Remote Sensing of Environment 127 (2012) 194–209

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

On timeliness and accuracy of wildfire detection by the GOES WF-ABBA algorithm over California during the 2006 fire season

Alexander Koltunov [a,*], Susan L. Ustin [a], Elaine M. Prins [b]

[a] Center for Spatial Technologies and Remote Sensing, Department of Land, Air, and Water Resources, University of California, Davis, Veihmeyer Hall, One Shields Avenue, Davis, CA 95616, USA
[b] University of Wisconsin-Madison Space Science and Engineering Center (SSEC), Cooperative Institute for Meteorological Satellite Studies (CIMSS)—Consultant, 1225 West Dayton Street, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 21 June 2011
Received in revised form 29 August 2012
Accepted 7 September 2012
Available online xxxx

Keywords:
Geostationary wildfire detection
WF-ABBA
Validation
Early wildfire detection
Detection timeliness
Remote sensing
GOES Landsat TM/ETM+
Burn scar detection

ABSTRACT

The Wildfire Automated Biomass Burning Algorithm (WF-ABBA) is a state-of-the-art algorithm for geostationary wildfire detection whose results have been increasingly used in a range of environmental applications. At present, the WF-ABBA validation activities and, in general, fire product validation methodologies are at a markedly less advanced stage than the algorithm itself. Particularly, little is known about detection timeliness, despite the value of such information for assessing the potential of geostationary observations to improve tactical decision making of first responders. This paper contributes to reducing this gap in two ways. Firstly, we describe a new methodology that is suitable for evaluating geostationary satellite wildfire detection in terms of incidents with regard to both timeliness and reliability. This methodology utilizes available official multi-agency wildfire reporting information and multitemporal Landsat imagery. Secondly, we apply the proposed validation method to temporally filtered GOES-West WF-ABBA (ver. 6.1) detections for the 2006 fire season over the State of California and present incident-wise and pixel-wise performance information. The results indicate highly reliable pixel-wise performance of WF-ABBA, with about 75% of fire pixels (or more) corresponding to actual recorded active wildfires. A substantial portion of wildfires were detected during their first hour of activity, and a few incidents—even before the initial reports from conventional sources. Although the WF-ABBA performs best at what it was designed for: consistently re-detecting (monitoring) active fires, we believe there is an additional potential for automated detection from current geostationary data to reduce wildfire ignition latencies in the Western U.S. Our results can serve as a guideline for algorithm developers and users of the WF-ABBA fire product.

1. Introduction and background

Although the Geostationary Operational Environmental Satellite (GOES) series have been used to monitor fire activity in the Western Hemisphere for more than two decades, the GOES-East Automated Biomass Burning Algorithm (ABBA) was first routinely applied in South America in 1995 to monitor spatial, diurnal and interannual trends in fire activity associated with deforestation and agricultural practices. In the year 2000 the modified GOES-E/-W WF-ABBA was implemented to provide near real time diurnal fire monitoring throughout the Western Hemisphere with increased value for operational agencies and scientific communities. Confident and informed use of the satellite detections, however, necessitates something that has been often offset to the margins of the product development budgets—algorithm validation efforts over a broad range of environmental conditions and geographic regions and with respect to appropriate performance measures. Furthermore, development of product validation methodologies is still far from complete. The objective difficulty of fire detection from geostationary platforms implies that satellite detections should not be considered the only source of information about wildfires. To this end, diverse and complete performance information is also necessary to find the appropriate place for satellite-derived fire products in diverse user toolboxes.

1.1. Active fire detection by the GOES WF-ABBA algorithm (Version 6.1)

The GOES-E/-W WF-ABBA processing system was developed as a collaborative effort between the National Oceanic and Atmospheric Administration (NOAA) Center for Satellite Application and Research (formerly, Office of Research and Applications) and the University of Wisconsin-Madison Cooperative Institute for Meteorological Satellite Studies (CIMSS). The WF-ABBA is a dynamic multispectral thresholding contextual algorithm that uses the visible (when available), 3.9 μm, and 10.7 μm infrared bands to locate and characterize hot spot pixels. When available, the 12 μm band is used in conjunction with the other bands to assist in identifying opaque clouds. The WF-ABBA algorithm is based on the sensitivity of the 3.9 μm band to high temperature subpixel

* Corresponding author. Tel.: +1 530 752 5092; fax: +1 530 752 5262.
E-mail address: akoltunov@ucdavis.edu (A. Koltunov).

0034-4257/$ – see front matter
http://dx.doi.org/10.1016/j.rse.2012.09.001

anomalies and is derived from a technique originally developed by Matson and Dozier (1981) for NOAA Advanced Very High Resolution Radiometer (AVHRR) data. Once the WF-ABBA locates a hot spot pixel, it incorporates ancillary data to reduce false positives and correct for water vapor attenuation, surface emissivity, solar reflectivity, diffraction, and semi-transparent clouds. The AVHRR derived Global Land Cover Characteristics (GLCC) data base (version 2.0) is used to assign surface emissivity and to screen for false positives. The National Centers for Environmental Prediction (NCEP) Aviation model total column precipitable water products are utilized to correct for water vapor attenuation. Numerical techniques are used to determine instantaneous estimates of subpixel fire size and average temperature.

The GOES Imager oversamples the infrared band instantaneous ground field of views (IGFOV) along a scan line by a factor of 1.75. The impulse response is smeared over three samples and can result in 3 fire detections for the same fire pixel. The WF-ABBA evaluates adjacent fire pixels along a scan line and filters out multiple detections of the same fire. If multiple detections occur, fire pixels are filtered according to confidence classification with the 3.9 μm brightness temperature used in case of a tie (highest temperature fire pixel is retained). An in-line temporal filter is applied to reduce false positives associated with noise in the imagery and cloud edge issues. The temporal filtering technique uses a time series of GOES fire products from previous hours to compare with the current fire product. A fire pixel must appear at least twice (within 0.1°) within the past 12 h in order to be retained in the final filtered fire product. The buffer of 0.1° compensates for geolocation inaccuracies of GOES pixels, which may reach several kilometers, due to satellite navigation errors. The filtered fire product can result in delayed identification of a fire start time and may eliminate short-lived agricultural management fires. For more information on the algorithm and the determination of subpixel fire characteristics, refer to Prins and Menzel (1992, 1994) and Prins et al. (1998, 2001, 2003).

Since the year 2000 the WF-ABBA has been executed at a minimum of every half-hour for both GOES-E and GOES-W, detecting fires within a satellite zenith angle of 80° (covering the better part of the visible hemisphere). For over 10 years both temporally filtered and non-filtered fire product ASCII files have been made available to the user community via anonymous ftp at UW-Madison CIMSS and the U.S. Navy's Fire Locating and Modeling of Burning Emissions (FLAMBE) project web site (http://www.nrlmry.navy.mil/flambe/, decommissioned on October 31, 2011). For each processed GOES *frame* (i.e. one of the many multispectral images composing the GOES image sequence), the WF-ABBA (Version 6.1) fire product ASCII files contain information about the detected fire pixels including location, observed 3.9 and 10.7 μm brightness temperatures, estimates of instantaneous sub-pixel fire size and temperature, ecosystem type, and fire confidence flag. The fire confidence flags consist of processed ("0"), saturated ("1"), cloudy ("2"), high ("3"), medium ("4"), and low possibility ("5") fire pixels. The latter category represents the largest number of false detections as it has the least stringent requirements for fire identification and were not used in this study.

1.1.1. WF-ABBA intended and extended use

The South American ABBA and next generation WF-ABBA algorithm pioneered fire monitoring from geostationary platforms. To date, the GOES WF-ABBA fire product has been successfully utilized by a broad user community, including applications in fire weather analysis and forecasting, climate change research, land-use/land-cover change studies, resource management, biomass burning emissions modeling, diagnostic and prognostic aerosol and trace gas modeling, and policy and decision making (Brioude et al., 2009; Cardoso et al., 2003; Freitas et al., 2007; McNamara et al., 2004; Nepstad et al., 2001, 2006; Schmidt & Prins, 2003; Wang et al., 2006; Weaver et al., 2004). Applications of the GOES WF-ABBA in model data assimilation studies have demonstrated the utility of integrating diurnal fire products in aerosol transport and air quality models to locate and predict air pollution (Al-Saadi et al., 2005; Freitas et al., 2007; Longo et al., 2010; Reid et al., 2009, 2004; Wang, et al., 2006). WF-ABBA detections and subsequent flux estimates have been assimilated under the FLAMBE project into the Naval Research Laboratory Aerosol Analysis and Prediction System (NAAPS) aerosol transport forecast model since the year 2000 (http://www.nrlmry.navy.mil/aerosol/). Version 6.1 of the WF-ABBA software was transitioned to the NOAA National Environmental Satellite Data and Information Service (NESDIS) operations in 2002 and incorporated into the Hazard Mapping System (McNamara et al., 2004).

Another potentially attractive application of a fire detection algorithm is to assist in earlier identification of ignitions, with the ultimate objective to minimize the time to initial detection. Earlier identification of wildfires increases situational awareness and decision making confidence of fire management agencies and may improve their response times and resource allocation. It is important to emphasize that the WF-ABBA was not specifically designed as an early warning tool, but for monitoring active fires. However, with a steep increase in the wildfire activity in the continental U.S. in recent decades (NICC, 2007) and unsustainable costs of managing wildfires to operational agencies (USFS, 2007), one should explore every opportunity to maximize utilization of currently available operational satellite fire detection assets. Furthermore, individual case studies (e.g. Feltz et al., 2003; Weaver et al., 2004) suggested the potential for geostationary satellites to provide early alarms about new ignitions. Given that MODIS (Moderate Resolution Imaging Spectroradiometer) (Giglio, 2010; Giglio et al., 2003) and next generation VIIRS (Visible Infrared Imager and Radiometer Suite (VIIRS) fire products from polar-orbiting satellites are available only twice a day in mid-latitudes, there is no alternative to using GOES for routinely providing much needed high temporal fire information at consistently low cost during the first hours of fire activity.

1.2. WF-ABBA product validation status and methodologies

Assessment of the quality of satellite wildfire detection products depends on numerous factors, including the study period and region, which determine the fire regime, satellite observation characteristics, dominant land cover type, climate, etc. Therefore, fire products must be validated for different geographic regions and seasons.

1.2.1. Performance measures

Whether a detection should be deemed true or false is defined by an important parameter: the maximal acceptable geolocation error, ε_g. It is set by the end user of a fire product regardless of actual (and never known) geolocation error of the fire product. For example, a hypothetical user for which geolocation error of more than 10 meters renders a fire detection useless should consider nearly every geostationary detection as a false detection. In practice, however, geostationary detections mislocated by several kilometers are still very valuable for many applications mentioned in Section 1.1.1. In general, a validation study which presents algorithm performance information for different choices of ε_g addresses needs of a broader spectrum of potential users.

Furthermore, the performance of a detection algorithm is a combination of two mutually related metrics: the number of true positive objects and the number of false positive objects. Neither one is fully informative without the other. It is also important to remember that the objects used in the above two dual metrics must be the same objects, unless the application explicitly defines customized detection quality statistics. Users or applications requiring maximization of the number of true positive pixels are also interested in minimizing the number of false positive pixels. In turn, when the number of detected actual wildfire incidents is of interest, this information should be accompanied by the information about falsely detected incidents (not merely pixels). For example, when detection output is used for

early warning, false positive fire pixels may have very different costs to a user and therefore should not equally contribute to the overall quality statistic. For example, detections identified in locations where confirmed false positives have been recently observed may be assigned a low priority for verification or simply ignored at low risk.

To date, performance of WF-ABBA has been primarily assessed with respect to the following conventional metrics:

- pixel-wise true positive and false positive rates or, equivalently, pixel-wise omission and commission errors (Feltz et al., 2003; Schroeder et al., 2008a, 2008b; Soja et al., 2009);
- percent of detected actual wildfire incidents (Feltz et al., 2003).

We could not find any published information on the number of false positive incidents produced by any fire detection algorithm. One of the objective reasons is that current operational fire detection algorithms do not output possible fire incidents, only fire pixels. Therefore, additional non-trivial analysis is needed to aggregate fire pixels across time and space into incident-candidates. As part of the overall focus to address the timeliness and fidelity of WF-ABBA wildfire detection in California, this paper presents a method (Section 3.3) for extracting incident-candidates and interval estimation of the number of false positive incidents.

