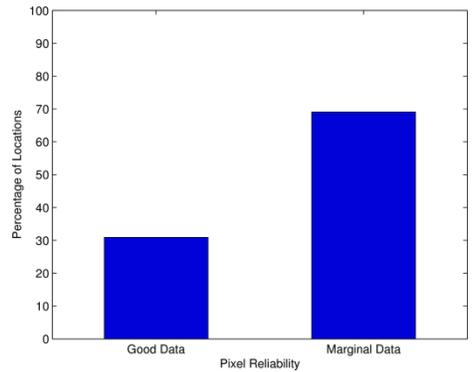
588 Figure 4. A comparison of histograms for pixel reliability of the BA algorithm and sensitivity to
589 MODIS data quality. The top histogram shows TPs of the high-quality MODIS pixels, and the
590 bottom histogram shows the FNs of the MODIS high-quality pixels.
591



592

593

TP in BA



594

595

FN in BA

596

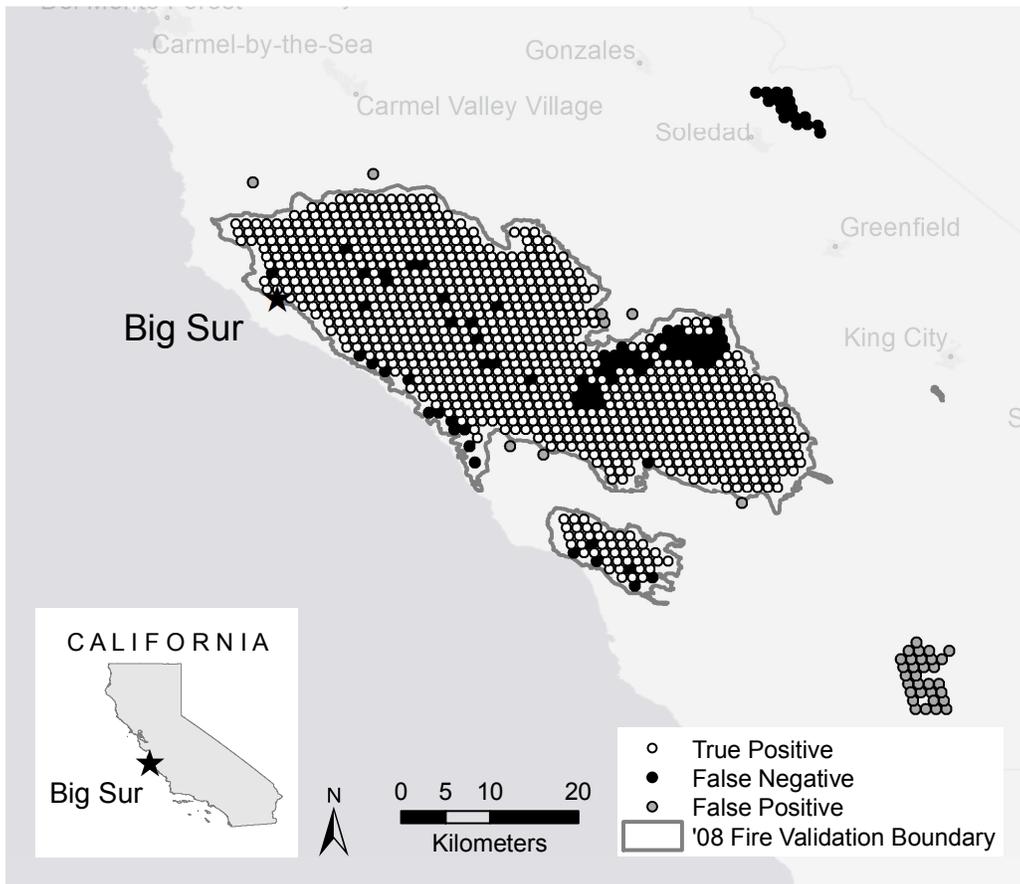
597

598

598
599
600
601
602
603
604
605
606

