Remote Sensing of Environment 276 (2022) 113043

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Timeliness in forest change monitoring: A new assessment framework demonstrated using Sentinel-1 and a continuous change detection algorithm

Eric L. Bullock[a,*], Sean P. Healey[a], Zhiqiang Yang[a], Rasmus Houborg[b], Noel Gorelick[c], Xiaojing Tang[d], Carole Andrianirina[e]

[a] US Forest Service, Rocky Mountain Research Station, Ogden, UT, United States
[b] Planet Labs, San Francisco, CA, United States
[c] Google Switzerland, Zurich, Switzerland
[d] Department of Earth and Environment, Boston University, Boston, MA, United States
[e] Laboratoire d'Observation des Forêts de Madagascar (LOFM), Bureau National des Changements Climatiques et de la REDD+ (BNCCREDD+), Ministère de l'Environnement et du Développement Durable (MEDD), Antananarivo 101, Madagascar

ARTICLE INFO

Editor: Marie Weiss

Keywords:
Forest monitoring
Disturbance
Timeliness
Near real-time
Land change monitoring

ABSTRACT

The development of near-real time forest monitoring systems, which are used to create alerts for events such as logging or fire, often involves tradeoffs between accuracy and timeliness. In the context of forest monitoring, timeliness is measured by the lag between when a change occurs in the forest and the creation of an alert about the event. Conventional accuracy assessments quantify both false negative and false positive errors in change maps, but do not specify those rates as a function of lag. Recent near real-time (NRT) accuracy assessment methods summarize the relationship between lag and correctly identified change events, but do not integrate consideration of changes mapped where they have not occurred. Here, we propose an assessment framework that we call "Sigmoid Accuracy", which characterizes forest change detection accuracy as it relates to timeliness. Using a change monitoring algorithm that applies the Continuous Change Detection and Classification algorithm (CCDC) to Sentinel-1 radar data in Madagascar, we demonstrate how the assessment framework can be used to calculate a lag-dependent F_1-Score to create a sigmoidal curve describing system performance as it relates to the time since a change event. We introduce two new performance metrics that define key moments along this curve: "Initial Delay", or the minimum time required to create an alert as observed in reference data, and the "Level Off Point", or the lag at which accuracy stabilizes. Our framework accommodates varying assessment designs, as demonstrated using pixel-, block-, and event-based agreement for the response design. These metrics provide a holistic way to evaluate and compare the usefulness of forest monitoring systems in real-world applications.

1. Introduction

Satellite remote sensing has proven invaluable for monitoring forest change. Early (1970s–2000s) approaches to mapping land cover or land use change were mostly retrospective in that they used historical data (over one year prior) to identify change after it occurred. However, advances in data acquisition and availability, processing environments, and change detection algorithms have enabled a shift towards monitoring of change in near-real time (Woodcock et al., 2020; Wulder et al., 2018). Operational systems now provide alerts for disturbances such as deforestation, in which a rapid response is needed (Hansen et al., 2016; Reiche et al., 2021; Shimabukuro et al., 2012). These systems have raised an important consideration for the development of forest change monitoring systems: not only is accuracy important, but so is timeliness.

1.1. Timeliness of a monitoring system

Timeliness refers to the latency between a change event, such as the felling of trees, and the creation of an alert from a monitoring system. Timeliness is important because there is often a limited time window that responsive action from the alert will be useful (Finer et al., 2018). For example, one application of forest monitoring is to identify illegal

* Corresponding author.
E-mail address: eric.bullock@usda.gov (E.L. Bullock).

https://doi.org/10.1016/j.rse.2022.113043
Received 18 June 2021; Received in revised form 14 March 2022; Accepted 6 April 2022
Available online 19 April 2022
0034-4257/Published by Elsevier Inc.

logging, in which the logging operations will often be conducted over a few hours or days and direct interdiction can require rapid (under one day) detection (Pendleton, 1998). A second example is to identify wildfires, where suppression efforts are most successful when starting soon after ignition (INSA, 2000). In both cases, the responses to an alert are most useful if they occur within a finite window, although the size of that window is application specific.

Most forest change monitoring systems create the most accurate maps when there are many pre- and post-disturbance data available. Retrospective analysis of change reduces sensitivity to ephemeral noise from faulty data; including stability of a change signal in the classification process helps to both eliminate spurious changes and increase algorithm sensitivity to more subtle shifts (Abercrombie and Friedl, 2016; Cohen et al., 2018). There is growing evidence that accuracy tends to improve with greater lag times (Reiche et al., 2018; Tang et al., 2019). Therefore, we argue that timeliness, alongside accuracy, is an integral component of forest monitoring that must be considered when developing and evaluating a near-real time forest monitoring system.

1.2. Causes for differences in timeliness

We address timeliness as a characteristic of a monitoring system, instead of as a trait of a map that the system might produce on a given day. The various components of a monitoring system are influenced by different sources of delay that together determine timeliness. Here, we define the three primary types of delay as observation lag, data lag, and algorithm lag (Fig. 1). Observation lag is the discrepancy between the time of data acquisition and the exact time when a change occurs. Since remote sensing data used for forest monitoring is usually acquired systematically and change events can happen at any time, there will likely be a period between the forest change event and when it is first observed by a remote sensing platform. For satellite sensors with standard orbital revisit times and no interference from clouds or cloud shadows the observation lag will be predictable. However, for Landsat or other optical sensors, clouds can increase the time between clear observations and therefore increase observation lag for variable lengths based on cloud cover.

The time between data acquisition and when it is useable for an application is data lag. For newly acquired data to be usable, it must be downloaded to a receiving station, pre-processed into a data product available to the public, and ingested into the workflow that runs the monitoring system. Data lag is systematic and under a week for most data. After data has been acquired and ingested, they can be used by a change detection algorithm to initiate alerts. Since most algorithms will not "confirm" a change the first time it is flagged, it can take multiple observations after a change event before an alert is created (Zhu et al., 2020). Therefore, algorithm lag will vary depending on the change detection algorithm. Unlike data and observation lag, algorithm lag can be partially controlled by the user through algorithm configuration.

1.3. History of timely forest monitoring through remote sensing

The concept of timeliness of forest change monitoring is relatively new. Early approaches to change monitoring made use of limited observations and technically arduous procedures for mapping change. Timely map creation was hindered by data lag due to unstandardized data processing pipelines, costs, and lag from computational requirements. These efforts focused primarily on improving map accuracy and timely change alerts were not seen as practical.

For example, Holmgren et al. (1989) evaluated Sweden's SPOT-based clearcut monitoring program and concluded that annual monitoring at the national level was infeasible due to clouds and costs associated with image acquisitions (Holmgren and Thuresson, 1998). Similarly, Brazil's National Institute for Space Research (INPE) has mapped deforestation using manual interpretation of Landsat imagery since 1988 (Instituto Nacional de Pesquisas Espaciais, 2000). However, analysis is performed for the preceding year, and persistent cloud coverage makes annual wall-to-wall mapping infeasible. Early land cover monitoring programs like Sweden's and Brazil's did not produce alerts that could be used to instigate early action, but instead performed retroactive (>two years) change detection to inform long term policy and/or resource management.

Early applications for near-real time (<one year) monitoring were primarily for active fire detection using coarse resolution sensors such as

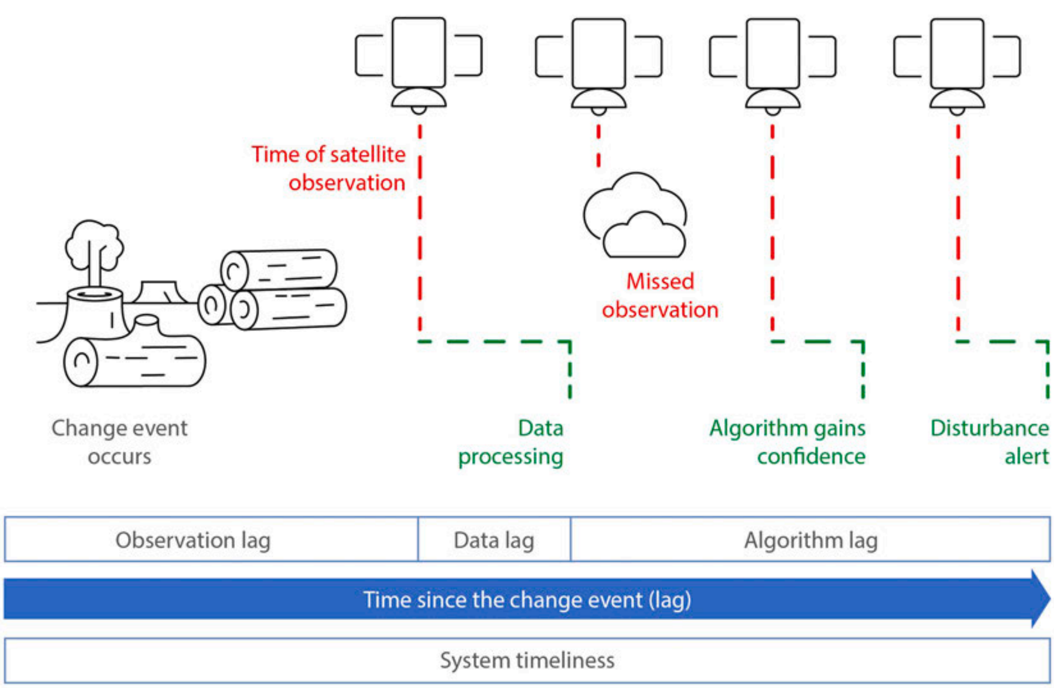

Fig. 1. Timeliness of a change alert is primarily controlled by three types of lag: observation, data, and algorithm lag. In systems where an alert can be issued upon the first observation, data lag and algorithm lag are the same.

the Advanced Very High Resolution Radiometer (AVHRR) and Moderate Resolution Imaging Spectroradiometer (MODIS). While methods for timely detection of active fires were proposed in the 1980s (e.g. Matson et al., 1987), operational active fires products were not produced until the early 2000s (Justice et al., 2002; Li et al., 2000). Accuracy and precision of these products were low (Li et al., 2001), and their coarse resolution limited local-scale analysis. These approaches were designed to detect active fires, and thus land cover conversion (e.g., deforestation) was not targeted. Prior to the 2010s, timely forest monitoring was limited to multi-year land cover change detection using (then-)costly Landsat or SPOT imagery and quicker detection of active fires using coarse resolution MODIS or AVHRR data.