Also, it is important to be clear about the actual meaning of words "detected fire". For a given validation study, this meaning is a direct implication of the validation procedure used. In the cited studies, the statement "a fire [incident] is detected" meant: "a fire incident is detected eventually, i.e. at some time during the burning period (plus margins)". The percentage of fires eventually detected from space is helpful for assessing, for example, fire geographic and seasonal distributions and their dynamics (Prins & Menzel, 1994; Prins et al., 2003). Although there are uncertainties regarding the periods of fire activity and total emissions, satellite derived fire products offer improvements over traditional emissions data bases and have been used for emissions modeling and in diagnostic and prognostic aerosol and trace gas modeling efforts for over 10 years (Al-Saadi et al., 2005; Freitas et al., 2007; Longo et al., 2010; Reid et al., 2009, 2004; Wang et al., 2006; Zhang et al., 2008).

Quite obviously, these measures do not allow one to reasonably assess how soon after ignition or initial report the fires are detected, which is the central focus of this paper. Besides interesting but occasional case studies (Feltz et al., 2003; Prins et al., 2003; Weaver et al., 2004), the vacuum of information on this aspect of WF-ABBA performance was somewhat addressed by Feltz et al., 2003. They have analyzed a large number of recorded wildfire incidents that occurred in Acre (Brazil) during three summer months of 2002 and found that the unfiltered WF-ABBA identified 58 of 88 incidents within 6 h of activity. Temporally filtered WF-ABBA (Section 1.1) detected 76 incidents within a window ± 1 day. The number of false positive incidents was not provided.

1.2.2. Truth data sources

Most intensive validation activities of the WF-ABBA have focused on matching its detections against active fires observed in coincident higher resolution remote sensing imagery, e.g. ASTER and ETM+ (Schroeder et al., 2008a, 2008b). This approach is suitable for estimating the probability that a GOES fire pixel is a true positive and has been used primarily due to the lack of consistent and complete ground truth reporting procedures and documentation. Cross-sensor matching is not informative to assess fire detection timeliness, because the most critical information—the actual fire start or initial report times—is not available from infrequent high resolution imagery. Another limitation of this method is that only a tiny fraction of detected fire pixels can be analyzed. For example, Schroeder et al. (2008b) processed as many as 115 ASTER scenes acquired across the entire continental U.S. during four years (2003 to 2006), but were able to verify only 103 WF-ABBA fire pixels (not incidents). In general, it is not always obvious how well the available high-resolution images represent the diversity of environmental and other conditions (Section 1.2) influencing fire detection (Csiszar et al., 2006).

Detections could also be given to a human analyst to assess. The value of this validation method is not yet obvious, because the procedures, rules, and auxiliary information sources that analysts use to accept or reject detections have not been clearly defined and are not complete (Schroeder et al., 2008b).

Comparison with official fire reporting information is a validation method that is truly independent of the satellite-based fire detection, although it has limitations and is dependent on the reporting methodologies, diligence, and completeness of the reporting agencies. Using reasonably consistent geo-databases of fire ignition reports provides a large sample for analysis and is necessary to estimate timeliness of incident detection, because many records include incident initial report times. However, fire information is often divided between different databases, reports, and other sources, with differences in reporting protocol and entails considerable effort to combine this information (Soja et al., 2009). Also, the geo-databases may be lacking information about certain types of fires (e.g. urban or agricultural fires, controlled burns), and possibly other omitted fires (Schroeder et al., 2008b). Therefore, performance estimates obtained from matching satellite detections to geodatabases apply only to those types of incidents that are actually represented in the databases, e.g. wildland fires, whereas the other types of detected ignitions, such as managed agricultural burns, are regarded as false positives. Furthermore, the geo-databases sometimes omit incidents of type "wildfire", which is the type of incidents addressed in this paper. In this case, the obtained performance statistics are actually lower bounds on method quality. Particularly, the false positive rates are overestimated significantly, when the validation databases miss a large fraction of wildfire incidents of substantial duration, which are also not under cloud cover, and whose intensity of burning is high enough to produce a large magnitude thermal anomaly at satellite pixel scales (~15 km^2 for GOES Imager over California). How likely this combination of circumstances is, depends in particular on a specific validation database. Additional issues include uncertainties about fires, such as actual end times, especially for large incidents. For example, a false detection may be counted as correct when the incident actually ended earlier than the fire reporting information indicates. Previous studies showed, however, that the opposite situation is more common: a fire is still burning after the recorded end time (sometimes this is the containment time, other times—control time). Addressing this problem, the recorded end time is extended for validation purposes, e.g. by 5 days (Soja et al., 2009) or by 1 day (Feltz et al., 2003).

It is clear that each method of fire product validation has its own advantages and limitations depending on the intended detection performance measures, and that validation activities should include both comparisons with higher resolution satellite or airborne data and fire reporting information. This is the methodology proposed in this paper (Section 3). We also emphasize again, that using official fire reporting information is necessary for statistically assessing timeliness of incident detection.

1.3. Study objectives

Motivated by substantial gaps in the information on wildfire detection quality with respect to spatial and temporal coverage and performance metrics, this study is aimed at the following objectives:

1. Propose a validation methodology in which fire geodatabases and high-resolution imagery are combined to compute detection performance measures that are appropriate for early warning applications.

2. Assess the timeliness and reliability of detection of wildland fire incidents for the State of California during the 2006 fire season.
3. Validate geostationary satellite WF-ABBA detections of wildfires over the State of California during the 2006 fire season with respect to pixel-wise metrics.

The first two objectives will contribute to addressing the question: What is the potential of current operational geostationary fire detection utilizing the GOES Imager to contribute to reducing latency of fire detections? Understanding the timeliness of detected fire incidents during periods of substantial wildfire activity is the first step necessary to assess the potential and increase the informed use of geostationary fire detection, determine the research needs and suggest ways of improvements in algorithmic and data acquisition/delivery aspects. We hope that the developments, results, and discussion presented below in this paper, will be valuable not only for the fire scientific community, but also for operational fire management agencies who recognize that timely information about ignitions facilitates rapid and prioritized response, leading to reduction of societal losses due to wildfires.

2. Datasets

2.1. Study area, test period, and WF-ABBA detections

Our study area, the entire State of California (Fig. 1), represents a broad range of ecosystems, ranging from semiarid shrublands, conifer dominated forests, annual grasslands, intensive agriculture, to wetland ecosystems (in order of approximate water availability). This diversity, small-scale mosaic of ecosystems, and significant elevation differences often result in high thermal contrasts, preventing actual fires from standing out from the natural background and causing some background pixels to appear as hot spots in the shortwave thermal infrared (TIR—3.9 μm) imagery. In addition, fog and low-altitude reflective clouds, regularly appearing, especially in Pacific coastal and Sierra Nevada mountain areas shortly after sunrise, can present spectral characteristics closely resembling those of wildfires. All these circumstances create substantial challenges to the multispectral fire detection approach used by the WF-ABBA. Finally, the study area is one of the most densely populated regions in the continental U.S. In the mobile phone era, California fires are normally reported shortly after they start. Therefore, only for a small fraction of wildfires can satellite observation provide the earliest alarm for this region.

Our analysis included the GOES WF-ABBA (version 6.1) temporally filtered fire product (confidence flags 0–4, excluding low possibility fire pixels) available from FLAMBE. We utilized the WF-ABBA processed data for 8486 half-hourly frames acquired by GOES-West during the *test period*: April 30, 2006 through October 31, 2006. Fig. 2 shows fragments of sample GOES-West visible and TIR 3.9 μm band images with a clearly observable "Day" fire—one of the largest wildfire incidents in California history. The ground pixel size for the study area varies from pixel to pixel up to 6 km in the north–south direction for the 3.9 μm, 10.7 μm, and 12 μm thermal bands.

3. Validation methodology

3.0. Outline and performance metrics

In the proposed methodology, WF-ABBA fire pixels are matched in space and time (Section 3.2) to records in geo-spatial databases of official wildfire incident reports (Section 3.1) and aggregated into incident-candidates that we term *events* (Section 3.3). Furthermore, the unmatched events are additionally cross-checked against new burn scars in Landsat imagery. The resulting information about timeliness and accuracy of detecting wildfire incidents (not pixels) is presented using the following metrics:

- relative detection latency of incidents with respect to their initial report times;
- estimated lower bound on the number of correctly detected wildfire incidents; and

Fig. 1. Study area and wildfires in California during the 2006 wildfire season. Study area and the 310 California wildfires (white polygons approximating burned areas) reported during the test period: April 30, 2006 through October 31, 2006.

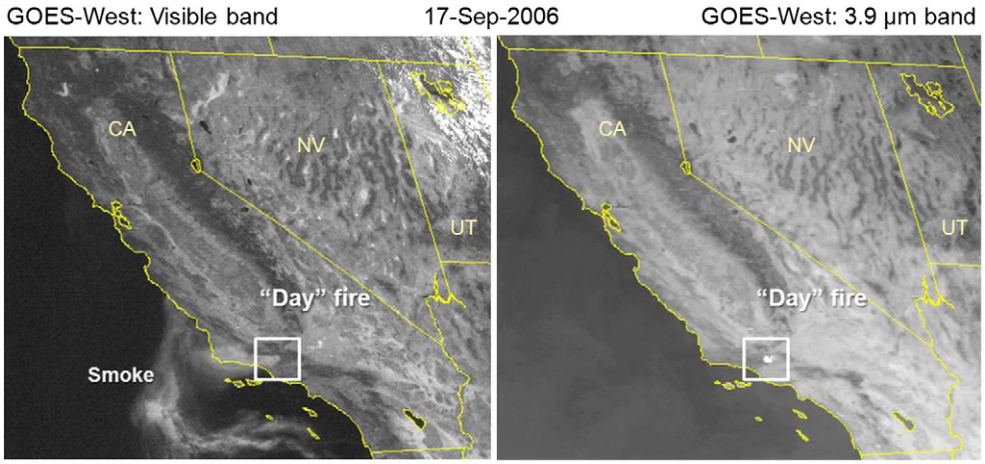

Fig. 2. GOES-West Image example. Examples of GOES-West visible and thermal infrared (3.9 μm) intensity images that show the study area—the State of California. GOES ground pixel size is variable (up to 6 km in north-west direction in the thermal image). The Day Fire shown in the box inset was the largest California wildfire incident in 2006. The Day Fire escaped out of control and over two months burned nearly 66,000 ha, with the estimated costs exceeding $75,000,000.

– estimated lower and upper bounds on the number of events that are not wildfire incidents.

The pixel-wise performance metrics that our method computes include:

- an estimated lower bound on the number of correctly detected wildfire pixels; and
- an estimated upper bound on the number of detected pixels that are not wildfires.

3.1. Wildfire geospatial records and test samples

We merged two separate multiagency wildfire databases for the 2006 wildfire season in California: a fire perimeter polygon database compiled by the California Department of Forestry and Fire Protection (CAL FIRE) and point and polygon databases created by the Geospatial Multi-Agency Coordination (GeoMAC) group. Both databases are based on the individual incident report data from multiple U.S. agencies, including CAL FIRE, US Forest Service, the Bureau of Land Management, National Park Service, and other federal and local agencies and departments. The incidents in these databases are limited to wildland fires, i.e. they do not include agricultural or urban fires. For fires represented only by geolocated points in the GeoMAC database we created pseudo-perimeter polygons as circular buffers with the area equal to the recorded final fire size which is the cumulatively burned area. In the course of merging the databases, we carefully eliminated instances of duplicate records of the same incident (e.g. different spellings, agency coding conventions), thereby making the commission error negligible and ensuring that the merged database is biased toward incident omission. This step was important for constructing intervals for detection performance, as will be further discussed in this paper.

3.1.1. Selection of test samples

The National Interagency Coordination Center (NICC) defines significant fires primarily as those exceeding 40 ha in timber fuel types and 120 ha in grass and brush fuel types (details are available in ILWDP, 2008). We limited our analysis to wildfires with final size exceeding 2 ha, assuming that it is not reasonable to expect detection of wildfires from geostationary satellites that did not burn more than 2 ha over their lifetime (*cf.* Hawbaker et al., 2008 who used a threshold of 18 ha). In the following, these fires are denoted by f_k and their respective final size polygons are denoted by p_k, where k is an enumerating subscript.