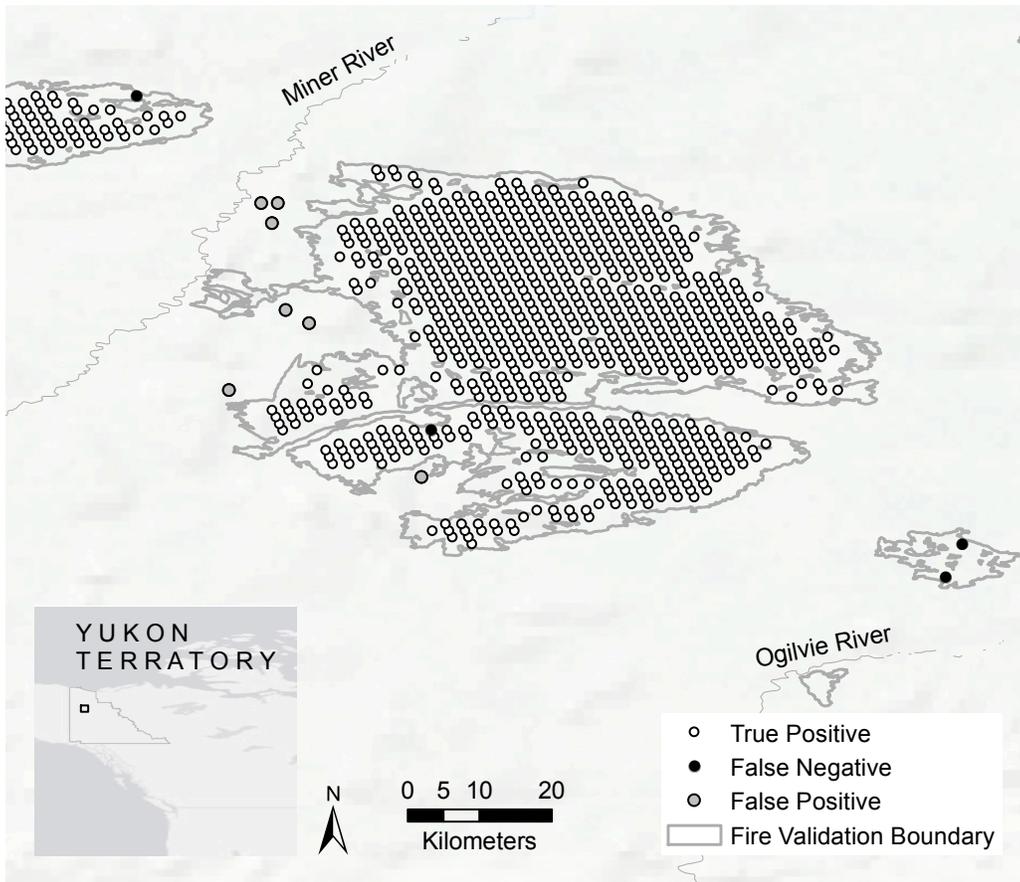
Figure 5. Map of fire polygon perimeters and L3 algorithm results (Table 3) in (a) California (US) and (b) Yukon Province (Canada).

a.



607
608
609

609 b.



610

611

612

613

614

615

616

617

618

619

620

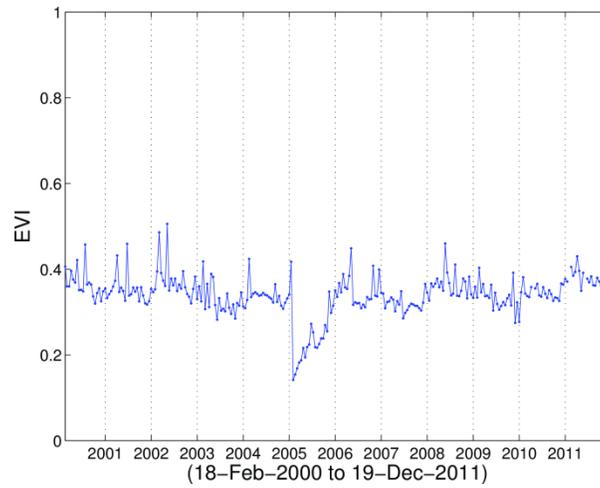
621

622

623

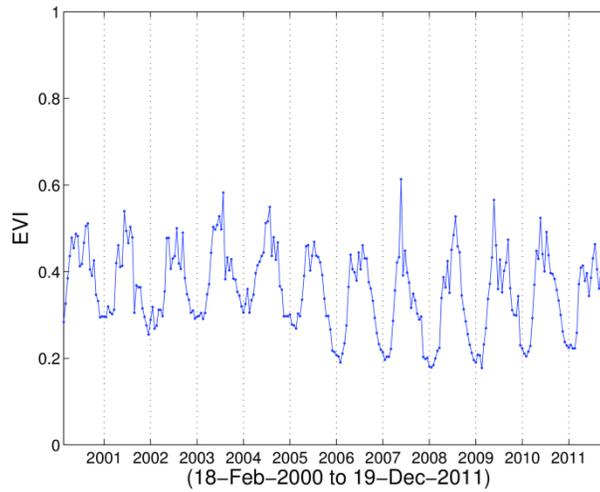
624
625
626
627

Figure 6. False positives (FP) results from the new forest fire detection approach.