The opening of the Landsat archive in 2008 (Woodcock et al., 2008) enabled methodologies based on dense time series data and, consequently, led to timely land cover change monitoring at moderate spatial scales. However, early (2010–2015) time series approaches to medium resolution (10-30 m) land cover change analysis were still performed years after the study periods, due in part to the computational requirements for data pre-processing and change detection. Landsat-based time series algorithms originally required local processing of cloud-masked surface reflectance data and computationally-intensive change detection processes (Huang et al., 2010; Kennedy et al., 2010; Verbesselt et al., 2010; Zhu and Woodcock, 2014). Furthermore, some methodologies did not operate iteratively or in a manner that could be updated with new observations, instead utilizing “offline” change detection on data stacks with a predefined end date. As a result, algorithm lag was usually high and the frequent updating of change maps remained challenging.

At the same time, approaches to land cover change monitoring in near-real time were being developed and operationalized using coarser resolution sensors such as MODIS and the Advanced Wide Field Sensor (AwiFS) platform. The near-daily repeat cycle and systematic processing of data from these sensors helped to improve system timeliness by reducing observation and data lag. The relative ease of processing coarse resolution (100-1000 m pixel resolution) imagery has the further advantage of reducing computation, thereby decreasing algorithm lag. Examples of early near-real time forest change monitoring systems include the Brazilian Detecção de Desmatamento em Tempo Real (DETER) program, Terra-I, and BFAST-monitor (Reymondin et al., 2012; Shimabukuro et al., 2007; Verbesselt et al., 2010). It is worth noting the diversity in change detection methods used by these systems: DETER uses manual interpretation, Terra-I uses machine learning, while BFAST-monitor is based on statistical break detection. Importantly, each approach was designed to update with new observations.

Continued developments in data availability, cloud computing, and time series algorithms have enabled the recent (2015-) deployment of near-real time monitoring systems at medium-to-high spatial resolutions. In addition to the freely available pre-processed data from Landsat, and Sentinel-1 and -2, cloud computing platforms such as Google Earth Engine and Amazon Web Services have also helped reduce algorithm lag by processing time series data more quickly and with fewer computing resources required by the user. Furthermore, change monitoring algorithms have been developed that accommodate a continuous influx of data to identify change at the end of the time series. Consequently, there are now operational systems for creating medium resolution alerts in near-real time (Hansen et al., 2016; Reiche et al., 2021, 2018).

For example, the Global Land Analysis & Discovery (GLAD) Alert System provides Landsat-based forest change alerts at 30 m resolution, while the RAdar for Detecting Deforestation (RADD) alert system uses cloud-penetrating C-band radar to create 10 m alerts (Hansen et al., 2016; Reiche et al., 2021, 2018). Notably, the RADD and GLAD systems were implemented on Google Earth Engine. Both systems are operationally automated, meaning that alerts are created from an algorithm rather than through manual interpretation.

Recent research on timely forest monitoring has shown potential of using high resolution (3 m) PlanetScope data (e.g. Francini et al., 2020), and for new applications such as a monitoring floods (Martinis et al., 2018; Twele et al., 2016), avalanches (Eckerstorfer et al., 2019), and pest damage (Pasquarella et al., 2017). There are also timely forest monitoring applications with less stringent latency requirements, for example evaluating compliance for payments for conservation of forest ecosystems (GFOI, 2020; Pérez-Sánchez et al., 2021).

1.4. Temporal assessment frameworks

The transition towards change monitoring has inspired new thinking about algorithm performance assessment. For near-real time systems to be useful they must produce timely and accurate results. Traditional accuracy assessments focus on quantifying errors of omission (false negatives) and commission (false positives) for a single, static map, omitting the element of timeliness. More recently, studies have moved beyond map accuracy to provide estimates of temporal accuracy or precision (Mardian et al., 2021; Zhu et al., 2012; Zhu and Woodcock, 2014). These approaches typically summarize the differences between change dates in a map and reference data as a measure of temporal performance (Boschetti et al., 2010; Campagnolo et al., 2021).

In the context of near-real time monitoring, the date labeled by the change detection algorithm is less important than the time the alert was created. The difference between the alert date and change event is the alert lag, which serves as the basis for evaluating timeliness. Reiche et al. (2015) proposed a mean time lag (MTL) metric using the average time between change events in a reference dataset and the confirmation of change alerts by a monitoring system. One limitation of MTL is that it is calculated as the average lag for the samples correctly labeled as change regardless of the error rates of the full dataset. MTL was proposed for consideration alongside spatial accuracy metrics, but this comparison could be misleading if lag is not also considered in the calculation of spatial accuracy. Map accuracy changes over time, and the accuracy at the end of a study period is not reflective of the accuracy at the average lag time. Accuracy will eventually stabilize at a saturation point, but the time it takes to reach that point is not characterized by a measure of central tendancy like MTL. While this information is not relevant for applications that require a latency faster than the saturation point, it is a useful measure for comparing system performance in applications with moderate (<one year) response requirements. Similarly, MTL does not provide an estimate on the minimum possible alert lag, which would be useful for applications with fast (1–2 day) response requirements.

Tang et al. (2019) introduced a near-real time assessment framework (henceforth, “the Tang framework”) as a way to incorporate lag in the calculation of omission error. The calculations in the Tang framework are designed so that accuracy can be calculated for any lag x. The result is a “alert-lag relationship” curve that describes the producer’s accuracy of the change class (“Detection Rate”) at specified lag times. Since Detection Rate (DR) is calculated using all samples that contain change in the reference data (not just correctly identified samples), it encapsolutes both spatial and temporal information without needing to compare to an independent assessment of omission error. However, like MTL, commission error in the Tang Framework is not expressed dynamically as it relates to lag and is instead calculated for a single map. It is therefore necessary to report DR alongside a traditional estimate of commission error, which can make interpretation challenging. Furthermore, the framework does not explicitly provide an estimate for the time the performance of a system saturates, or the minimum time required to create an alert.

There is a need for evaluation metrics that encapsulate spatial-temporal performance (lag, omission error, and commission error) at key moments in time after a disturbance. Lag averages (MTL) provide a measure of central tendency for correctly identified change samples, but interpretation can be subjective due to the independent calculation of spatial and temporal accuracy. The Tang framework provides a statistical approach for incorporating lag into the estimation of omission error. Yet neither approach incorporates commission error and do not explicitly provide estimates of the quickest alert time or the lag at which accuracy stabilizes.

1.5. Pixel-, block- and event-based agreement

The Tang framework formalized the concept of event-based agreement, in which a disturbance is considered correctly detected if a part of the event is identified, as opposed to measuring accuracy according to detection of randomly chosen individual pixels (i.e., pixel-based agreement). Choice of an agreement protocol for determining successful detection, often referred to as the response design, is necessary for building an error matrix and for computing area and accuracy under sample-based assessment designs (Olofsson et al., 2014). The protocol defines the properties of the map and reference labels that are needed to be considered agreement, such as the allowable discrepancies in size, location, timing, or classifications of the target map class (Stehman and Wickham, 2011).

An important decision when defining agreement between the map and reference data is the choice of a spatial unit for evaluating each sample. Common spatial units for reference samples include pixels, blocks (or windows), and polygons. Under a pixel-based design, a map pixel is used as the spatial domain of a sample, and the reference label of the area within each sample unit is compared to the classification of the corresponding map pixel. While commonly implemented and unambiguous from a sampling standpoint, pixel-based spatial units have been criticized as being arbitrary in size and too sensitive to geolocation errors (Congalton and Green, 2019). Tang et al. (2019) further critiqued pixel-based designs in the near real-time context since spatial precision is often less important than successful identification of change events. Block-based designs use a cluster of pixels as the spatial domain of each sample and can alleviate issues related to spatial mis-registration. Polygon-based designs make use of spatially varying and homogenous areas as the spatial domain of the sample units. For near real-time change monitoring, use of polygons representing the perimeters of forest loss patches may form the basis of an event-based design. Tang et al. (2019) suggested that event-based agreement be used for evaluating timely monitoring systems but noted that the arduous process of developing reference data for change events would likely prohibit its widespread adoption.

While the choice of response design will influence the results of an accuracy assessment, there is not one technique that is suitable for all situations. The spatial domain is one choice of many that go into a response design which can make comparison between independent accuracy assessments challenging. Accuracy metrics such as omission and commission error can be calculated under many different response designs and are only comparable across similarly designed assessments. Similarly, any metric for evaluating system timeliness should be flexible to design choices but comparable across assessments with matching designs.

1.6. Research objective

The rapid advancement of forest monitoring systems makes this an appropriate time to evaluate their usefulness in achieving real-life monitoring objectives. Previous research such as Reiche et al. (2015), Reiche et al. (2018), and Tang et al. (2019) were instrumental steps towards replicable characterization of system timeliness. As noted in the previous sections, however, they are limited in many regards, and there remains no agreed-upon standard for evaluating accuracy as it relates to timeliness.

The objective of this research is to propose a novel assessment framework for evaluating the timeliness and accuracy of a forest monitoring system. We demonstrate this framework using a case study in Madagascar that applies Continuous Change Detection and Classification (CCDC) to Sentinel-1 data. Our application of CCDC is well suited to demonstrate the concept of timeliness: it has tunable parameters that affect the algorithm's accuracy and timeliness; it is an "online" monitoring approach that accommodates new data as it becomes available; it operates on Google Earth Engine; and it is not affected by clouds. This research serves to link accuracy with timeliness, providing an assessment framework that can be used to choose from available change detection algorithms according to system needs and sensitivities.

2. Methodology

2.1. The Sigmoid Accuracy Assessment Framework

We propose an assessment framework named "Sigmoid Accuracy" that incorporates both omission and commission of change alerts as a function of lag. The framework graphs accuracy as a function of lag time. As will be seen, that function generally has a sigmoidal shape, allowing reduction of a system's time/accuracy relationship to the 2D coordinates of two metrics that may be comparable across monitoring systems. These metrics are: "Initial Delay", or the minimum lag with which an alert can be created, and "Level Off Point", or the lag at which gains in accuracy begins to reach an asymptote. The following procedures are used to evaluate a monitoring system using our proposed workflow:

1. Create a probability sample for a study area.
2. Define an agreement protocol for comparing change alerts with reference samples.
3. Interpret the reference sample so that it contains information on precise change timing and location.
4. Calculate agreement at a given lag x between the reference information and forest change dataset for each sample unit.
5. Use the reference-agreement data in the context of the sample design to calculate omission and commission of forest change at lag x.
6. Calculate an F_1-Score from omission and commission error at lag x.
7. Repeat steps 4–6 for incremental values of x until reaching a maximum lag time.
8. Compute the Initial Delay and Level Off Point using the relationship of lag to F_1 calculated in step 7.
9. Calculate standard errors and confidence intervals using bootstrapping.
10. Report the (lag, F_1) coordinates of the Initial Delay and Level Off Point with standard errors and confidence intervals.