According to the merged geodatabase, 402 fires were active during the test period, of which 354 fires were initially reported during the test period. Which of these fires should be included in the sample to test detection performance? The correct answer depends on what performance measure is being evaluated. In the case of pixel-wise metrics (Section 3.0), all 402 wildfires that were active during the test period can be included. In contrast, the incident-wise metrics require a more delicate approach. Indeed, two or more wildfires may overlap in time and also be geographically close enough to cause ambiguity in WF-ABBA detection attribution, if all the co-occurring incidents are included in the sample. Therefore, we analyzed incident co-occurrence and excluded some of the co-occurring incidents. The deatiled description of this procedure is given below in Section 3.2.2 after we define the rules for determining fire activity periods in Section 3.2.1. As a result, we have selected 310 of the 354 fires reported during the test period. Furthermore, of these 310 non-overlapping fires, the initial report time (in addition to the date) was available only for 164 fires. Obviously, only these 164 fires can be used for measuring detection timeliness with a precision of minutes. Thus, the two test sub-samples selected for assessing incident detection timeliness included:

1. 164 fires reported during the test period and for which the initial report hour was recorded, and
2. 310 fires (including the above 164 fires) that were reported during the test period; these are shown in Fig. 1.

Fires in these samples widely vary by their final size (Fig. 3A) and activity period durations (Fig. 3B). In Fig. 3, the abscissa axes are binned into seven intervals. The burned area intervals are: <5 ha, 5–20 ha, 20–100 ha, 100–500 ha, 500–1000 ha, 1000–5000 ha, and >5000 ha. The incident duration intervals (Fig. 3B) are: <6 h, 6–12 h, 12–24 h, 1–2 days, 2–7 days, 7–30 days, and >30 days. Furthermore, the 164 wildfires with a recorded report hour do not form a completely random subset of the 310 fires. Indeed, as can be seen in Fig. 3A, incidents with a greater final burned area are also more likely to have a record about the initial report hour. Because larger incidents are generally easier to detect from space, the sample with 310 fires yields better results for some statistics (e.g. the fractions of never detected fires or fires detected during the first 24 h of activity). Fig. 4 displays the joint distribution for recorded time of the day for initial reports versus the fire final size for the 164 wildfire

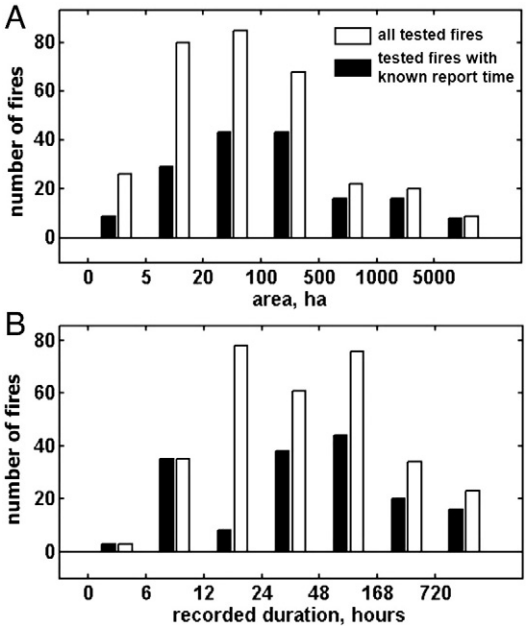

Fig. 3. Final size and recorded incident duration distributions for tested fires. Histograms of the final size distribution (A) and recorded incident duration distribution (B) for all 310 tested active fires and for 164 active fires with known initial report time. Both fire area and duration are binned into 7 groups each (see Section 3.1.1). The bar locations (or widths) on the area and duration axes represent group membership and not a within-group mean (or range).

incidents. According to these data, most incidents were reported between noon and 5 p.m. local time, and not a single one—between 3 a.m. and 6.a.m. These data are consistent with the premise that wildfires in California tend to start in the afternoon (the hottest part of the day) when more people are active and available to observe them, and so they are reported soon after ignition.

3.2. Spatio-temporal matching of geospatial wildfire records against WF-ABBA detections

3.2.1. Fire activity period

Matching a fire pixel in space and time is based on the spatio-temporal proximity of the pixel to final size polygons $\{p_k\}$, $k=1,\ldots, 402$, of active fires. For a k-th fire, its estimated activity periods Δt_k were determined differently, depending on availability of the fire initial report time (IRT) in the geodatabase:

- For a fire with IRT available: $\Delta t_k = [\text{IRT}-3\text{ h}, \text{date of containment} + 2\text{ days}]$;
- for other fires: $\Delta t_k = [\text{initial alarm date at 12 am}, \text{date of containment} + 2\text{ days}]$.

The fire start and end times are defined as the left and the right ends of the estimated fire activity intervals, respectively. Adding the above margins to the recorded fire activity periods was meant to compensate for possible delays in initial report and for possible residual burning after containment, similar to previous studies (Feltz et al., 2003; Soja et al., 2009). Given the high population density in the study region, we believe that further increasing the fire start margin is not justified. Keeping these margins on a conservative side also reduces the chance of a false match of a detection to an incident record. Below in Section 3.3 we discuss that conservative (smaller) margins are preferable for incident-wise analyses.

3.2.2. Spatio-temporal matching of fire pixels

Spatial matching of WF-ABBA fire pixels to geo-polygons is somewhat less than completely straightforward, due to a number of reasons. A fire polygon can be arbitrarily intersecting a perfectly geolocated pixel or even lie outside. Furthermore, GOES pixels are not perfectly geolocated due to satellite position drifts, leading to an unknown and spatially and temporally varying error of geographic coordinates for WF-ABBA detections. A simple approach to compensate for these uncertainties is to use a single large spatial buffer. However, as in the case of using overly wide temporal margins (Section 3.2.1), this approach may positively bias detection quality estimates, and shift geo-location accuracy beyond what is practically useful (see our discussion of geolocation error in Section 1.2.1). Another problem with selecting an overly large buffer is that a GOES fire pixel may coincide in time and be spatially close enough to more than one fire incident, making unclear which incidents are actually "responsible" for the fire pixel. Therefore, a new ignition within the buffer distance of an older active fire is more likely to be "detected" earlier than it actually started.

We address these issues by 1) analyzing groups of *co-occurring* incidents, 2) using a two-buffer pixel-to-fire matching rule, and 3) tracking dynamic detected *events*, as discussed just below and also in Section 3.3.

Two or more fire incidents are considered *co-occurring* at the spatial scale of the GOES pixel if their activity periods overlap and their final perimeter polygons p_k are less than 6 km (approximately one GOES-W TIR pixel over California) apart. The co-occurrence analysis is facilitated by extracting the following categories from the 354 active fire incidents that started during the test period:

a. Fires co-occurring with one or more active wildfire incidents that started earlier, according to the estimated activity periods (Section 3.2.1).
b. Fires not in category a) that co-occur with a larger fire started at the same (up to 10 min) time.

Fires falling in categories a) or b) are excluded from the test sample for assessing detection timeliness. This leads to a conservative assessment of detection timeliness, because if a fire pixel matches to a group of co-occurring fires, then only one (the older or larger) fire is assumed detected. However, all 402 fires are included in the pixel-wise true/false positive analysis. Thus, the test sample for assessing detection timeliness included 310 of the 354 fires reported during the test period of which the initial report time was recorded for 164 fires, as was mentioned above in Section 3.1.1.

The pixel-to-fire matching rule uses two distance thresholds b_1 and b_2, where $b_1 < b_2$, for example $b_1 = 5.6$ km, $b_2 = 11.2$ km. Below in Section 4.1 we discuss in detail different choices of validation

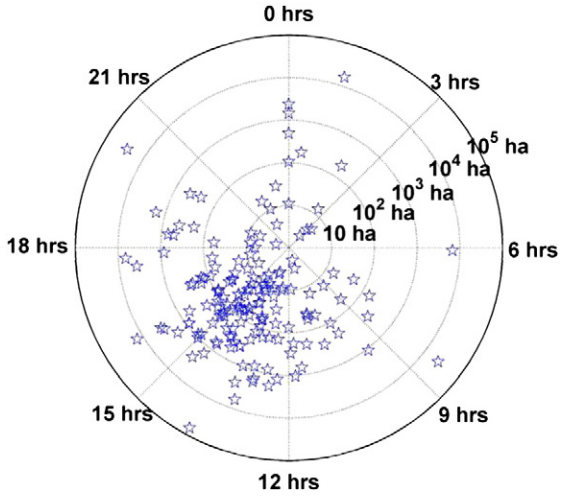

Fig. 4. Report time vs. final size. Scatter plot of the report time and final size distribution (on a logarithmic scale) for wildfires with known report time.

control parameters and their implications for the validation results. A fire pixel (x, t_{det}) centered at a spatial location x detected in a frame acquired at detection time t_{det} matches fire f_k, if and only if f_k is active at t_{det}, and at least one of the following conditions holds:

- x is within a buffer distance b_1 from the final fire perimeter polygon p_k;
- x is within a buffer distance b_2 from the final fire perimeter polygon p_k, and fire f_k is the closest active fire to the fire pixel (x, t_{det}).

3.3. Tracking and validating detected events

This section discusses the rationale and procedure for aggregating WF-ABBA pixel-wise detections in space and time into objects that we term *events*.

3.3.1. Events

Let a *detected event* (or simply an *event*, hereafter) be defined as a set of WF-ABBA designated fire pixels that are considered by the user as a possible single fire incident. As the definition suggests, a detected event may or may not correspond to an actual wildfire. Similar to actual fire incidents, detected events are dynamic objects: they may be identified in multiple (not necessarily continuous) frames and represented by one or more pixel locations per frame. We also postulate that every pixel that is flagged as a fire pixel by the WF-ABBA is a member of exactly one event. It is also understood from the above definition that pixel-members of an event should have similar spatial and temporal properties. At any given time, an event can be of one of the following two types:

a. an *existing event*, i.e. an event that was already detected by the algorithm in previous images and therefore assumed to have been reported to first responders; or
b. a *new event*.

To adequately evaluate and represent algorithm performance with respect to the number or timeliness of detecting wildfire incidents, a validation analysis should primarily focus on validating only new events. Indeed, a repeat detection of an existing event is expected to be acted upon by the user rather differently from detection of a new event. For example, a new event may be assigned a higher priority for verification and response. If an event is verified and flagged false positive, its re-detection (i.e. repeat detection) in subsequent images may not be as costly as its first appearance. If it is flagged true positive, its re-detection in subsequent frames may be of little informational value, as far as timeliness is concerned. In any case, a re-detection of an existing event could often be safely ignored by users interested primarily in maximizing timeliness of detection. Also, when an actual fire incident results in multiple fire pixels forming the same new event, the value of the detection information may not be directly proportional to the number of fire pixels matching this incident. Likewise, the user cost due to false information, i.e. a false positive new event, is not simply a multiple of the number of pixel-members of the event.

Since tactical operational decisions are generally made on a per-event basis and in near real time, it is helpful to imagine a simplified scenario in which an analyst is to make decisions about alerting first responders, solely based on WF-ABBA outputs. The analyst issues an alarm only if (s)he decides it is a new event, i.e. it will not be a repeat alert about the same possible wildfire. What should be the criteria for deciding whether this is a new event or an existing one? Obviously, the events analysis should be based only on the information that is available at the time of decision making. Ideally, one wants to simultaneously minimize the chance of ignoring a recurring or a new ignition and the chance of repeatedly announcing an alarm regarding the same phenomenon (whether an actual wildfire or a false alarm). In practice, different users would choose different balances between these two types of error. In the next section we provide what we believe is a reasonable example of an event analysis procedure, that we call an *event tracker*.

3.3.2. Event tracker

Our automated algorithm for extracting and tracking events across time is based on analyzing the temporal evolution of spatially *connected components* (c.c.) formed by WF-ABBA fire pixels in the GOES image coordinate system. A connected component is conventionally defined as a spatially connected group of pixels in the image. The algorithm has a control parameter which we call the history length h, which specifies the size of a temporal window during which spatially overlapping (up to the buffer distance b_2 defined in Section 3.2.2) connected components are considered the same event. For a frame collected at detection time t_{det}, the algorithm proceeds by the following steps:

1. Partition fire pixels into connected components.
2. Initially, when past WF-ABBA detections are not available, each c.c. is considered a new event.
3. When past WF-ABBA detections are available, for each fire pixel (x, t_{det}) find the nearest location y_x that was flagged as a fire pixel at least once during last h hours. If $\|x - y_x\| < b_2$, then pixel (x, t_{det}) is termed a *re-detected* fire pixel. Let $E(y_x)$ denote the event corresponding to y_x. Because $E(y_x)$ was detected before t_{det}, it is considered an existing event at time t_{det}.