628
629

a. Missed fire event in Victoria, Australia



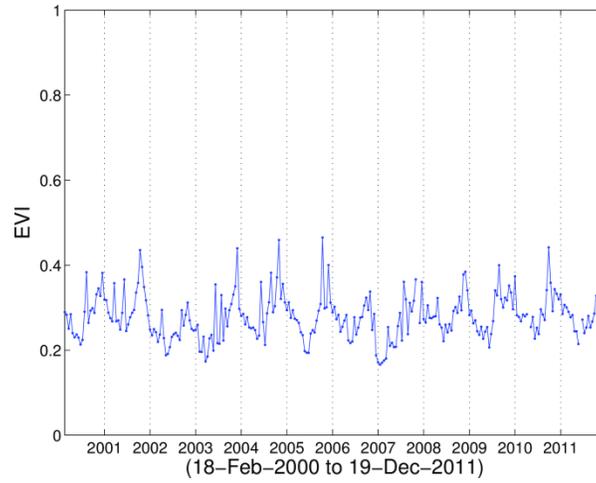
630
631

b. Gradual land cover change in Georgia (US)

632
633
634
635

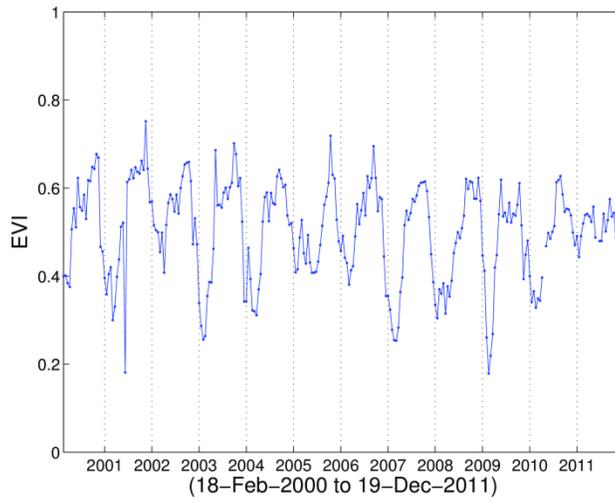
636 Figure 7. False negatives (FN) results in Victoria, Australia

637



638

639 a. Vegetation with high variability (event occurred in 2007)



640

641 b. Atypical characteristic of fire (event occurred in 2009)

642

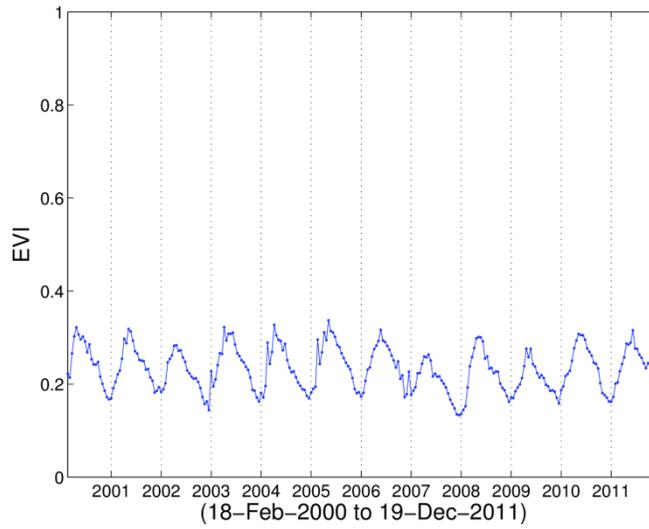
643

643

644 Figure 8. False negatives (FN) results due to weak fire characteristics

645

646

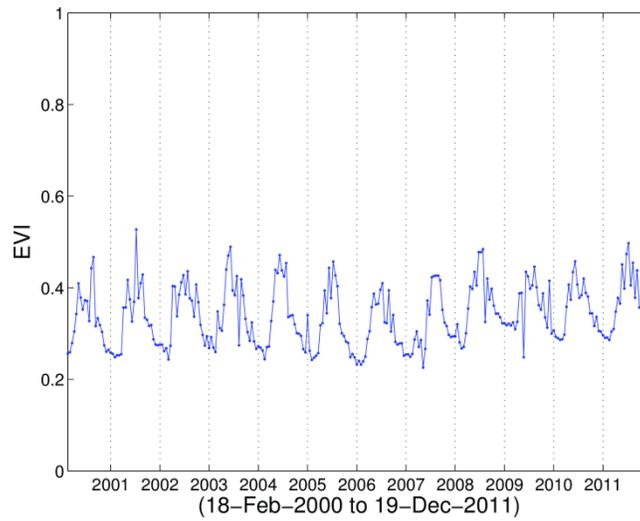


647

648 FN in California (event occurred in 2008)

648

649



650

651 FN in Georgia (event occurred in 2007)

651

652