The reference sample must be selected under a probability design and include examples of forest change. For that reason, we recommend using a stratified random sample to target samples in a forest change class (e.g., deforestation), although other forms of probability sampling can be accommodated (Olofsson et al., 2013). In the simplest case, the stratification can be created using a map of change alerts to define two strata, "Change", and "Not Change". However, there are techniques for improving the efficiency of the stratification such as incorporating additional strata or including a spatial "buffer" class (Arévalo et al., 2020; Bullock et al., 2020b, 2020c; Hansen et al., 2016; Olofsson et al., 2020; Potapov et al., 2017; Reiche et al., 2021; Tyukavina et al., 2013). These choices are application specific, and we recommend following "Good Practice" guidelines described in Olofsson et al. (2014) for designing a stratification and sample for evaluating land cover change data.

Reference data collected at each sample location must contain information on the date the change is first observed in the reference data (T_{ref}) and the date of an alert for that sample unit (T)(Reiche et al., 2018). Importantly, T is the time the alert is created, not the date that the algorithm attributes the change to (T_{map}) or the first date labeled as a possible (or low-probability) change (T_F). The difference between T_{map} and T is determined by a system's algorithm lag. For applications with near-daily reference data, T_{ref} is a good approximation of the "true" date of change. However, there may be applications in which T_{ref} cannot be precisely determined due to infrequent reference data. For such cases, Reiche et al. (2018, eq. 4) proposed an adjustment to T_{ref} ($T_{refadjusted}$), which is calculated as the time directly between the time the change is first observed in the reference data (T_{ref}) and the previous reference image ($T_{refprevious}$):

$$T_{refadjusted} = T_{ref} - \frac{T_{ref} - T_{refprevious}}{2}$$

Agreement between the reference data and an alert is defined by the agreement protocol (i.e., the response design). We do not propose a standardized protocol, but instead suggest users to define agreement based on their specific application. It is important to note that these decisions will influence the assessment results and comparison should only be made between assessments with similar designs. Here we define temporal agreement and test three approaches for defining spatial agreement. Temporal agreement is determined by a "Tolerance Window" parameter, defined here as a minimum and maximum lag before and after an event in which an alert is considered correct (for example, −20 to 365 days). Negative lags should be included to accommodate imprecise reference data or situations where an algorithm identifies early signs of change before it is visible in the reference data.

Our framework accommodates spatial agreement using pixel-, block-, and event-based spatial units. For pixel-based agreement, a change is considered correct if the pixel is classified as forest loss in the reference data and in an alert within the Tolerance Window. For block-based agreement, a change alert is considered correct if one pixel within a 3 × 3 window is classified as a change alert within the Tolerance Window, and the center pixel is forest loss in the reference data. For event-based agreement, a change alert is considered correct if the sampled pixel is forest loss in the reference data, and a predefined percentage of the overlapping patch of forest loss is classified as a change alert within the Tolerance Window. As discussed in Tang et al. (2019), change patches can be large, and therefore it is not sufficient to define agreement when a single pixel within the event is flagged as change. Instead, a Detection Threshold parameter (P_t) is used to define the minimum percentage of the event that must be identified for it to be considered a correctly identified event.

The reference/agreement data can then be used to calculate omission error, commission error, and F_1-Score for a given lag x (Chinchor, 1992; van Rijsbergen, 1979). An F_1-Score is the harmonic mean of a model's precision and recall, which in map terminology is equivalent to User's and Producer's accuracy or the inverse of Commission and Omission error:

$$F_1 - Score = 2^* \frac{(Precision^* Recall)}{(Precision + Recall)} = 2^* \frac{(Users^* Producers)}{(Users + Producers)} = 2^* \frac{(1 - Commission)^*(1 - Omission)}{(1 - Commission) + (1 - Omission)}$$

An F_1-Score provides a general accuracy metric for a specific map class, such as forest loss, and can be calculated at varying lags. In other words, the relation of lag to F_1-Score describes how well a system performs at detecting change within a specific amount of time after the change occurs.

To calculate the F_1-Score (referred to hereafter as F_1 for brevity), it is necessary to calculate omission and commission error at a specified lag x. To do so, the change date (T_{ref} or $T_{refadjusted}$) and alert date (T) must be accounted for when estimating accuracy. This can be done by treating each sample as an indicator observation with parameters y_u and z_u which can be either 1 or 0, where

$$y_{u,x} = \begin{cases} 1 \text{ if pixel is correctly labeled as change within lag } x \\ 0 \text{ if pixel not correctly labeled as change within lag } x \end{cases}$$

$$z_u = \begin{cases} 1 \text{ if reference label is change} \\ 0 \text{ if reference label is not change} \end{cases}$$

A sample is correctly classified as change within lag x if the difference between T_{ref} (or $T_{refadjusted}$) and T is less than x. The parameter x is bounded by the Tolerance Window. To avoid double counting a reference observation, in addition to simplifying the agreement protocol for defining $y_{u,x}$, we recommend excluding or replacing sample units with more than one change event or alert within the Tolerance Window (see Discussion for the disadvantages of this simplification). Samples with a lag before or after the bounds of the Tolerance Window are treated as commission and omission error, respectively. Following the logic explained in Stehman (2014), omission error can be calculated as:

$$omission_x = 1 - \frac{\sum_{u=1}^{N} y_{u,x}}{\sum_{u=1}^{N} z_u}$$

The parameters for commission error are expressed as:

$$y_{u,x} = \begin{cases} 1 \text{ if pixel is correctly labeled as change within lag } x \\ 0 \text{ if pixel not correctly labeled as change within lag } x \end{cases}$$

$$z_{u,x} = \begin{cases} 1 \text{ if pixel labeled as change within lag } x \\ 0 \text{ if pixel not labeled as change within lag } x \end{cases}$$

For commission errors, the detection lag x is an inherent property of the algorithm rather than a comparison of an alert and change date. Specifically, the commission lag is the minimum theoretical time within which an algorithm can create an alert and can be estimated using the input data and parameterized change detection algorithm. Commission error can be calculated using the same ratio function as omission error. Some algorithms are designed to reanalyze detected change as data are added to the time series (e.g. Bullock et al., 2020a), while others treat change alerts as immutable, and therefore the shape of the lag-commission curve can vary substantially among algorithms.

This form of ratio estimation can easily incorporate a stratified sample. For example, the parameters for omission errors for stratum s can be expressed as:

$$y_{u,x,s} = \begin{cases} 1 \text{ if sample is correctly labeled as change within lag } x \text{ and was selected from stratum } s \\ 0 \text{ if sample not correctly labeled as change within lag } x \text{ or was not selected from stratum } s \end{cases}$$

$$z_{u,s} = \begin{cases} 1 \text{ if reference label is change and the sample was selected from stratum } s \\ 0 \text{ if reference label is not change or the sample was not selected from stratum } s \end{cases}$$

Parameters $\overline{y}_{x,s}$ and $\overline{z}_s$ are the stratum and lag specific means of $y_{u,x,s}$ and $z_{u,s}$. Given area weights W_s and L number of strata, omission error could then be calculated as:

$$omission_x = 1 - \frac{\sum_{S=1}^{L} W_s \overline{y}_{x,s}}{\sum_{S=1}^{L} W_s \overline{z}_s}$$

F_1 at lag x is defined as:

$$F_{1,x} = 2 * \frac{(1 - Commission_x) * (1 - Omission_x)}{(1 - Commission_x) + (1 - Omission_x)}$$

F_1 ranges from 0 to 1 in a comparable manner to accuracy. Plotting F_1 as it relates to lag reveals two important time segments in the detection process: the lag before a first alert ("Initial Delay") and the lag at which the curve levels off ("Level Off Point"). The time of an Initial Delay defines the minimum time the system can create a change alert and is dependent on algorithm parameters that control the number of observations to confirm a change in addition to observation and data lag. Here we use the term "confirm" to refer to the assignment of a change alert by the algorithm, and "verify" when establishing an alert's accuracy through comparison to reference data. There may be cases when a single or a few events are detected earlier than observation, data, and algorithm lag would theoretically allow. Causes of premature detection include commission of change that overlaps later change by chance, or erroneous date attribution in the reference data. Therefore, we suggest assigning the Initial Delay to the date that 2% of the total change events are detected, thereby minimizing the impact of unrealistically quick and nonrepresentative samples. The lag at which the curve levels off signifies the time it takes for accuracy to stabilize. Under the assumption that F_1 will rise steadily as a function of lag (after a delay related to system lag and before an asymptote due to detection limitations), a graph of accuracy as a function of lag should reveal a sigmoidal accuracy function, as the framework's name implies (Fig. 2B).

There are numerous ways to calculate a curve "leveling off" or saturation point. In computer systems, this point is sometimes referred to as a knee, and is an important concept for parameter optimization and efficient system designs (Salvador and Chan, 2004; Satopää et al., 2011). Here we propose to use the F_1 curve knee as a way of measuring the Level Off Point. As described in Satopää et al. (2011), one generalizable approach to calculating the knee of a discrete set of points is to calculate the point of maximum curvature (Fig. 2). This point corresponds to x at the maximum value of y for a curve rotated 45 degrees around the line that connects the beginning and end points of the curve. To ensure even rotation, it is necessary to first rescale y to the bounds of the Tolerance Window.

The lag-dependent accuracy curves shown in Fig. 2 are an extension of the alert-lag relationship first introduced in the Tang assessment framework. The two derived metrics, Initial Delay and Level Off Point, fill in important gaps, allowing simultaneous consideration of timeliness and accuracy (omission error and commission error). A Python notebook demonstrating the calculations in Fig. 2 is included in the Appendix.