The following steps assign pixels to existing or new events. These steps are illustrated in Fig. 5, where re-detected pixels are light grey and other pixels are dark grey. The text inside each pixel denotes the event to which the pixel will be assigned by steps 4 and 5 below.

4. Connected components with no re-detected pixels become new events (Fig. 5a) that, as was discussed earlier, may or may not match true fire incidents.
5. For connected components that do include re-detected pixels (Fig. 5b–h) the analysis is slightly more complex:
   a) First, each re-detected pixel-member (x, t_{det}) is assigned to its corresponding existing event $E(y_x)$ determined by step 3 (i.e. all light-grey pixels in Fig. 5b–h are assigned a label).

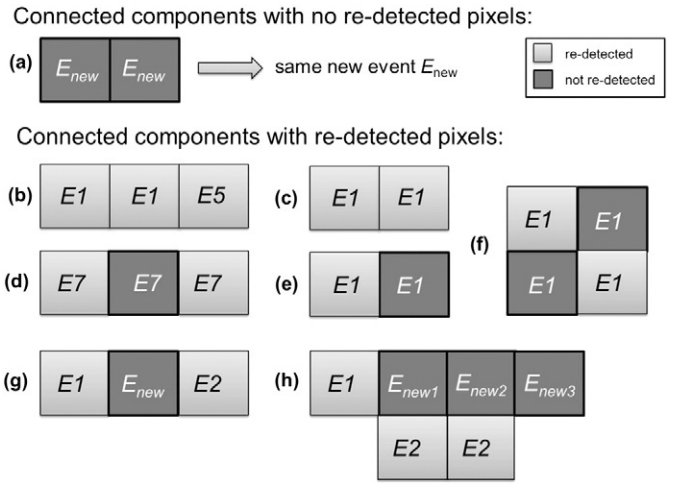

Fig. 5. Event tracking: assignment of pixels to events. Event tracking: an illustration for assignment of WF-ABBA fire pixels (light grey and dark grey square patches) to existing events (E1, E2, …) or new events (Enew, Enew1, Enew2, …), depending on the pixel spatial connectivity and proximity to recent historic WF-ABBA fire pixels. The text inside each square denotes the event to which this pixel is being assigned by steps 4 and 5 described in Section 3.3.2.

b) Next, pixel-members that are not re-detected pixels (dark pixels in Fig. 5) are assigned depending on the set of events to which re-detected pixel-members (light grey pixels) of this c.c. have been assigned by rule a) above. Specifically,
   i. if all re-detected pixel-members of this c.c. are assigned to the same event, then so are all other pixel-members of this c.c. (i.e. dark pixels of the connected components shown in Fig. 5d–f are labeled with the same label as the neighboring light-grey pixels);
   ii. otherwise, all pixel-members of this c.c. that are not re-detected pixels (dark pixels) are assigned to new events, one event per pixel (as illustrated in Fig. 5g and h).

An alternative option to ii) is to assign all of these pixels to one of the existing events to which the re-detected pixels were assigned by rule a): for example, in Fig. 5h all three dark grey pixels would be labeled with the same label, say "*E2*". These two options represent two opposite extreme choices allowing one to assess stability of the analysis results. Indeed, following the former option, as we did, produces a greater number of new events, thereby potentially biasing estimated performance toward more true positives, more timely detection, and more false positives. In turn, the latter option potentially increases incident detection latency and omission rate while potentially reducing the number of false positive new events (see Section 4.1 further discussing validation parameter choice).

The event-to-fire matching rule is based on the pixel-to-fire matching rule described in Section 3.2.2. Specifically, an event E matches fire f_k if at least one pixel assigned to E by the event tracker matches f_k.

The event tracker assigns a fire pixel to a new or existing event based on the recent (up to h hours back in time) history of WF-ABBA detections at adjacent pixel locations, which makes the event-wise performance measures less sensitive to the end time uncertainty than pixel-wise metrics. For example, without event tracking, if the beyond-containment margin (defined in Section 3.2.1) is too small for incident f_k and WF-ABBA keeps detecting fire pixels after Δt_k, these detections will be deemed false positives. In contrast, the event tracker may assign these pixels to an existing (i.e. previously detected) event that was matched to f_k by the above event-to-fire matching rule. In this case these pixels are not considered positives, and thus obviously not false positives. On the other hand, the event tracker assumes that the probability that a true positive new event (i.e. a new wildfire incident) is detected within h hours after a false positive was committed within the buffer distance b_2 is negligibly small. Therefore, when such a coincidence happens, the event tracker will fail to recognize a new event.

3.3.3. Interval estimation of false positive rates by matching with geo-spatial wildfire records and burn scar detection

The interval estimation for the $\mathbb{P}(F|E)$, which denotes the conditional probability that a new event E is a false positive, is based on the complete probability formula:

$$\mathbb{P}(F|E) = \mathbb{P}(F|U)\mathbb{P}(U|E) + \mathbb{P}(F|M)\mathbb{P}(M|E), \quad (1)$$

where "M" and "U" abbreviate "Matched" and "Unmatched", respectively, and denote outcomes of the event-to-fire matching according to the rule defined in Section 3.3.2. For example, $\mathbb{P}(M|E)$ is the probability that a randomly chosen WF-ABBA event is matched to a wildfire incident record in the geodatabase. The probability $\mathbb{P}(F|M)$ is assumed to be negligibly small, given that the database is compiled from official incident records and the fire activity period margins and spatial match buffers are narrow (Section 3.2.1). Thus, ignoring the second term on the right-hand side of Eq. (1), we obtain that for any number q, such that $\mathbb{P}(F|U) \geq 1\text{-}q$, the following inequality holds:

$$(1-q)\mathbb{P}(U|E) \leq \mathbb{P}(F|E) \leq \mathbb{P}(U|E). \quad (2)$$

Obviously, the probability $\mathbb{P}(U|E)$ is easily approximated by the fraction of new events detected by WF-ABBA that did not match any wildfire incident records. Thus, the problem of finding a non-trivial (i.e. non-zero) lower bound for the number of false positive new events is essentially reduced to the problem of finding a non-trivial lower bound for $\mathbb{P}(F|U)$, or equivalently, a value of q that is between 0 and 1. Our proposed approach to addressing this problem is to use temporal difference in the high resolution Landsat TM/ETM+ imagery to assess the likelihood that a new wild land burn scar appeared some time during the period between the chosen Landsat images before and after the WF-ABBA detected a new event. In mathematical terms, this can be written as:

$$\begin{aligned}\mathbb{P}(F|U) = {} & \mathbb{P}(F|\text{new burn})\mathbb{P}(\text{new burn}|U) \\ & + \mathbb{P}(F|\text{no new burn})\mathbb{P}(\text{no new burn}|U).\end{aligned} \quad (3)$$

A chosen procedure that detects possible new burns should minimize the chance of omitting wildfire burns that are detectable by WF-ABBA, in which case the probability $\mathbb{P}(F \mid \text{no new burn}) \approx 1$ (see Section 4.1), and therefore:

$$\begin{aligned}\mathbb{P}(F|U) \approx {} & \mathbb{P}(F|\text{new burn})\mathbb{P}(\text{new burn}|U) \\ & + \mathbb{P}(\text{no new burn}|U) \geq \mathbb{P}(\text{no new burn}|U).\end{aligned} \quad (4)$$

It follows from Eq. (4) that one can use $q = \mathbb{P}(\text{new burn}|U)$ in Eq. (2), leading to the following bounds:

$$\mathbb{P}(F|E) \geq 1 - \mathbb{P}(\text{new burn}|U) \quad \text{(a lower bound)} \quad (5)$$

$$\mathbb{P}(F|E) \leq \mathbb{P}(U|E). \quad \text{(an upper bound)} \quad (6)$$

In practice, we implemented $\mathbb{P}(\text{new burn}|U)$ estimation as follows. We selected a random sample $\{E_n\}$, $n = 1,\ldots,N$ of new events detected by WF-ABBA that were not matched by incidents in the geodatabase. Figs. 6, 7, and 8 illustrate our approach. Fig. 6 shows the locations of suspected WF-ABBA false positive new events $\{E_n\}$ and the footprints of the Landsat scenes used to evaluate them. For each of these events we defined a region of interest (ROI) that is a circular buffer of radius b_2 around the WF-ABBA event being tested. Without loss of generality, Fig. 7 displays rectangular ROIs centered at events E_1 and E_2. In each ROI, we computed an adjusted difference in the scaled Normalized Burn Ratio index (dNBR, van Wagtendonk et al., 2004; Key & Benson, 1999) between the chosen Landsat images before and after the WF-ABBA detected the new event. The algorithm for computing the Adjusted Normalized Burn Ratio Difference (dNBRA) is described in the Appendix. The dNBRA retains high sensitivity of dNBR to actual burns, and markedly reduces the number of falsely detected burns by dNBR over various surfaces (Figs. 7 and 8), which we found helpful for subsequent visual interpretation. Areas with dNBRA > 100 (brighter areas in dNBRA images in Figs. 7 and 8) were candidates for new scars and were subject to additional visual inspection in both pre-event and post-event RGB-composite images of TM/ETM+ band combinations 7,4,3 (Fig. 7). The new burn scar-candidates with area less than $a_{\min}$ hectares were rejected as unlikely to have caused the WF-ABBA detection being validated (see Section 4.1 discussion of specific values of $a_{\min}$ and other parameters). Otherwise, the burn scar-candidates were considered new burns. If a new active fire was observed in the post-E_n Landsat image, it was also considered a new burn for the purpose of this section.

Initially, the criteria for selecting a Landsat image pair for an unmatched WF-ABBA event E_n detected at time t_n were the following:

- Landsat 5, and
- 8 to 64 days between acquisitions, and

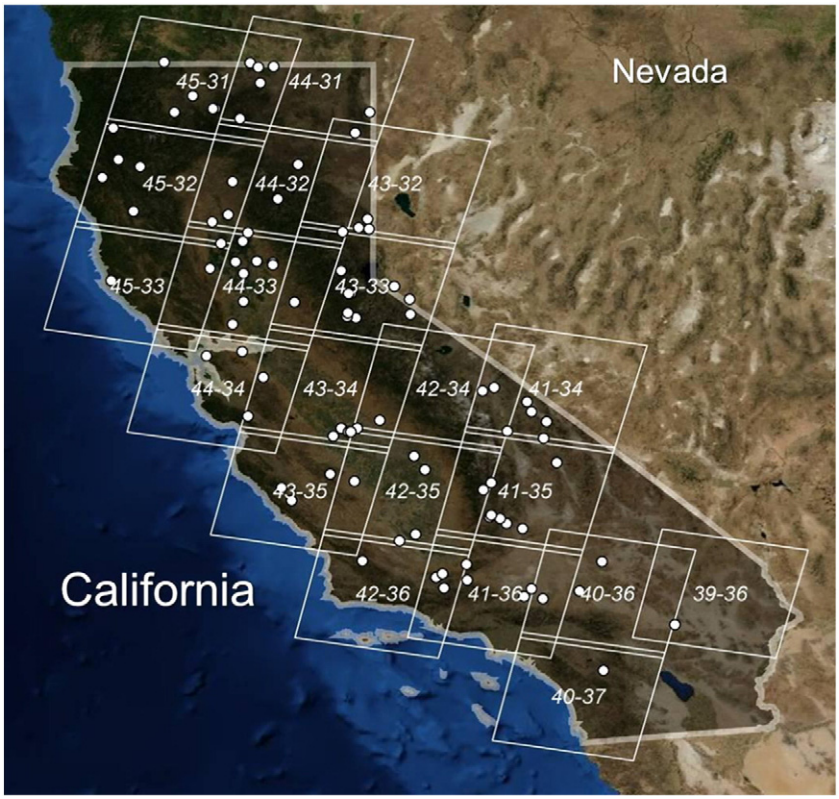

Fig. 6. Unmatched events and Landsat tiles. Locations of the suspected false positive new events (white circles) that were verified using Landsat images (see Section 3.3.3 for details). Labeled rectangles show the geographic coverage and the path-row numbers of the Landsat multitemporal scenes used for verification.