For interpretation of the Initial Delay and Level Off Points it is important to consider both lag and accuracy. Therefore, we recommend using a coordinate syntax (x, y), or (lag, F_1). For example, the Level Off Point in Fig. 2 has a lag of 80 days, at which the F_1 is 0.82 (80, 0.82). Coordinates for the Initial Delay in Fig. 2 are (10, 0.05). Alternatively, F_1 can be substituted for error rates (omission or commission error) or

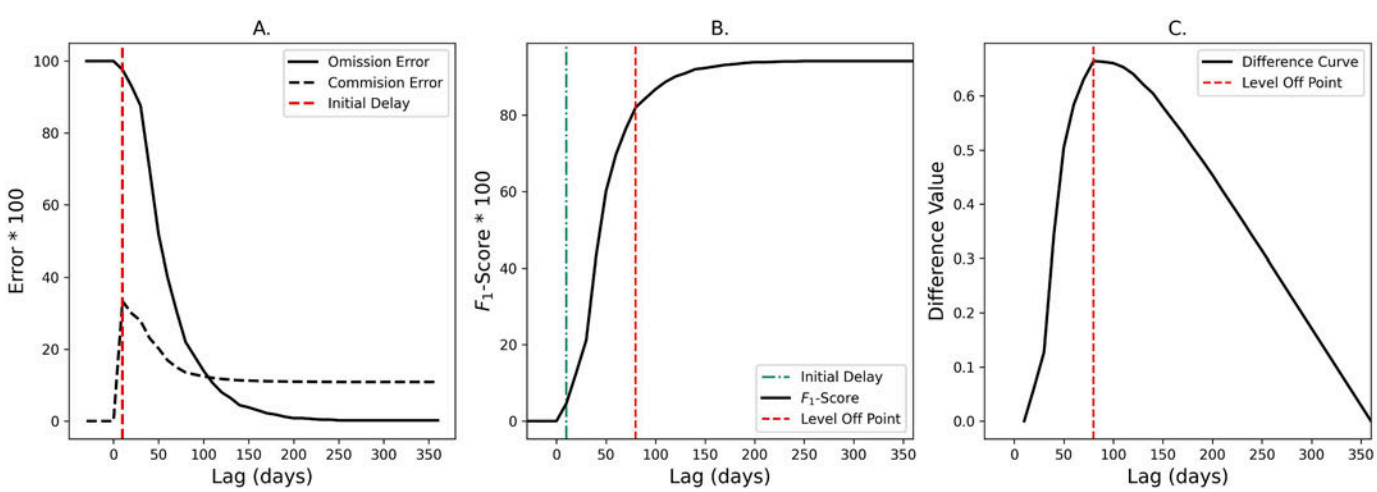

Fig. 2. F_1 is calculated from omission and commission error as a function of time (A.). The minimum time that an alert can possibly be created is the Initial Delay (B.). The Level Off Point is calculated as the point of maximum curvature in F_1 curve (B.), which can be found at the maximum value after rotating it 45 degrees around the line connecting the beginning and end points of data rescaled to the Tolerance Window (C.).

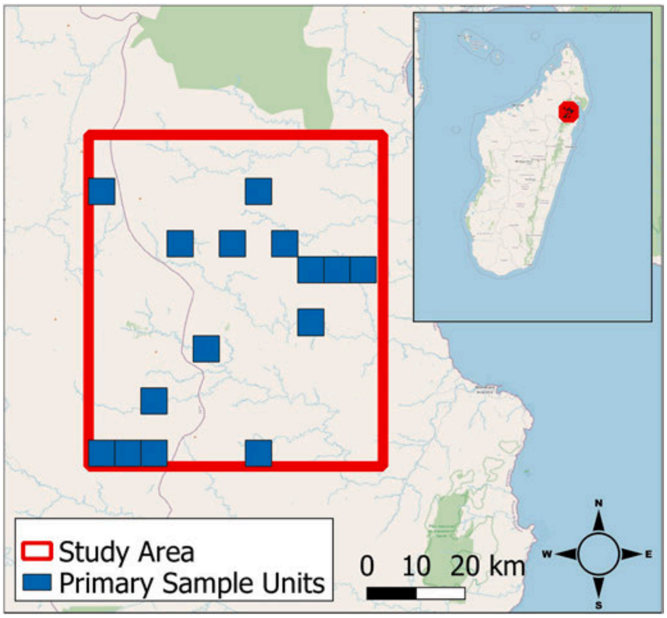

Fig. 3. The study area in Northeastern Madagascar. Assessment was performed within 15 3,000-ha grid cells that were randomly selected from our study area.

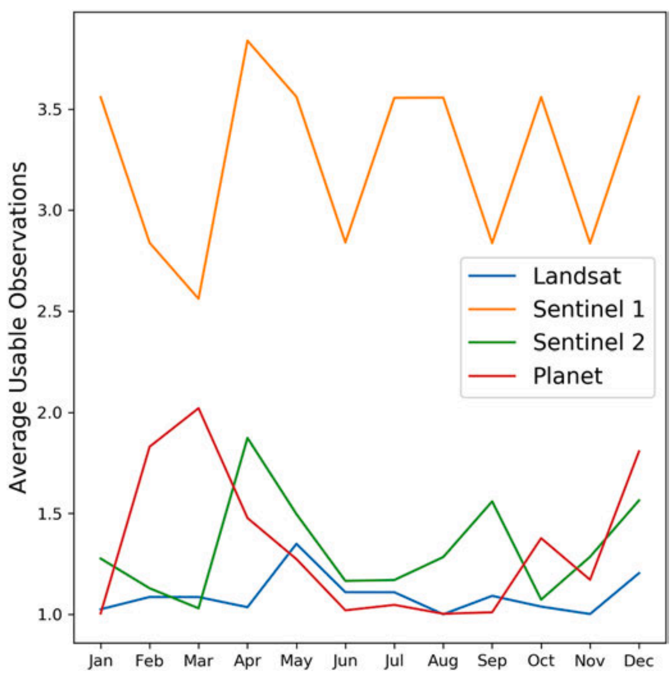

Fig. 4. Average usable (i.e. cloud-free) observations for each pixel in the study region for Sentinel-1, Sentinel−2, PlanetScope, and Landsat in 2020. Cloud-free observations for Sentinel-2 and Landsat are defined by the cloud masks delivered with the surface reflectance products.

accuracy metrics (user's, producer's, or overall accuracy) in coordinate syntax.

Bootstrap standard errors and confidence intervals can be calculated for the Sigmoid Accuracy metrics to provide approximate measures of uncertainty. The bootstrap standard error is the standard deviation of the sampling distribution derived from repetitions of steps 6–8. For each repetition, a random sample of the reference dataset is selected with replacement and used to build a bootstrap distribution of lags and F_1s for each metric (Efron and Tibshirani, 1986). There are various bootstrap confidence intervals, and we recommend the bias-corrected and accelerated (BCa) approach to account for asymmetric sampling distributions (Efron, 1987). While tests revealed the bootstrap distributions of F_1 to be normally distributed and centered on the F_1 coordinate of the Initial Delay and Level Off Point, the lag distribution was found to be slightly asymmetric in some situations. The BCa confidence intervals accounts for bias and skew in the bootstrap distribution and is therefore recommended. The standard errors of the estimates of lag and F_1 for the Initial Delay in Fig. 2 are 1.796 and 0.007, respectively, while the 95% confidence intervals are (7, 13) and (0.039, 0.065). For the Level Off Point, the standard errors are 8.670 and 0.022, while the confidence intervals are (73, 93) and (0.795, 0.830), respectively. All calculations in this section are demonstrated for the example in Fig. 2 in the supplemental Python notebook.

2.2. *Case study: Northeastern Madagascar*

In 2021 the US Forest Service and LOFM embarked on a collaboration to develop a national forest warning system in Madagascar. As part of this effort, a pilot location in northeastern Madagascar was selected for testing, which is the same area used here as a case study in system timeliness (Fig. 3). The region contains montane tropical forests interspersed with heavily modified non-forest land covers. The area is persistently cloudy and has a rainy season from October to April. Primary drivers of deforestation in the region include smallholder agriculture, mining, timber extraction, and livestock (SalvaTerra, 2017). The area is remote, has minimal formal economic activity, covers diverse terrain, and experiences relatively small (<1 ha) change events, making it an ideal location for testing the performance of a forest monitoring system.

2.3. *Data selection: Sentinel-1*

Our study area is frequently cloudy and, consequently, is a challenging setting for forest monitoring without cloud-penetrating radar data. For 2020, Sentinel-1 averaged over three times more usable observations than Landsat, Sentinel-2, or PlanetScope (Fig. 4). We therefore decided to use Sentinel-1 data as the basis for our analysis.

The Sentinel-1 constellation of C-band radar sensors have been collecting dual-polarization backscatter data since 2014 and are processed as gridded data suitable for time series analysis. This study used all available Sentinel-1 descending-orbit and Interferometric Wide (IW) swath mode Ground Range Detected (GRD) data, processed using the Sentinel-1 Toolbox available on Google Earth Engine (Gorelick et al., 2017). Data retrieved along the ascending orbit was infrequent in our study region, and the differences between orbital directions introduces variability due to differences in view angle, which is heightened by uneven terrain. Therefore, only data collected on the descending orbital pass was used for our analysis.

Further pre-processing of the Sentinel-1 data was performed on Google Earth Engine. Backscatter data were transformed linearly and topographically corrected using the radiometric slope correction described in Vollrath et al., 2020. The VV and VH polarization data were used to calculate a ratio index and all data were transformed using the Refined Lee filter (Yommy et al., 2015). Depending on the algorithm configuration (see next section), further data smoothing (e.g., "speckle" removal) was performed using a simple spatial mean or a temporal average within a multi-observation window.

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

Continuous Change Detection and Classification (CCDC) is a land cover change monitoring algorithm that was originally developed to use Landsat time series data (Zhu and Woodcock, 2014). As the name suggests, CCDC has two primary components: change monitoring using structural break detection and classification using a machine learning

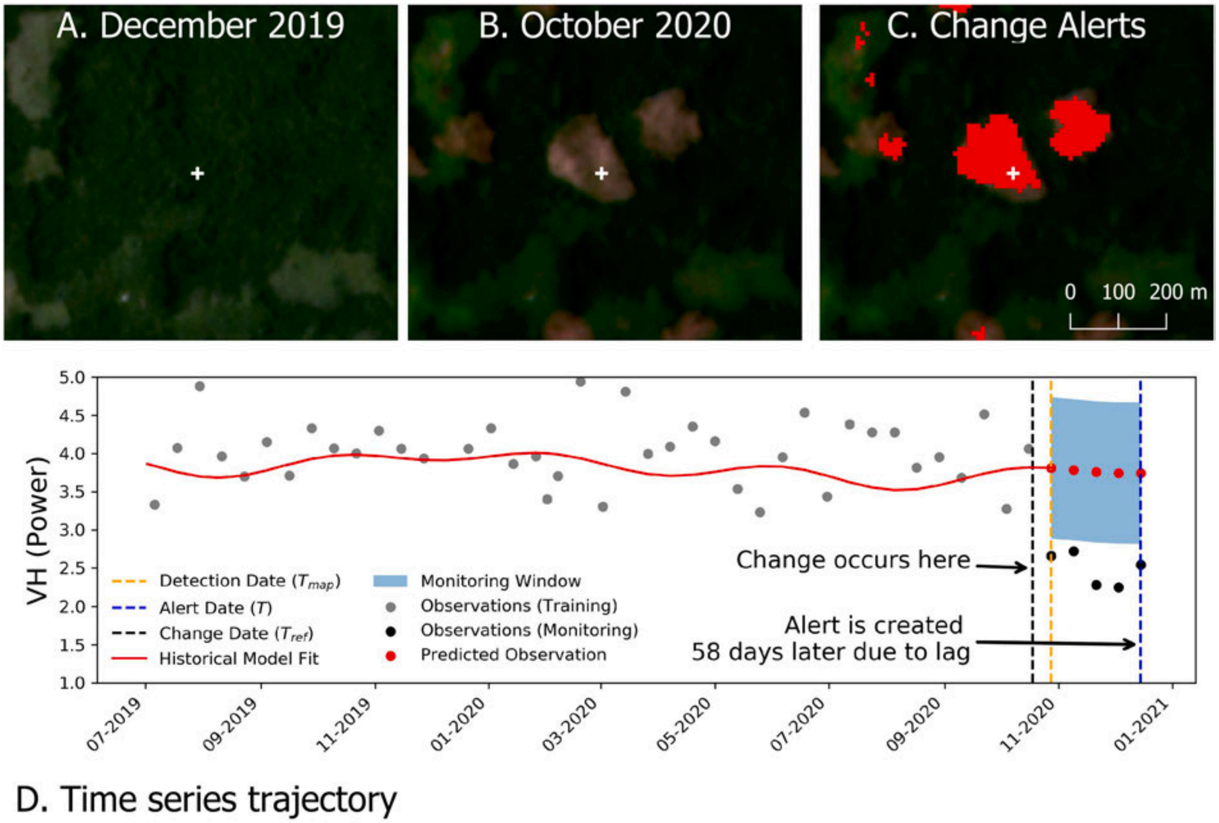

Fig. 5. An example forest change event detected by CCDC. PlanetScope Biennial Mosaics are displayed for December 2019 (A.), and October 2020 (B.). Change alerts for the year 2020 are shown in C. The VH backscatter time series for a location shown in the white crosses is shown in D. The difference between the change date and detection date is observation lag, while the difference between the detection date and alert date is algorithm lag.

Table 1
Four selected parameter configurations for CCDC. The four configurations were chosen to simulate differences in lag.

Configuration	Chi^2 Probability	Minimum Observations	Temporal Average (weeks)	Spatial Average (Pixels)	Relative Lag Properties
P1	0.95	3	None	None	Low algorithm lag
P2	0.99	6	None	None	High algorithm lag
P3	0.95	6	4	4	High algorithm, data lag
P4	0.97	4	None	3	Moderate algorithm lag

classifier. CCDC was applied in the context of near-real time forest monitoring in Tang et al. (2019) using daily MODIS data.

CCDC operates in near-real time using a statistical test within a moving window at the end of a time series (Fig. 5). In this way, CCDC progresses through time to find change and can update the data within the moving window as new acquisitions become available. CCDC monitors for change by comparing predictions, resulting from a model fit during a training period, to subsequent observations during a monitoring period. When the model residuals exceed a statistical boundary for a predefined number of consecutive observations, a change is detected. The Chi^2 probability parameter (set by the operator) determines the size of the statistical boundary, while the Minimum Observations threshold controls the number of consecutive observations needed to flag a change (see Sayler and Zanter (2020) for details on CCDC parameters). The regression model parameters are then classified using land cover training data. For this implementation, all changes away from a forest classification resulting in a decrease in VH backscatter were assigned a Forest Disturbance label.

Our application of CCDC was modified to use Sentinel-1 data. First, the multi-temporal cloud masking of CCDC could not be performed because it uses spectral bands that are unique to optical sensors. Second, the two backscatter bands and the ratio layer were used for change monitoring, whereas the original CCDC used a subset of optical bands.

We have designed our assessment framework to enable comparison between monitoring systems. To simulate a comparison between systems we developed four implementations of CCDC by varying parameters that control pre-processing of the Sentinel-1 data and change detection (Fig. 4).

The implementations are referred to here as P1, P2, P3, and P4 and were designed to simulate different types of lags (Table 1). Four selected parameter configurations for CCDC. The four configurations were chosen to simulate differences in lag. P1 is aggressive in identifying change due to a low Chi^2 Probability and Minimum Observation parameters and therefore has low algorithm lag relative to the other configurations. P2 has high algorithm lag relative to P1 as it requires twice as many observations (Minimum Observations = 6) to exceed a greater statistical boundary (Chi^2 Probability = 0.99). While P1 and P2 use all data from the descending orbital direction, P3 uses temporal composites within four-week windows and spatial averaging within a four-by-four-pixel window to reduce noise in the time series. Therefore, P3 simulates a system with high data lag. P4 has moderate settings for each parameter and does not use temporal composites. Of consequence of the aggressive change detection parameters, the P1-alert map appears “noisier” than the other configurations, which tend to detect changes in patches rather

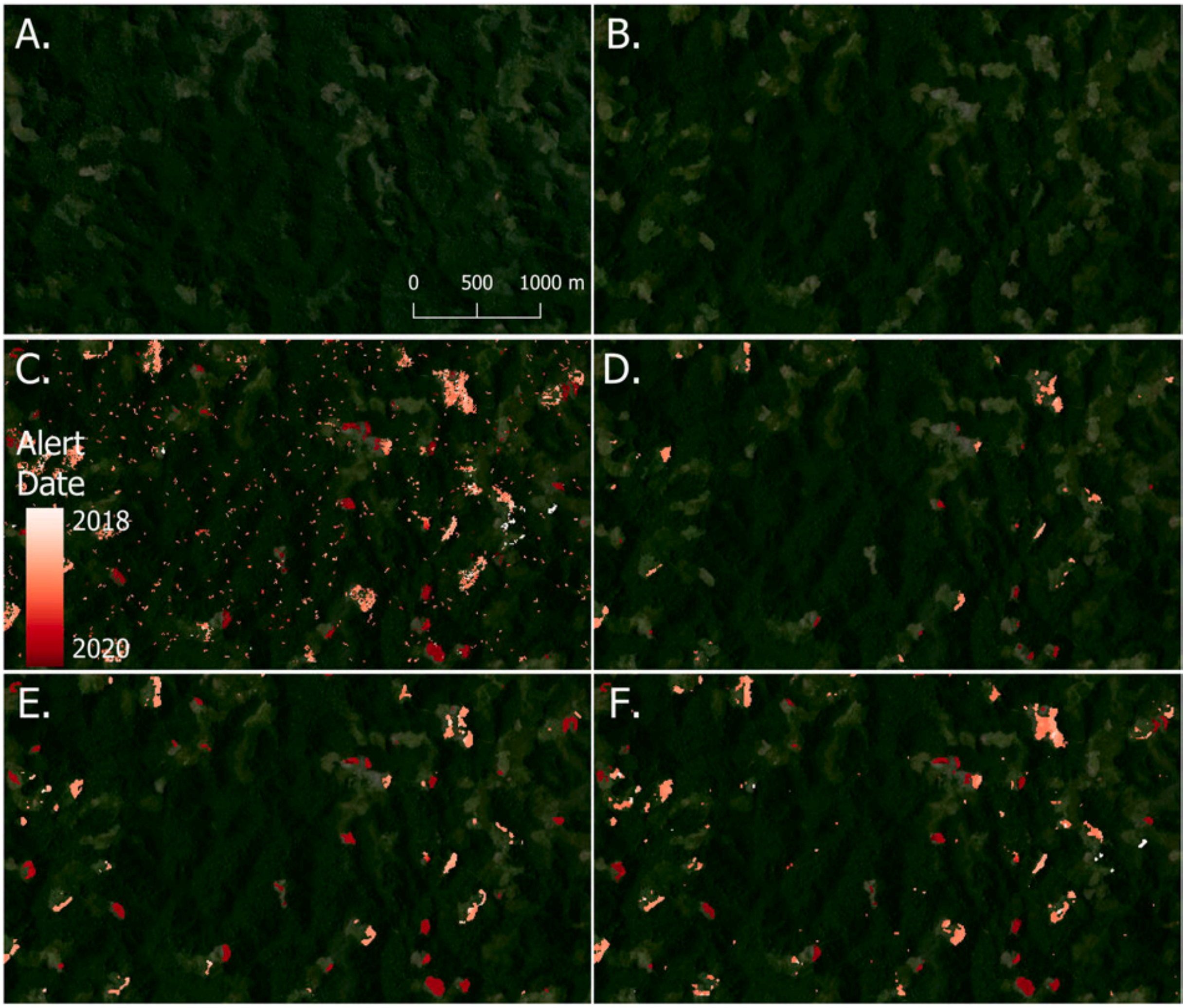

Fig. 6. Planet biennial mosaics from June 2017 (A.) and December 2019 (B.), and change alerts from P1 (C.), P2 (D.), P3 (E.) and P4 (F.).

than isolated pixels (Fig. 6)

2.5. Reference data interpretation

To perform the accuracy assessment, we randomly selected 15 3,000-ha grid cells that we treated as representative of the study area (Fig. 3). Within those areas, we manually established two exhaustive strata using daily PlanetScope imagery: disturbed at any point during the study period ("Change") and not disturbed within that period ("Not Change"). Limiting the assessment to 15 sub-regions reduced the time needed to create this stratification, while randomly selecting grid cells ensured ample spatial coverage of our study area. All visible patches of tree cover loss as small as one PlanetScope pixel were delineated. A stratification was created by reprojecting the change patches to the grid of Sentinel-1, and any pixel containing a change event was assigned to the Change stratum. A stratified sample of disturbance date observations was made within the sampled sub-regions using all available imagery, such that near-equal numbers of Change (475) and Not Change (485) samples were collected. Stratification weights were carried through subsequent calculations of accuracy.

The first date of tree cover loss (T_{ref}) was identified using all available Planet, Sentinel-2, Landsat, and Sentinel-1 imagery. One interpreter collected the entire reference dataset but evaluated each location twice: the first time to identify change patches within the study period and the second time to attribute T_{ref} to the sample units. We created a tool in Google Earth Engine for visual attribution of T_{ref}. After experimental interpretation using Sentinel-1 and daily PlanetScope imagery, we concluded that use of all available medium- and high-resolution data was needed for precise date attribution. Since the reference dataset had daily temporal resolution, we determined that the imprecision in T_{ref} would be negligible (<5 days) and we therefore did not calculate $T_{refadjusted}$. We defined the Tolerance Window as −5 to 365 days to allow for a maximum five-day discrepancy between T_{ref} and the "true" date of change.