- t_n is no less than two days apart from the image acquisition dates, and
- low-to-none cloud cover over the corresponding ROI.

When no new burns were detected anywhere in the ROI for the initial Landsat image pair but a portion of the ROI was under cloud cover, that portion was additionally checked with an alternative Landsat pair or pairs, including Landsat 7. Any new burns detected in an image pair spaced by more than 32 days, were checked with alternative images, where available, for evidence that these new burns appeared more than one day before or after E_n was detected, in which case that new burn was rejected. Events for which insufficient cloud-free Landsat data were considered unverified and removed from the sample. Missing values in Landsat images, e.g. due to scan line correction failure in Landsat 7, were treated as clouds. Section 4.1 provides additional information on the interval estimation implementation.

4. Results and discussion

4.1. Validation experiment parameters

We feel it is important to mention up-front that the validation results presented and discussed in the following sections, figures, and tables, are stable with respect to the choice for free parameters in our validation experiments:

- buffers b_1 and b_2 used for spatial match, event tracking, and false positive rate correction;
- assumed margins for the recorded fire activity period (Section 3.2.1);
- history length, h, for tracking detected events (Section 3.2.2);
- the choice between rules for assigning pixels to events in step 5.b).ii) in Section 3.3.2.

Among these parameters, b_1, b_2, and h were found the most influential. The minimal values for b_1 and b_2 can easily be suggested as approximately half a pixel and one pixel linear size, respectively. The buffer size indirectly indicates the maximal acceptable error of geolocation for WF-ABBA detections (Section 1.2.1), with smaller values leading to more conservative estimates of detection quality but implying more accurate geolocation of true positives. Increasing history length for event tracking essentially turns off alarms over longer periods, biasing the results toward fewer false alarms and fewer detected actual wildfire incidents. In the following, the performance statistics are presented for the six options given below:

1) $b_1 = 3.4$ km, $b_2 = 6.8$ km, $h = 48$ h. The distance of 3.4 km is approximately 0.6 times the linear size (in the north–south direction) of the largest GOES-W TIR pixel in California.
2) $b_1 = 3.4$ km, $b_2 = 6.8$ km, $h = 72$ h.
3) $b_1 = 5.6$ km, $b_2 = 11.2$ km, $h = 48$ h;
4) $b_1 = 5.6$ km, $b_2 = 11.2$ km, $h = 72$ h;
5) $b_1 = 7.5$ km, $b_2 = 15.0$ km, $h = 48$ h;
6) $b_1 = 7.5$ km, $b_2 = 15.0$ km, $h = 72$ h;

Incident detection latency distributions (Fig. 9) and fire final size and duration distributions for detection timeliness groups (Figs. 10 and 11) are presented for $b_1 = 5.6$ km, $b_2 = 11.2$ km, and $h = 72$ h. We will discuss these results below in Section 4.3.

Given one of the above six choices for b_1, b_2, and h, different reasonable combinations of other validation parameters do not change any of the presented statistics by more than 3% of their respective absolute values. For example, when we use the alternative version of rule 5.b).ii) in Section 3.3.2 for assigning pixels to events, the upper bound for the average number of false positive new events per 24 h

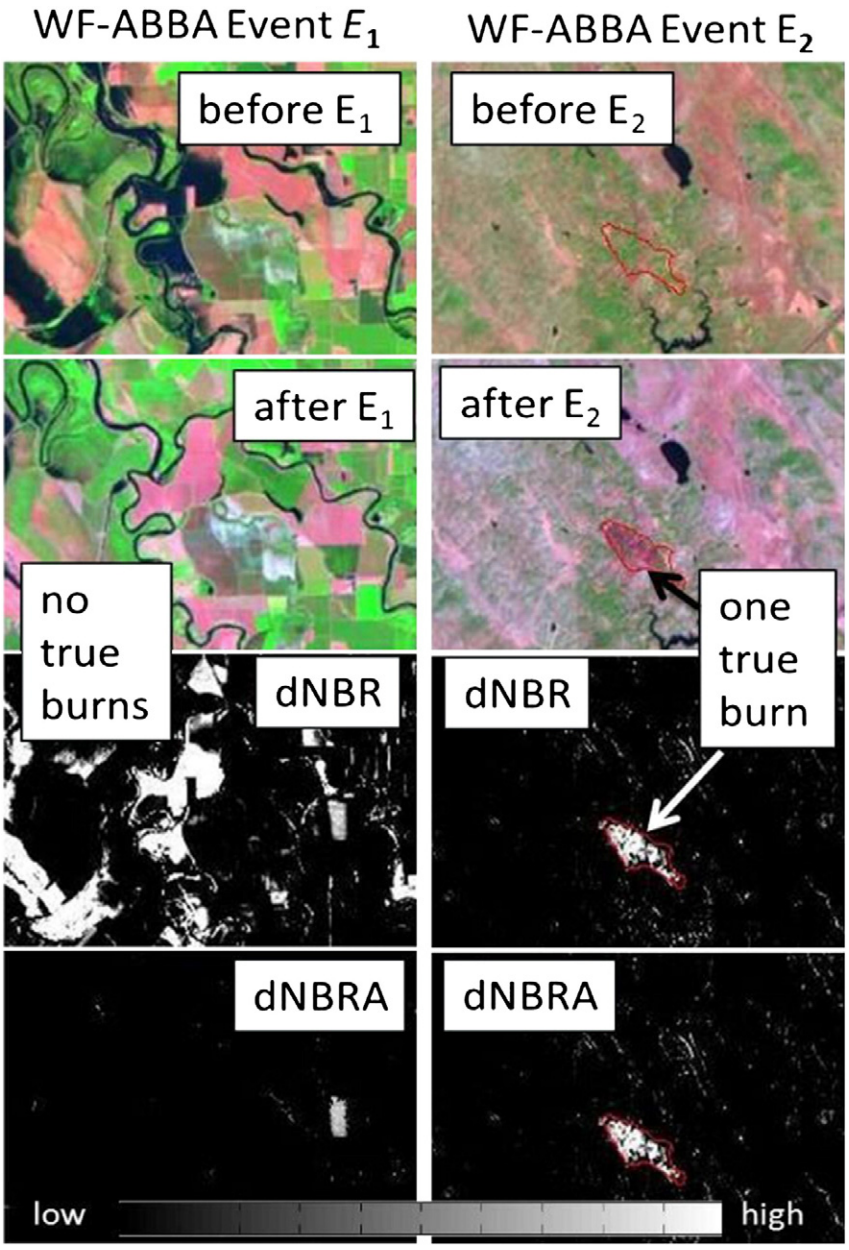

Fig. 7. New burn analysis illustration. Conceptual illustration of new burn analysis for two suspected false positive new WF-ABBA events, E_1 and E_2, using pre- and post-event Landsat images (see Section 3.3.3 and Appendix for details). The red outlined polygon represents an actual new burn. The color RGB-composite images use Landsat bands 7, 4, and 3, respectively.

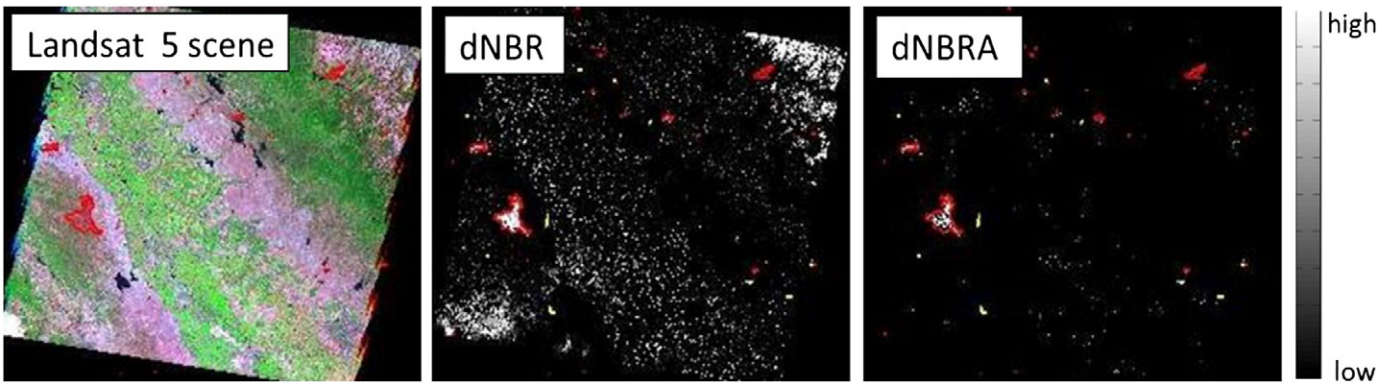

Fig. 8. dNBRA vs. dNBR. New burn pre-screening with dNBRA vs. dNBR in a Landsat scene (path-row 43-34). Red outlined polygons represent actual burned areas. Nearly all other high values in both dNBR and dNBRA images are not burns. The color RGB-composite image on the left uses Landsat bands 7, 4, and 3, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

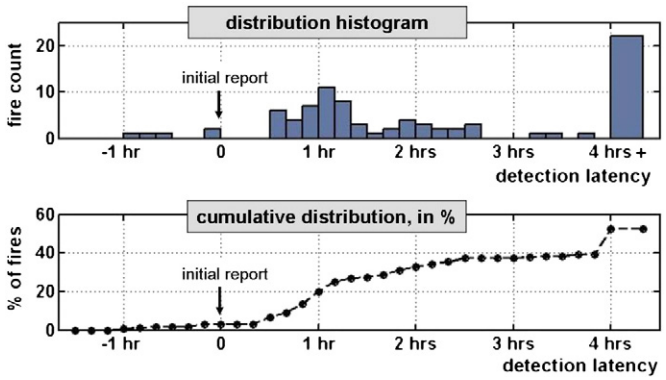

Fig. 9. Incident detection latency relative to initial report. WF-ABBA detection latency relative to recorded initial reports (the difference between WF-ABBA first alarm time and the initial report time) for new events detection. Incidents included in the sample are 164 tested incidents with known report times (Sections 3.1.1 and 3.2.2). The top graph displays a histogram of the relative latency distribution. In the bottom graph, for each latency value, $\Delta\tau$, the value of the cumulative distribution function represents the estimated proportion of incidents detected earlier than $\Delta\tau$ minutes since the initial report.

(Table 1) changes from 5.26 to 5.2596, and all other results presented in the figures and tables remain the same. Similarly, a bounding box (a minimal bounding rectangle) was used as a computationally efficient surrogate for fire perimeter polygons. Because a fire perimeter can be a very irregular shape, such a simplification could potentially affect our validation results. It turns out it did not: neither for pixels nor for events, as it is very rare that a buffered area of a GOES fire pixel overlaps with the bounding box but does not overlap with the perimeter itself.

For interval estimation of the false positive rates with Landsat imagery (Section 3.3.3), we randomly selected $N=122$ (~14%) of suspected false positive events, and the burned area threshold $a_{\min}=2$ ha, 5 ha, and 10 ha. Based on Landsat data availability, as discussed in Section 3.3.3, we were able to verify 99 of 122 events (Fig. 6). We applied our analysis to false positive events obtained with the values of b_1, b_2, and h, as listed in options 2) and 4) above. Furthermore, the assumption that $\mathbb{P}(\text{F}|\text{no new burn})\approx 1$, which justifies the validity of the bounds (5)-(6), was experimentally confirmed based on an ancillary experiment with 30 randomly chosen known actual wildfire incidents using 19 bi-temporal Landsat images. In this experiment, the new burns were clearly visible for 96.7% of the incidents. The probability $\mathbb{P}(\text{F}|\text{no new burn})$ is even greater, which can be established via Bayes' theorem using the fact that both new burns and wildfires are rare events (further details omitted).