For each sample point, the interpreter viewed two time series: VH backscatter from Sentinel-1, and a combined NDVI time series produced from surface reflectance data from Landsat, Sentinel-2, and PlanetScope. While Landsat has a coarser spatial resolution than the Sentinel-1 change alerts, we determined it acceptable to be included as reference data for disturbance date attribution since higher resolution (3 m) PlanetScope data was used for spatial delineation. The interpreter determined the disturbance date by first visually inspecting the time series trajectory for noticeable downward shifts in NDVI or backscatter, which would suggest an abrupt decrease in vegetation. When the interpreter clicked on an observation in the time series, the underlying image would load on the map, similar to the operation of TimeSync (Cohen et al., 2010). The interpreter viewed as many images as needed to identify the first date of tree loss.

In addition to the T_{ref}, each sample labeled as "Forest Change" was labeled as "Center" or "Edge" and "Stand Replacing" or "Partial

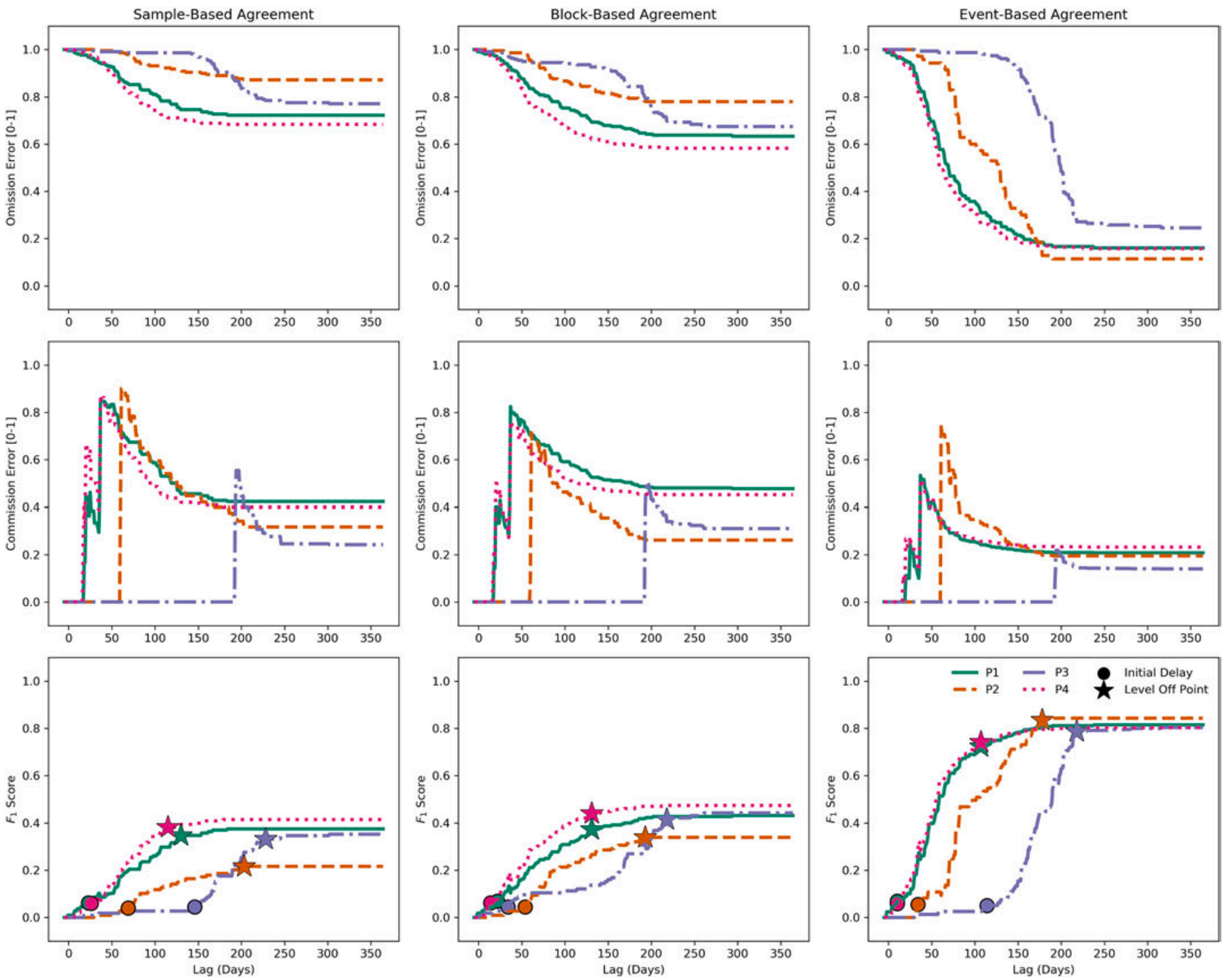

Fig. 7. Lag/accuracy curves for the four algorithm configurations (P1, P2, P3, and P4) and three spatial agreement protocols (sample, block, and event). Omission error (top), commission error (middle), and F_1 are plotted as a function of lag within a tolerance window of −5 to 365 days. Note that accuracy tends to increase with time since a change event but eventually reaches an asymptote. The Initial Delay (circles) and Level Off Points (stars) are two key moments along the lag/accuracy curves that are quantified with our assessment framework. Uncertainties are reported as standard errors and confidence intervals in Table 2.

Clearing". An "Edge" sample was defined as a sample unit in which the boundary overlapped the exterior of the corresponding change patch, and "Partial Clearing" was defined when there were visible signs of trees after the disturbance in the PlanetScope imagery. The areas of the change events associated with each sample were calculated from the reference change polygons. Our monitoring system will eventually develop into a national alert system for Madagascar, so only stand-replacing changes over one-half hectare in size were used for the assessment to be consistent with Madagascar's national monitoring and REDD+ program (BNC REDD, 2018). However, performance is explored for all samples and by different disturbance characteristics in the Discussion section "Errors of Commission and Omission".

The resulting reference dataset was applied in the Sigmoid Accuracy Assessment Framework for each parameter configuration and using sample-, block-, and event-based agreement. We used a Detection Threshold for event-based agreement of 15% ($P_t = 15\%$), which means that 15% of an event must be labeled as change by CCDC to be considered correct (Tang et al., 2019).

3. Results and discussion

3.1. Accuracy assessment

The relation of lag to F_1 describes the accuracy of forest change detection as a function of time. Fig. 7 illustrates the relationship between accuracy and timeliness: accuracy is dynamic and will usually improve with time. The accuracy at the Initial Delay date will be low because it corresponds to the lowest amount of data needed to flag a change. For CCDC, the Initial Delay date is controlled primarily by the size of the temporal moving window used to detect change in addition to data lag between Sentinel-1 observations. As time goes by, the stability of the change signal aids in its differentiation from noise in the data, enabling better identification by CCDC. Therefore, F_1 gradually increases after the Initial Delay as the algorithm can better characterize change events. Eventually F_1 will reach the Level Off Point and, given enough time, plateau.

Fig. 7 highlights four algorithm configurations, P1, P2, P3, and P4, and three spatial units for the agreement protocol. In all configurations, the Initial Delay indicated quick, inaccurate mapping of change. For

Table 2
Estimates of the (lag, F_1) coordinates and uncertainties for Initial Delays and Level Off Points. Uncertainties are expressed as bootstrap standard errors (SE) and 95% BCa confidence intervals (CI).

		Initial Delay (Lag in Days)				Initial Delay (F1)			
Configuration	Spatial Unit	Estimate	SE	CI Low	CI High	Estimate	SE	CI Low	CI High
P1	Sample	23	7.82	7	34	0.062	0.009	0.042	0.072
	Window	22	7.99	6	26	0.070	0.008	0.061	0.111
	Event	10	7.32	-2	26	0.069	0.012	0.059	0.108
P2	Sample	69	6.33	30	71	0.040	0.008	0.035	0.049
	Window	54	6.29	54	54	0.045	0.008	0.036	0.065
	Event	34	9.04	31	40	0.056	0.021	0.039	0.090
P3	Sample	146	39.56	18	154	0.045	0.007	0.039	0.056
	Window	34	37.51	18	154	0.045	0.007	0.039	0.055
	Event	114	29.90	30	138	0.050	0.011	0.040	0.077
P4	Sample	26	6.72	10	34	0.060	0.010	0.042	0.071
	Window	14	6.72	14	14	0.062	0.010	0.044	0.081
	Event	10	3.97	-2	10	0.059	0.012	0.058	0.058
		Level Off Point (Lag in Days)				Level Off Point (F1)			
Configuration	Spatial Unit	Estimate	SE	CI Low	CI High	Estimate	SE	CI Low	CI High
P1	Sample	130	13.99	107	167	0.347	0.033	0.275	0.406
	Window	131	14.43	130	184	0.373	0.032	0.374	0.410
	Event	107	15.09	83	124	0.723	0.036	0.651	0.789
P2	Sample	203	28.56	136	203	0.216	0.035	0.157	0.295
	Window	193	27.93	110	203	0.339	0.031	0.339	0.339
	Event	178	12.57	83	178	0.836	0.041	0.755	0.897
P3	Sample	228	9.89	218	246	0.332	0.037	0.263	0.407
	Window	218	14.57	172	228	0.415	0.035	0.415	0.415
	Event	218	2.21	203	218	0.787	0.028	0.732	0.837
P4	Sample	115	12.11	95	145	0.382	0.034	0.306	0.444
	Window	131	12.42	130	154	0.440	0.031	0.440	0.441
	Event	107	12.12	75	107	0.742	0.032	0.684	0.803

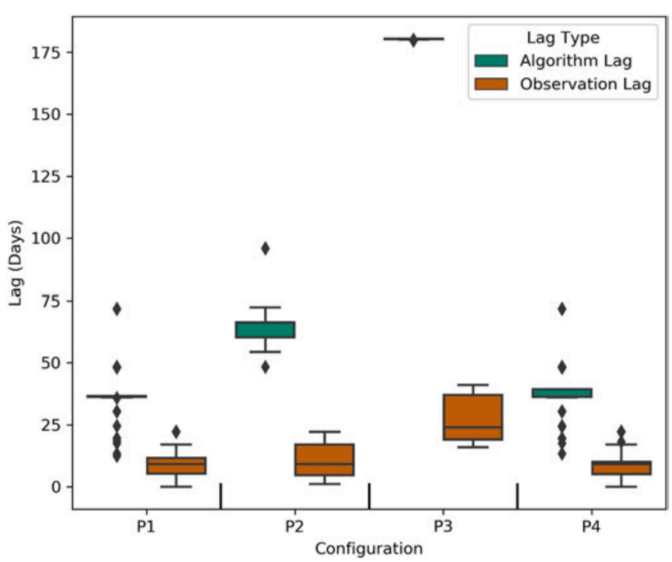

Fig. 8. Pixel-based lags for correctly identified samples. Lag was calculated as the days between the actual change date and the first observed image (Observation Lag) and the days between the initial detection and the date that an alert is created (Algorithm Lag).

example, the (lag in days, F_1) coordinates of the Initial Delay of P1 were (23, 0.06, 22, 0.07, 10, 0.07), for pixel-, block-, and event-based agreement, respectively (uncertainties are reported in Table 2). In other words, P1 was able to identify a change within around 18 days of it occurring, but omission was high. Over time, detection improved, and accuracy began to saturate, with Level Off Points of (130, 0.35, 131, 0.37, 107, 0.72). Notably, accuracy leveled off around the same time for each agreement type (107–131 days), but F_1 was considerably higher for event-based agreement (0.72) than using sample (0.35) or block (0.37) spatial units.

P2 simulated high algorithm lag by using a higher Minimum Observation threshold than P1. Fig. 8 shows the delay that can be attributed to algorithm and observation lag using pixel-based agreement, and as can be seen the observation lag for P2 was equivalent to P1 but the algorithm lag was double. Consequently, the Initial Delay coordinates were (69, 0.04, 54, 0.04, 34, 0.06) for sample-, block-, and event-based agreement, which were two to three weeks slower than P1. The Level Off Points for sample- (203, 0.22) and block-based agreement (193, 0.34) were also slower and less accurate, while the event-based agreement had a slower but more accurate Level Off Point at (178, 0.84). The discrepancy between agreement protocols highlights the importance of considering the response design when evaluating system performance.

P3 simulated an increase in observation lag relative to the other configurations by using monthly composites rather than all available observations. On average, P3 requires over a month of data after a change before it is first used by the algorithm (i.e., observation lag) (Fig. 8). Since P3 requires six consecutive observations to flag a change and each observation is a monthly composite, there was an average of 180 days of algorithm lag between the first usable observation and the change alert. As a result, the Initial Delay and Level Off Points for pixel-based agreement were (146, 0.04) and (228, 0.33), respectively. Initial Delay for block- and event-based agreement were lower than the estimated lag in Fig. 8, which suggests that surrounding pixels may have changed prior to the disturbance dates that were attributed to the original sample pixels (T_{ref}). Change events grow over time (see Tang et al., 2019), and it is likely that pixel-based change date attribution may be inconsistent with block- or event-based agreement protocols. While the date of change can be attributed for each pixel in a block or event, the process is tedious and not realistic for most assessments. Sample-based agreement alleviates the inconsistencies that arise due to evolving change events and therefore provides a more precise estimate of the Initial Delay when using pixel-based date attribution.

P4 had moderate change detection parameters compared to the other configurations. The algorithm lag was in between P1 and P2 due to a Minimum Observation threshold of four, while observation lag was limited by using all observations in the descending orbit. P4 was also designed to be less noisy than P1 or P2 by using a spatial mean as an additional method for speckle-removal. Consequently, P4 performed best under sample- and block-based agreement, having the quickest and most accurate Level Off Points.

The time-dependent performance metrics proposed here can be useful for selecting a monitoring system. Based on the above comparison, there is no situation in which a user should select P3, as it is both slower and less accurate than the others. An approach that requires pixel - or block-based agreement should choose P4, as it had the quickest Initial Delays and most accurate Level Off Points. If using event-based agreement, then P2 was shown to be most accurate, having Level Off coordinates of (178, 0.84). Low accuracy is not detrimental if the alerts are observed manually, possibly with real-time high-resolution imagery, prior to responsive action. However, if the alerts are used to automatically trigger a response, then they should be more accurate, and in this case the user would be better off selecting P2. As with the illustrated choice among different configurations of CCDC with Sentinel-1 imagery, Initial Delay and Level Off coordinates can be used to choose among any available change alert systems.

The concept that accuracy varies as a function of time is not unique to CCDC. For example, the RADD alerts use a Bayesian updating approach that "builds" probability of change across observations and an alert is only created once the probability crosses a threshold (Reiche et al., 2021, 2018). In the 2018 study, the alert user's accuracy improved from 8.8% to 93.3% for use case's with MTLs of 12 and 39 days, respectively. A second example is Fusion2, which identifies changes in a combined MODIS-Landsat time series, and had a producer's accuracy of the disturbance class that increased from 30% at a lag of 23 days to 50% after 82 days (Tang et al., 2019). Finally, an assessment of the Stochastic Continuous Change Detection (S-CCD) algorithm revealed an increase in the Detection Rate from 50% after 126 days to 75% after 500 days (Ye et al., 2021).

Just as algorithms can optimize omission or commission error, they can also be tuned to maximize timeliness. Users often must choose algorithm parameters that control change detection sensitivity, which can necessitate fewer observations to confirm a change and thereby directly influence algorithm lag and timeliness. For CCDC, two parameters that strongly influenced the timeliness of CCDC were Minimum Observations, which controls the size of the moving window used to identify change, and the Chi^2 Probability, which controls the size of the statistical boundary (Fig. 5). Using a low Minimum Observation threshold resulted in shorter Initial Delays but lower accuracies of those alerts. For example, P1 had a Minimum Observation threshold of three and had a pixel-based Initial Delay at 23 days, while P2 had a threshold of six and had an Initial Delay of 69 days.

3.2. Agreement protocol

As part of our agreement protocol, we defined the Tolerance Window parameter so that alerts would not be correctly identified unless the dates they were detected (T) were between −5 and 365 days of the change date in the reference data (T_{ref}). Similarly, Reiche et al. (2018) included samples that were first flagged by the algorithm (T_F) prior to T_{ref} as commission errors. Consequently, a shorter agreement time will result in lower accuracy and will change the shape of the relationship between lag and F_1. Fig. 9 demonstrates the (lag, F_1) coordinates for P1 using different Tolerance Windows. Note that using a lower Tolerance Window will result in a quicker Level Off Point, but at a low accuracy.

In some cases, F_1 will not saturate within the Tolerance Window, therefore the Level Off Point cannot be calculated. In such a case the system performance will continue to improve beyond the allowable time for a user, and the maximum (lag, F_1) coordinates can be used instead. For example, application of our framework for choosing a monitoring system that requires monthly latency should use a Tolerance Window of 0 to 31 days. Assuming no system performance saturates within that time, the system with the highest F_1 at 31 days should be preferred.

The choice of an appropriate spatial unit for an accuracy assessment has been debated in the literature (e.g. Stehman and Wickham, 2011). Standardized spatial units (pixels or blocks) are easy to apply in a probability-based inference framework, in which variables such as area or accuracy are estimated for a population. However, a near-real time monitoring system might only need to focus a user on the general vicinity of a change event to be useful, which is not encapsulated in traditional pixel-based approaches. Event-based agreement was recommended in the Tang framework as an alternative to standardized spatial

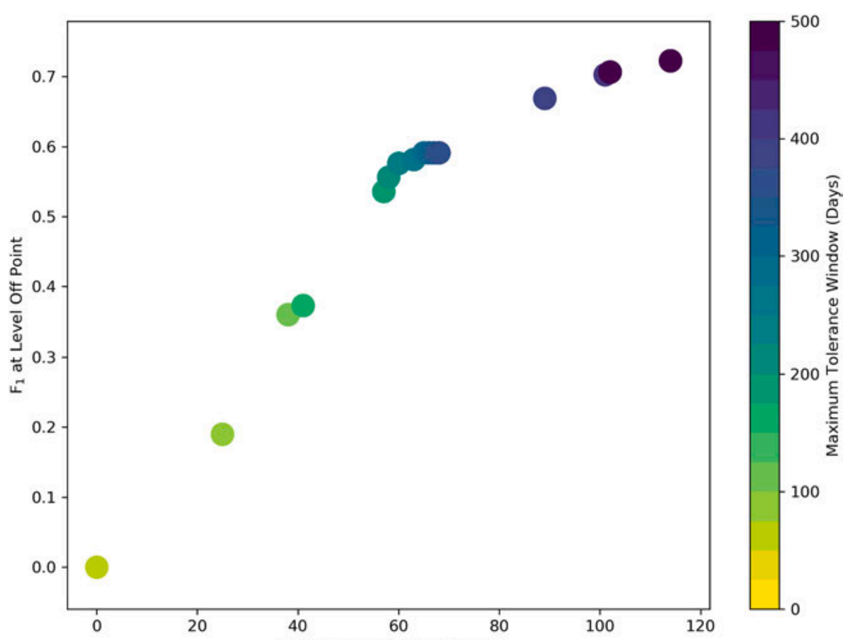

Fig. 9. The lag/accuracy coordinates of the Level Off Point for 25-day increments for the P1 algorithm configuration. Note that Tolerance Window parameter influences the calculation of timeliness metrics, and therefore these metrics should only be compared for assessments with consistent designs.

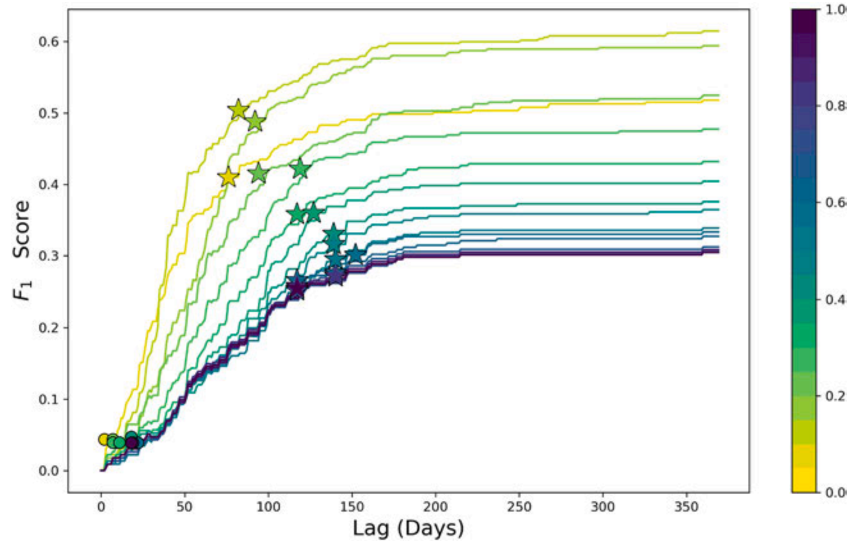

Fig. 10. The lag/accuracy curve for P1 using event-based agreement with varying Detection Thresholds. Note that F_1 at the Level Off Point (stars) tend to increase with lower Detection Thresholds, while the lag of the Initial Delay increases with Detection Thresholds.

units, but the required reference data must be collected through arduous interpretation of high-quality reference data. Statistical estimation of commission error is also complicated due to the varying size of reference polygons. While it is relatively straightforward to assess omission of events, there is no such analogy for calculating commission since it is not possible to collect reference polygons of committed change. We chose not to limit our framework to a specific spatial agreement protocol, and instead demonstrated its applicability using three different approaches.