4.2. Pixel-wise statistics

The pixel-wise performance statistics are summarized in the first two blocks of rows in Table 1. These pixel-wise statistics were obtained without event tracking. The last two blocks of rows show the event-wise statistics. These data are presented for different choices of the validation parameters and WF-ABBA fire confidence levels. Dividing a lower bound for the number of true positive pixels in this table by the total number of detections (e.g. $40.75/(40.75+15.08)$) shows that 73% to 74% of fire pixels detected at all confidence levels are matched in the recorded wildfire incidents. The actual pixel-wise performance is expected to be higher, because, as we discussed in several places above, some of the unmatched and therefore deemed false positive pixels may actually represent wildfires omitted in the incident report database we used. Additionally, unmatched detections may include, for example, agricultural fires and controlled burns.

4.3. Incident detection and its timeliness

For presentation purposes, we defined six non-exclusive groups of fire incidents:

1. Never detected fires.
2. Eventually detected fires.
3. Detected on the same date (if the IRT is unknown) or in less than 12 h (if IRT is known).
4. Detected in less than 2 h after the initial report.

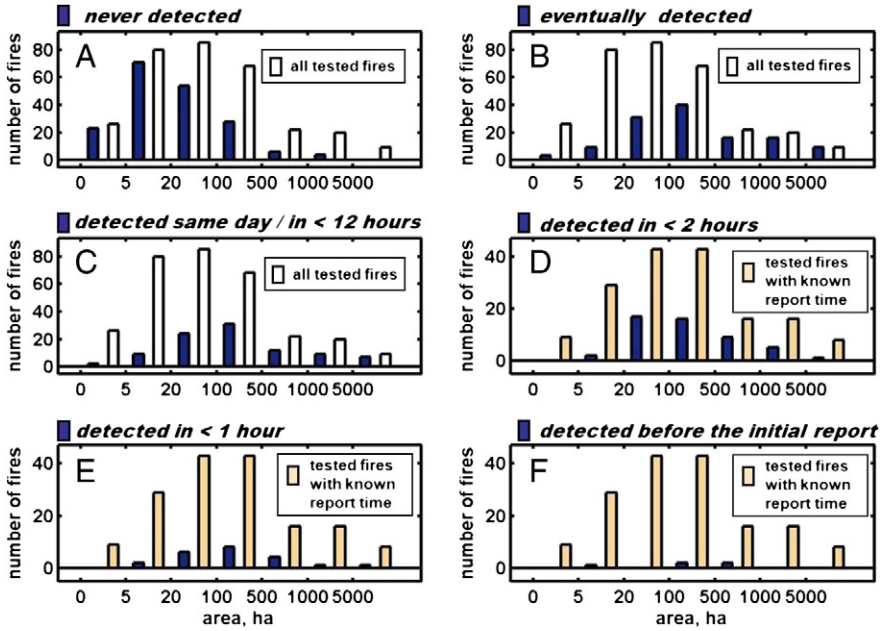

Fig. 10. Fire final size distribution for detection timeliness groups. New event detection by WF-ABBA: Histograms of final size distributions (dark blue bars) for undetected fires (A) and for fires in five (B)–(F) detection timeliness categories. White bars in all plots display all tested active fires, and beige color bars depict tested fires with known initial report time. Fire area is binned into 7 groups (see Section 3.1.1). The bar locations (or widths) on the area axis represent area group membership and not a within-group mean (or range). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

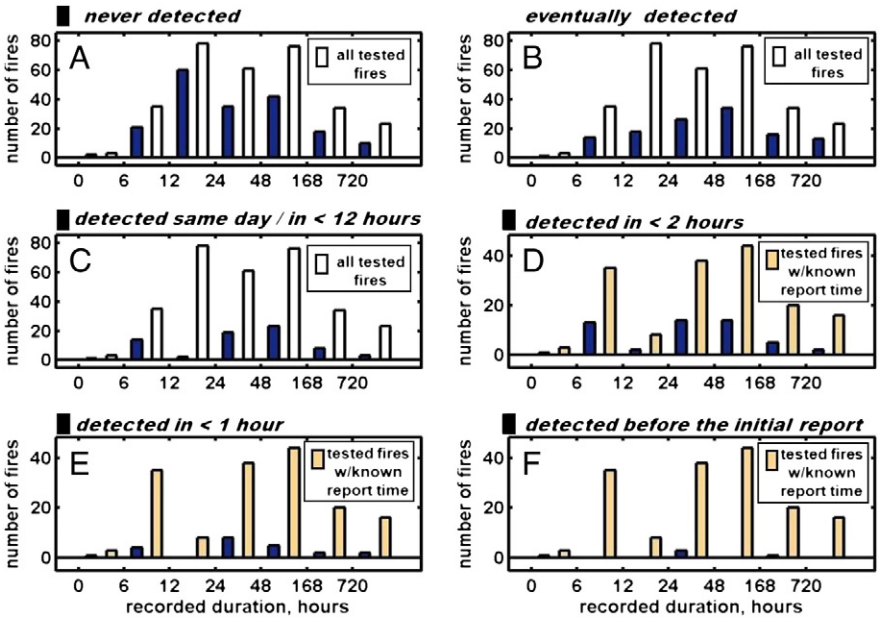

Fig. 11. Recorded incident duration distribution for detection timeliness groups. New event detection by WF-ABBA: Histograms of the recorded incident duration distributions (black bars) for undetected fires (A) and for fires in five (B)–(F) detection timeliness categories. White bars in all plots display distributions of all analyzed fires, as detailed in the respective plot legends. Fire duration is binned into 7 groups (see Section 3.1). The bar locations (or widths) on the area axis represent group membership and not a within-group mean (or range). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Detected in less than 1 h after the initial report.
6. Detected before the initial report.

The statistics for these detection timeliness groups are presented in Figs. 9 through 11, summarized in Table 2, and discussed below in this subsection.

Detailed information on detection timeliness is presented in Fig. 9, which shows a histogram of the WF-ABBA detection latency relative to the initial report from conventional sources. The relative latency is defined as the difference between the time of the first alarm from temporally filtered WF-ABBA products and the recorded time of initial reports. For example, negative values of latency indicate WF-ABBA detection prior to initial report. The latency does not account for GOES data delivery and processing time. Included in the sample are 164 tested incidents with known report times (Sections 3.1.1 and 3.2.2). As was discussed in Section 3.1.1, this sample does not appear to be an entirely random subset of the 310 recorded fires. In particular (Fig. 3A), smaller incidents are underrepresented in the 164-fire sample. Therefore, we also present the timeliness information in a stratified form in Figs. 10 and 11. Figs. 10 and 11 display burned area and incident duration distribution histograms for the above six groups, respectively. The area and duration axes are binned into the same seven intervals as histograms in Fig. 3 (see Section 3.1 for details). Overall, WF-ABBA detected ~30% of the fires within 2 h of the initial report time (see Table 2). It took the operational agencies and firefighters from 6 h to more than a month to contain these fires (Fig. 11D) that burned from less than 20 ha to more

Table 1
Pixel-wise and event-wise detection statistics. Pixel-wise and event-wise detection statistics under different choices of validation parameters (see Section 4.1). Event tracking history length h = 0 stands for "no event tracking".

Performance statistic	Buffer sizes and events tracking history length (b_1, b_2, h)	Values of the performance statistics per WF-ABBA fire detection confidence class					
		Medium flag = 4	High flag = 3	Partly Clouded flag = 2	Saturated flag = 1	Processed flag = 0	All fire classes with flags: 0–4
Pixel-wise:	3.4 km, 6.8 km, 0 h	1.10	6.17	9.95	0.32	23.22	40.75
Lower bound for mean true positive pixels per 24 h	5.6 km, 11.2 km, 0 h	1.13	6.22	9.98	0.33	23.44	41.10
	7.5 km 15.0 km, 0 h	1.13	6.24	10.06	0.34	23.58	41.35
Pixel-wise:	3.4 km, 6.8 km, 0 h	0.51	1.08	2.09	0.83	10.58	15.08
Upper bound for mean false positive pixels per 24 h	5.6 km, 11.2 km, 0 h	0.49	1.03	2.05	0.81	10.36	14.73
	7.5 km 15.0 km, 0 h	0.48	1.00	1.97	0.81	10.21	14.48
Event-wise:	3.4 km, 6.8 km, 48 h	0.062	0.096	0.317	0.034	0.600	1.11
Lower bound for mean true positive new events per 24 h	3.4 km, 6.8 km, 72 h	0.050	0.079	0.300	0.034	0.533	1.00
	5.6 km, 11.2 km, 48 h	0.062	0.096	0.329	0.050	0.634	1.17
	5.6 km, 11.2 km, 72 h	0.050	0.079	0.312	0.050	0.566	1.06
	7.5 km 15.0 km, 48 h	0.062	0.101	0.355	0.058	0.679	1.26
	7.5 km 15.0 km, 72 h	0.050	0.084	0.338	0.058	0.605	1.14
Event-wise:	3.4 km, 6.8 km, 48 h	0.242	0.566	1.046	0.492	2.909	5.26
Upper bound for mean false positive new events per 24 h	3.4 km, 6.8 km, 72 h	0.233	0.566	1.001	0.457	2.614	4.89
	5.6 km, 11.2 km, 48 h	0.242	0.566	1.034	0.475	2.873	5.19
	5.6 km, 11.2 km, 72 h	0.233	0.566	0.989	0.458	2.580	4.83
	7.5 km 15.0 km, 48 h	0.242	0.559	1.008	0.470	2.830	5.11
	7.5 km 15.0 km,72 h	0.233	0.559	0.962	0.454	2.539	4.75

Table 2
Detection timeliness summary. Fire detection timeliness summary statistics under different choices of validation parameters (see Sections 4.1 and 4.3).

Detection timelines group	Number of tested fires	Buffer sizes and events tracking history length (b_1, b_2, h)	When detecting by pixels		When detecting by new events only	
			Fires in group	% of tested fires in group	Fires in group	% of tested fires in group
Eventually detected	310	3.4 km, 6.8 km, 48 h	123	39.7%	119	38.4%
		3.4 km, 6.8 km, 72 h	–	–	117	37.7%
		5.6 km, 11.2 km, 48 h	131	42.3%	126	40.7%
		5.6 km, 11.2 km, 72 h	–	–	124	40.0%
		7.5 km 15.0 km, 48 h	138	44.5%	130	41.9%
		7.5 km 15.0 km, 72 h	–	–	128	41.3%
Detected same day or in less than 12 h	310	3.4 km, 6.8 km, 48 h	93	30.0%	92	29.7%
		3.4 km, 6.8 km, 72 h	–	–	91	29.4%
		5.6 km, 11.2 km, 48 h	96	31.0%	95	30.7%
		5.6 km, 11.2 km, 72 h	–	–	94	30.3%
		7.5 km 15.0 km, 48 h	100	32.3%	98	31.6%
		7.5 km 15.0 km, 72 h	–	–	97	31.3%
Detected in less than 2 h	164	3.4 km, 6.8 km, 48 h	52	31.7%	51	31.1%
		3.4 km, 6.8 km, 72 h	–	–	50	30.5%
		5.6 km, 11.2 km, 48 h	52	31.7%	51	31.1%
		5.6 km, 11.2 km, 72 h	–	–	50	30.5%
		7.5 km 15.0 km, 48 h	52	31.7%	52	31.7%
		7.5 km 15.0 km, 72 h	–	–	51	31.1%
Detected in less than 1 h	164	3.4 km, 6.8 km, 48 h	22	13.4%	22	13.4%
		3.4 km, 6.8 km, 72 h	–	–	22	13.4%
		5.6 km, 11.2 km, 48 h	22	13.4%	22	13.4%
		5.6 km, 11.2 km, 72 h	–	–	22	13.4%
		7.5 km 15.0 km, 48 h	22	13.4%	22	13.4%
		7.5 km 15.0 km, 72 h	–	–	22	13.4%
Detected before initial report	164	3.4 km, 6.8 km, 48 h	5	3.1%	5	3.1%
		3.4 km, 6.8 km, 72 h	–	–	5	3.1%
		5.6 km, 11.2 km, 48 h	5	3.1%	5	3.1%
		5.6 km, 11.2 km, 72 h	–	–	5	3.1%
		7.5 km 15.0 km, 48 h	5	3.1%	5	3.1%
		7.5 km 15.0 km, 72 h	–	–	5	3.1%

than 5000 ha each (Fig. 10D). A substantial number of fires, ~13% of the tested fires, were detected in less than 1 h after the initial report (Table 2). Remarkably, a few fires were detected even before the initial report. These include fires named: "Bassets", "Butte 2", "Marysville", 'Rollins", "Skyway". Figs. 10F and 11F show final size and duration distributions for these incidents. As might be expected and is demonstrated in Fig. 11A, the highest fraction of undetected fires are short-lived fires, with less than 24 h activity period, comprising ~44% of never detected fires. We observe in Fig. 11B that the fraction of eventually detected fires increases with fire duration. Furthermore, 80% of never detected fires are fires with final size under 100 ha (Fig. 10A). Analogously, as illustrated in Fig. 10B, WF-ABBA eventually detects larger fires with a high probability: 80% of fires with final size exceeding 500 ha were detected in our experiment. The smallest fire eventually detected by WF-ABBA was 2.1 ha ("Beauty"), and the smallest fire detected before the initial report burned 12.1 ha ("Rollins").

As mentioned at the end of Section 3.3.1, more than one WF-ABBA event can match a wildfire incident, depending on the chosen algorithm for extracting events from pixel-wise detections. On average, our event tracking algorithm identified 1.65 new events per eventually detected incident (under $b_1 = 5.6$ km, $b_2 = 11.2$ km, $h = 72$ h). However, protracted incidents tended to be detected as new events more often due to temporal gaps in detection of these fires (we believe, mainly from variable cloud cover, not accounted for in this paper, and fire regime dynamics) and when detected pixels for an incident do not form a connected component. For instance, the "Perkins" fire resulted in 5 new WF-ABBA events during one month of its activity, and the "Cottonwood" incident of August 5, 2006 was matched by two simultaneously detected new events, because the corresponding WF-ABBA detections did not represent adjacent pixels in the GOES image.

Furthermore, the tracking algorithm may ignore an actual wildfire incident that can be successfully detected by pixel-to-wildfire matching (Section 3.2.2). This happens when the first detection of the incident occurs within h hours after the same or an adjacent pixel location was flagged as a fire by the WF-ABBA. Table 2 demonstrates that this omission error (i.e. the difference between columns 4 and 6) is very small for our dataset.

Substantial differences between pixel-wise and event-wise performance statistics in Table 1 suggest that the WF-ABBA is better at re-detecting active fires than at initial detection. By its design, the event tracker eliminates many re-detected false positive pixels and counts multi-pixel false positives only once per event. Thus, many of the remaining false positive new events (last row and last column in Table 1) are represented by a small number of WF-ABBA fire pixels each, and they tend to recur after more than 3 days, if ever. The event-wise performance of WF-ABBA could be improved, by using a more advanced event tracking algorithm to take into account detection black-out periods due to e.g. cloud cover. This could be a subject for future research.

Table 3 presents the estimated lower and upper bounds for false positive new events. These bounds were obtained by combining wildfire official reporting information and multitemporal Landsat images (Section 3.3.3) under different choices for the validation control parameters discussed in Section 4.1. The width of these intervals and their variability with respect to control parameters reflect the various uncertainties in the validation process and datasets. The second column of Table 3 shows that an estimated 14% to 33% of the new events detected by WF-ABBA but unmatched by the incident reports, were not confirmed to be false positive wildfires. In other words, for this fraction of WF-ABBA events the results of Landsat multitemporal analysis were not inconsistent with the possibility that WF-ABBA had correctly detected a wildfire omitted in the incident geodatabase. Following the proposed validation methodology, narrower intervals could be obtained with additional time investment into Landsat image analysis and compilation of fire reporting information.

We acknowledge the fact that official incident reporting information may never be a perfect truth dataset, including systematic biases and

Table 3
Interval estimation of incident-wise false positive rates. Interval estimation of the average number of false positive new events per 24 h, based on verification of 99 randomly chosen suspected false positive new events, using multitemporal Landsat images (see also Sections 3.3.3, 4.1, and 4.3).

Buffer sizes, events tracking history length, and minimal new burn area threshold (b_1, b_2, h, a_{min})	Unmatched new events for which a new burn was found after Landsat image analysis (q)	Mean false positive new events per 24 h	
		Lower bound	Upper bound (from Table 1)
3.4 km, 6.8 km, 72 h, 2 ha	17 of 99 = 17%	4.06	4.89
3.4 km, 6.8 km, 72 h, 5 ha	16 of 99 = 16%	4.11	4.89
3.4 km, 6.8 km, 72 h, 10 ha	14 of 99 = 14%	4.21	4.89
5.6 km, 11.2 km, 72 h, 2 ha	33 of 99 = 33%	3.24	4.83
5.6 km, 11.2 km, 72 h, 5 ha	27 of 99 = 27%	3.53	4.83
5.6 km, 11.2 km, 72 h, 10 ha	20 of 99 = 20%	3.86	4.83

random errors. Importantly, our proposed validation methodology is consistent with this reality. In addition, we have recognized that while any kind of error in the database is possible, not all kinds of errors are equally probable. Furthermore, after proper preprocessing steps, we often can make reasonable assumptions about the statistical direction of the remaining major sources of error (e.g. underestimation bias or overestimation bias) and also about the types of errors the probability of which is negligibly small. In this way, we have been able to derive performance measures that are both useful and reliable, even though the wildfire report database is not to be completely trusted. We want to emphasize again that our method is specifically designed to evaluate timeliness and reliability of wildfire incidents. Having said that, we also believe that some of these ideas can be productive for validation of active fire products with respect to pixel-wise metrics.

We conclude this section by several additional remarks regarding our analyses and the presented results. Analysis of reasons for fire omission by WF-ABBA during the test period is beyond the scope of this paper. These reasons may include the lack of WF-ABBA or GOES Imager sensitivity, cloud cover or topography preventing the thermal signal from reaching the sensor, low instantaneous fire signal, or other reasons. Examples of other interesting questions that future follow-up studies could address include: performance of the temporally unfiltered WF-ABBA product, the effect of including low-possibility fire pixels, performance evaluation that is conditional on availability of cloud-free observations, and studying other potentially complex feedbacks between cloud cover and WF-ABBA performance. Consistent with its emphasis on detection timeliness, this paper has analyzed WF-ABBA performance over a season of substantial wildfire activity and danger in the study area. Furthermore, quite obviously, timeliness of wildfire identification from GOES benefits from high frequency of algorithm applications. However, for version 6.1, the WF-ABBA was only applied to GOES imagery acquired every 30 min, which contributes to delays in detection with respect to initial on-site reports. If a fire is immediately detected by WF-ABBA after the ignition, the minimal latency is 30 min + data delivery lag + processing time. Finally, a relatively high number of false positive new events (Table 3) suggests that geostationary detections should not be expected to replace other means of rapid wildfire identification but used as a complementary tool.

4.4. GOES-R prospect

The next generation of geostationary satellites will offer significant advancements in fire detection and monitoring (Schmidt et al., 2010). The GOES-R platform is scheduled for launch during the fourth quarter of 2015 and will include the Advanced Baseline Imager (ABI) offering improved temporal, spatial, spectral, and geolocation capabilities. Enhanced temporal resolution will include full disk coverage every 15 min and CONUS coverage every 5 min to capture short-lived fire activity. The ABI will have improved spatial resolution (2 km) in the short and long-wave infrared window bands (3.9, 10.8, and 11.2 μm) and a new channel centered at 10.3 μm. The elevated saturation temperature of 400 K in the 3.9 μm band will limit the number of saturated fire pixels to less than 5% of all observed fires. The enhanced capabilities of GOES-R are expected to significantly improve the timeliness and reliability of geostationary wildfire detection.

5. Summary and conclusions

The main contributions of the present paper include:

1) a new methodology for validating a geostationary fire product with respect to detection timeliness, including the following new features:
   - evaluating incident detection timeliness and reliability by combining geospatial fire records from operational agencies and multitemporal Landsat image analysis to identify new burn scars; which enables detection timeliness analysis with minute accuracy, makes validation independent of the satellite active fire detection products, provides large test samples for estimating true positive rates, and most importantly, yields reliable intervals (lower and upper bounds) for the number of false positive incidents;
   - an automatic algorithm for extracting potential new ignitions (event tracking) from WF-ABBA fire pixels;
2) systematic validation of WF-ABBA ver. 6.1 temporally filtered product with respect to detection timeliness, incident-wise metrics, and conventional performance measures over a large region in Western United States with significant fire activity and a wide range of biophysical conditions; and result interpretation;
3) results stability information and analysis under different choices of the validation parameters.

Our test site, the State of California represents a densely populated and well-monitored region, in which conventional fire identification normally comes before satellite detection. Nevertheless, our results suggest that even though WF-ABBA was not specifically designed to minimize the time to initial detection of a wildfire, a substantial fraction of wildfire incidents were detected within 1 h after the first initial report from conventional sources and on a few occasions—even before the recorded initial reports.

In the next few years we expect a substantial increase in the utilization of geostationary fire products over the continental U.S. and around the world. In particular, the WF-ABBA has undergone significant upgrades with the release and implementation of the global geostationary WF-ABBA (version 6.5.006) in 2010 (Prins et al., 2010). Improvements include the following:

1) Opaque cloud product indicating where fire detection is not possible.
2) Fire Radiative Power and Dozier instantaneous estimates of fire size and temperature.
3) Meta data on processing region; opaque cloud coverage; block-out zones due to solar reflectance, clouds, extreme view angles, biome type, etc.
4) Fire/meta data mask.
5) Revised ASCII fire product output: latitude; longitude; satellite view angle; pixel size; observed 4 and 11 μm brightness temperatures; instantaneous estimates of fire size, temperature, and FRP; biome type; fire confidence flag.
6) Processing of all available GOES-E/-W imagery including Rapid Scan and Super Rapid Scan mode when imagery is available as frequent as every minute over select regions of the continental U.S.

Furthermore, complementary to that effort, new approaches for earlier fire detection are being actively developed (Koltunov & Ustin, 2007; Koltunov et al., 2009). The improved information from WF-ABBA and the future new complementary systems that are

optimized specifically toward achieving the earliest alarm, will have a greater impact for the user communities. Finally, with the scheduled launch of GOES-R in 2015, we expect to see major improvements in geostationary fire detection and monitoring, including more rapid detection.

Our work emphasizes the potential of geostationary imagery for reducing the latency of wildfires in the Western U.S. and the importance of timely delivery of the images to the processing centers. We hope that the presented methodology, results, and discussion will increase WF-ABBA users' confidence, encourage much needed future validation efforts over different regions, and support expansion of remote sensing fire monitoring activities and applications.

Acknowledgments

We wish to acknowledge USDA FS and UC Davis support under contract # 0-IA-11130400-009, "Evaluating Operational Potential of Geo-Stationary Early Fire Detection Capabilities at Regional Level". We especially thank Mr. Brian Schwind and Mr. Brad Quayle from USDA FS Remote Sensing Application Center (RSAC) for their valuable support, information, and insightful discussions. We are grateful to Ms. Mui Lay (UC Davis) for help with Landsat image processing and analysis. We thank Mr. George Scheer (UC Davis) for hardware and system administration and other computation support that enabled us to analyze these data. We are thankful to Mr. Mark Rosenberg (CAL FIRE, Sacramento, CA) for kindly providing us with geospatial fire perimeter databases and other information related to incident reporting.

Appendix A. New burn scar detection in Landsat imagery with the Adjusted Normalized Burn Ratio Difference (dNBRA)

The NBR difference index (dNBR, van Wagtendonk et al., 2004) has been extensively used by the wildfire community, primarily for assessing burn severity and regeneration (French et al., 2008; Key & Benson, 2006). Its scaled version is defined by

$$\mathrm{dNBR} = 1000 \times \left(\mathrm{NBR}_{\mathrm{prefire}} - \mathrm{NBR}_{\mathrm{postfire}}\right), \tag{7}$$

where NBR is the Normalized Burn Ratio index (Key & Benson, 1999) computed from Landsat TM/ETM+ reflectance bands 4 and 7 by the following formula:

$$NBR = \frac{B4 - B7}{B4 + B7}. \tag{8}$$

To rapidly detect candidates for new burn scars in TM/ETM+ imagery, we propose an Adjusted Normalized Burn Ratio Difference (dNBRA) that is obtained by combining the radiance-based scaled dNBR (7) with additional filters to eliminate false scar detections as described below. In the Eqs. (9) through (13) we are using scaled radiance values (in the range of [1, 255].) of Landsat TM/ETM+ bands.

1. dNBRA = scaled dNBR,
2. dNBRA = 0, if any of the conditions (9) through (13) hold:

$$B_4^{\mathrm{post}} - B_7^{\mathrm{post}} > 1, \tag{9}$$

$$\frac{B_1^{\mathrm{post}}}{B_2^{\mathrm{post}} + B_3^{\mathrm{post}}} - \frac{B_1^{\mathrm{pre}}}{B_2^{\mathrm{pre}} + B_3^{\mathrm{pre}}} < 0.01, \tag{10}$$

$$\frac{B_4^{\mathrm{pre}}}{B_7^{\mathrm{pre}}} > 1.5 \ \text{ and } \ \frac{B_3^{\mathrm{pre}}}{B_7^{\mathrm{pre}}} > 1.5, \tag{11}$$

$$B_3^{\mathrm{pre}} > B_4^{\mathrm{pre}} \ \text{ and } \ B_3^{\mathrm{pre}} > B_7^{\mathrm{pre}} \tag{12}$$

$$B_3^{\mathrm{post}} > B_4^{\mathrm{post}} \ \text{ and } \ B_3^{\mathrm{post}} > B_7^{\mathrm{post}}. \tag{13}$$

In this way, according to our preliminary experiment using 50 new wildland burns that occurred between May and November 2006, dNBRA allows one to detect 96% of the actual burns detectable by dNBR, and drastically (by several orders of magnitude) reduces the number of falsely detected burns by dNBR over clouds, bright soils, urban areas, areas of phenological change (e.g. dry-out of grasslands), agricultural fields (e.g. harvest), and other land cover changes (Figs. 7 and 8). Small size forest fires where only understory is burning while canopies are not affected are likely to be missed by both dNBR and dNBRA. The remaining false detections by dNBRA are primarily due to misregistration and certain types of land cover changes. The dNBRA values are interpreted in the same way as those of scaled dNBR.

References

Al-Saadi, J., Szykman, J., Pierce, R. B., Kittaka, C., Neil, D., Chu, D. A., et al. (2005). Improving national air quality forecasts with satellite aerosol observations. *Bulletin of the American Meteorological Society*, *86*, 1249–1261.

Brioude, J., Cooper, O. R., Feingold, G., Trainer, M., Freitas, S. R., Kowal, D., et al. (2009). Effect of biomass burning on marine stratocumulus clouds off the California coast. *Atmospheric Chemistry and Physics*, *9*, 8841–8856.

Cardoso, M. F., Hurtt, G. C., Moore, B. I., Nobre, C. A., & Prins, E. M. (2003). Projecting future fire activity in Amazonia. *Global Change Biology*, *9*, 656–669.

Csiszar, I. A., Morisette, J. T., & Giglio, L. (2006). Validation of active fire detection from moderate-resolution satellite sensors: The MODIS example in northern Eurasia. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1746–1757.

Feltz, J. M., Moreau, M., Prins, E. M., McClaid-Cook, K., & Brown, I. F. (2003). Recent validation studies of the GOES Wildfire Automated Biomass Burning Algorithm (WF_ABBA) in North and South America. *Proceedings of the 2nd International Wildland Fire Ecology and Fire Management Congress and AMS 5th Symposium on Fire and Forest Meteorology, Orlando, Florida, November 16–20, 2003* 6 pp.

Freitas, S. R., Longo, K. M., Chatfield, R., Latham, D., Silva Dias, M. A. F., Andreae, M. O., et al. (2007). Including the sub-grid scale plume rise of vegetation fires in low resolution atmospheric transport models. *Atmospheric Chemistry and Physics*, *7*, 3385–3398.

French, N. H. F., Kasischke, E. S., Hall, R. J., Murphy, K. A., Verbyla, D. L., Hoy, E. E., et al. (2008). Using Landsat data to assess fire and burn severity in North American boreal forest region: An overview and summary of results. *International Journal of Wildland Fire*, *17*(2008), 443–462.

Giglio, L. (2010). MODIS Collection5 Active Fire Product User's Guide, Version 2.4. Available at: http://modis-fire.umd.edu/Documents/MODIS_Fire_Users_Guide_2.4.pdf

Giglio, L., Descloitres, J., Justice, C. O., & Kaufman, Y. J. (2003). An enhanced contextual fire detection algorithm for MODIS. *Remote Sensing of Environment*, *87*, 273–282.

Hawbaker, T. J., Radeloff, V. C., Syrphard, A. D., Zhu, Z., & Stewart, S. I. (2008). Detection rates of the MODIS active fire product in the United States. *Remote Sensing of Environment*, *112*(5), 2656–2664.

ILWDP (2008). *Assessing progress towards an integrated risk and cost fire management strategy*. Independent Large Wildfire Cost Panel (ILWDP). USDA 70 pp.

Key, C. H., & Benson, N. C. (1999). Measuring and remote sensing of burn severity. In L. F. Neuenschwander, & K. C. Ryan (Eds.), *Proceedings Joint Fire Science Conference and Workshop*, *vol. II*. (pp. 284)Moscow, ID: University of Idaho and International Association of Wildland Fire.

Key, C. H., & Benson, N. C. (2006). Landscape assessment: Sampling and analysis methods. *USDA Forest Service Gen Tech. Rep RMRS-GTR-164-CD*.

Koltunov, A., Ben-Dor, E., & Ustin, S. L. (2009). Image construction using multitemporal observations and dynamic detection models. *International Journal of Remote Sensing*, *30*(1), 57–83.

Koltunov, A., & Ustin, S. L. (2007). Early fire detection using non-linear multitemporal prediction of thermal imagery. *Remote Sensing of Environment*, *110*(1), 18–28.

Longo, K. M., Freitas, S. R., Andreae, M. O., Setzer, A., Prins, E., & Artaxo, P. (2010). The Coupled Aerosol and Tracer Transport model to the Brazilian developments on the Regional Atmospheric Modeling System (CATT-BRAMS)—Part 2: Model sensitivity to the biomass burning inventories. *Atmospheric Chemistry and Physics*, *10*, 5785–5795.

Matson, M., & Dozier, J. (1981). Identification of subresolution high temperature sources using the thermal IR. *Photogrammetric Engineering and Remote Sensing*, *47*, 1311–1318.

McNamara, D., Stephens, G., & Ruminski, M. (2004). The Hazard Mapping System (HMS)—NOAA multi-sensor fire and smoke detection program using environmental satellites. *Preprints, 13th Conf. on Satellite Meteorology and Oceanography*. Norfolk, VA: Amer. Meteor. Soc CD-ROM, 4.3.

Nepstad, D., Carvalho, G., Barros, A., Alencar, A., Capobianco, J., Bishop, J., et al. (2001). Road paving, fire regime feedbacks, and the future of Amazon forests. *Forest Ecology and Management*, *154*, 395–407.

Nepstad, D., Schwartzman, S., Bamberger, B., Santilli, M., Ray, D., Schlesinger, P., et al. (2006). Inhibition of Amazon deforestation and fire by parks and indigenous lands. *Conservation Biology*, *20*, 65–73.

NICC (2007). *Wildland fires and acres*. National Interagency Coordination Center (NICC) available at http://www.nifc.gov/fire_info/fires_acres.htm

Prins, E. M., Feltz, J. M., Menzel, W. P., & Ward, D. E. (1998). An overview of GOES-8 diurnal fire and smoke results for SCAR-B and 1995 fire season in South America. *Journal of Geophysical Research*, *103*(D24), 31.821–31.835.

Prins, E. M., & Menzel, W. P. (1992). Geostationary satellite detection of biomass burning in South America. *International Journal of Remote Sensing*, *13*, 2783–2799.

Prins, E. M., & Menzel, W. P. (1994). Trends in South American biomass burning detected with the GOES visible infrared spin scan radiometer atmospheric sounder from 1983 to 1991. *Journal of Geophysical Research*, *99*, 16719–16735.

Prins, E., Schmetz, J., Flynn, L., Hillger, D., & Feltz, J. (2001). Overview of current and future diurnal active fire monitoring using a suite of international geostationary satellites. In F. J. Ahern, J. G. Goldammer, & C. O. Justice (Eds.), *Global and regional wildfire monitoring: Current status and future plans* (pp. 145–170). The Hague, Netherlands: SPB Academic Publishing.

Prins, E. M., Schmidt, C. C., Brunner, J. C., Hoffman, J. P., Lindstrom, S. S., & Feltz, J. M. (2010). The global geostationary Wildfire ABBA fire monitoring network. *17th Conference on Satellite Meteorology and Oceanography, Annapolis, MD, September 2010*. Amer. Meteor. Soc P9.13.

Prins, E. M., Schmidt, C. C., Feltz, J. M., Reid, J. S., Westphal, D. L., & Richardson, K. (2003). A two-year analysis of fire activity in the Western Hemisphere as observed with the GOES Wildfire Automated Biomass Burning Algorithm. *12-th Conference on Satellite Meteorology and Oceanography, CD-ROM: Combined Preprints. 83rd AMS Annual Meeting, Paper P2.28., Long Beach, CA, February 9–13, 2003*.

Reid, J. S., Hyer, E. J., Prins, E. M., Westphal, D. L., Zhang, J., Wang, J., et al. (2009). Global monitoring and forecasting of biomass-burning smoke: Description of and lessons from the Fire Locating and Modeling of Burning Emissions (FLAMBE) program. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *2*(3), 144–162.

Reid, J. S., Prins, E. M., Westphal, D. L., Schmidt, C. C., Richardson, K. A., Christopher, S. A., et al. (2004). Real-time monitoring of South American smoke particle emissions and transport using a coupled remote sensing/box-model approach. *Geophysical Research Letters*, *31*(6).

Schmidt, C. S., Hoffman, J., & Prins, E. (September 27). *GOES-R Advanced Baseline Imager (ABI) algorithm theoretical basis document for fire/hot spot characterization, Version 2.0.* NOAA NESDIS Center for Satellite Applications and Research.

Schmidt, C. S., & Prins, E. (2003). GOES wildfire applications in the Western hemisphere. *Proceedings of the 2nd International Wildland Fire Ecology and Fire Management Congress and AMS 5th Symposium on Fire and Forest Meteorology, Orlando, Florida, November 16–20, 2003* 4 pp.

Schroeder, W., Prins, E. M., Giglio, L., Csiszar, I., Schmidt, C., Morisette, J., et al. (2008a). Validation of GOES and MODIS active fire detection products using ASTER and ETM plus data. *Remote Sensing of Environment*, *112*(5), 2711–2726.

Schroeder, W., Ruminski, M., Csiszar, I., Giglio, L., Prins, E., Schmidt, C., et al. (2008b). Validation analyses of an operational fire monitoring product: The Hazard Mapping System. *International Journal of Remote Sensing*, *29*(20), 6059–6066.

Soja, A. J., Al-Saadi, J., Giglio, L., Randall, D., Kittaka, C., Pouliot, G., et al. (2009). Assessing satellite-based fire data for use in the National Emission Inventory. *Journal of Applied Remote Sensing*, *3*, 1–29.

USFS (2007). FY2008 Forest Service Budget Overview, USDA Forest Service. Available at http://www.fs.fed.us/publications/budget-2008/fy2008-forest-service-budget-overview.pdf

van Wagtendonk, J. W., Root, R., & Key, C. (2004). Comparison of AVIRIS and Landsat ETM+ detection capabilities for burn severity. *Remote Sensing of Environment*, *92*, 397–408.

Wang, J., Christopher, S. A., Nair, U. S., Beid, J. S., Prins, E. M., Szykman, J., et al. (2006). Mesoscale modeling of Central American smoke transport to the United States. I. "Top-down" assessment of emission strength and diurnal variation impacts. *Journal of Geophysical Research*, *111*, http://dx.doi.org/10.1029/2005JD006416.

Weaver, J. F., Lindsey, D., Bikos, D., Schmidt, C. C., & Prins, E. (2004). Fire detection using GOES Rapid Scan imagery. *Weather and Forecasting*, *19*, 496–510.

Zhang, X., Kondragunta, S., Schmidt, C., & Kogan, F. (2008). Near real time monitoring of biomass burning particulate emissions (PM2.5) across contiguous United States using multiple satellite instruments. *Atmospheric Environment*, *42*.