As demonstrated in Fig. 7, omission error is lower in the assessment designs with less stringent spatial agreement requirements. Detecting an

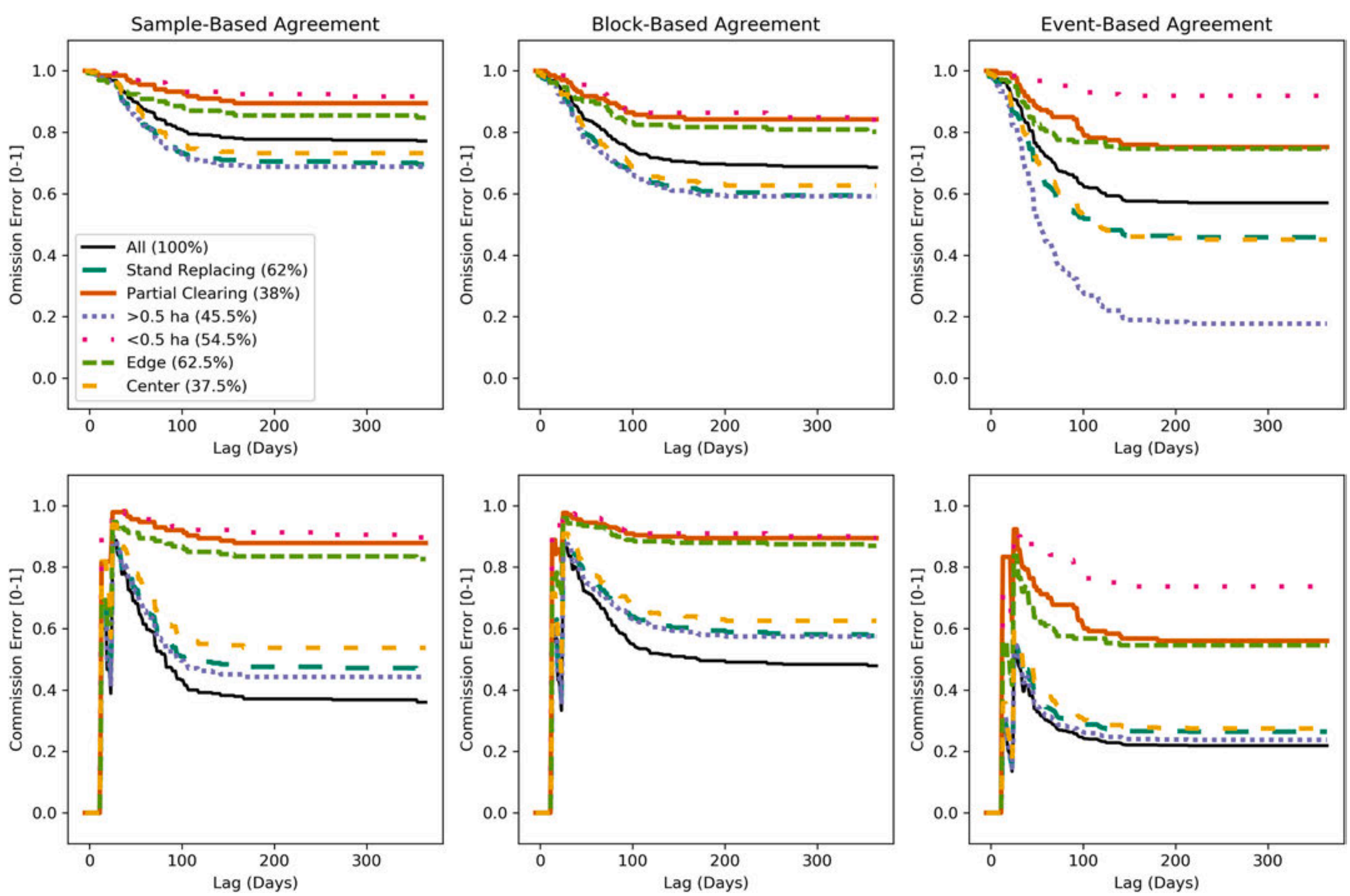

Fig. 11. Errors of omission for P4 for data filtered by three exclusive categories: stand replacing tree cover loss at the Sentinel-1 scale versus not-fully stand replacing, above or below 0.5 ha patches, and the location of the sample pixel being in the center or edge of a patch. The percentage of the reference samples that belong to each change type are listed in parenthesis in the legend.

event requires less precision in a change alert than detecting a specific pixel or block of pixels. Therefore, F_1 at the Level Off Point was consistently highest under event-based agreement. An important consideration for use of event-based agreement is the choice of a Detection Threshold parameter. For our analysis we chose a threshold of 15%, meaning that 15% of an event must be detected within the Tolerance Window to be considered correct. The Detection Threshold will influence the speed that an event is detected and will therefore alter the lag/accuracy curve and timeliness metrics, highlighting the importance in considering the agreement protocol when comparing timeliness metrics from different assessments (Fig. 10).

3.3. Errors of commission and omission

Our assessment framework uses both commission and omission error to characterize accuracy as it relates to time, which is a more holistic representation of performance than methods that rely on only omission error (Tang et al., 2019) or independent lag and accuracy metrics (Reiche et al., 2018). Commission error can be problematic for monitoring systems if the alerts instigate an automated response. Errors of commission tend to peak around the Initial Delay, since that is the minimum time the system can create an alert and, while some initial alerts will be correct, others will be false alerts (Type I errors). As changes are correctly identified the numerator increases for the calculation of User's Accuracy (the inverse of commission error) and, consequently, commission error decreases over time.

Applications that penalize commission error may not wish to use (lag, F_1) coordinates, since F_1 is calculated using both commission and omission error. Alternatively, (lag, commission) or (lag, omission) coordinates can be used for the Initial Delay and Level Off Points. For

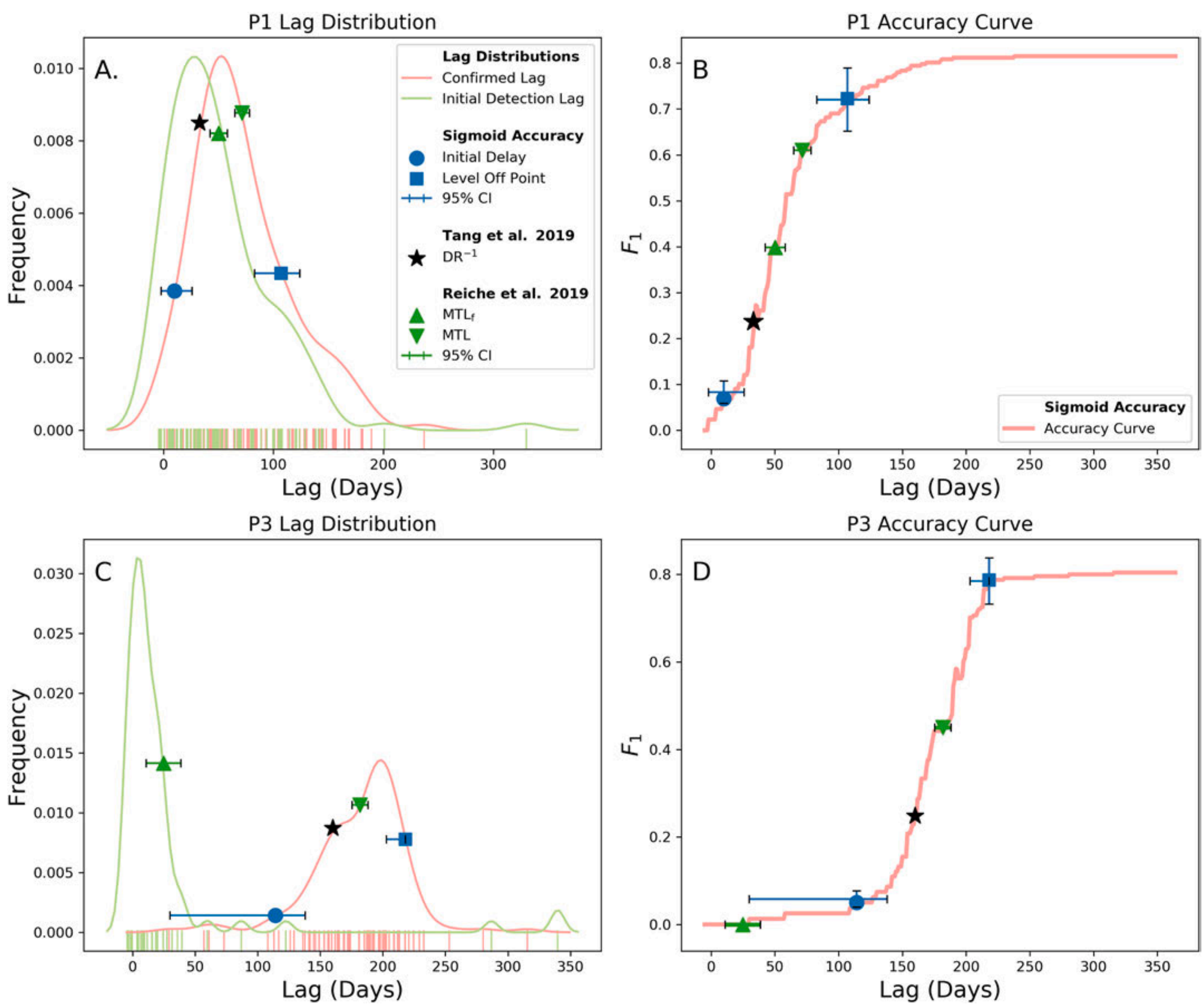

Fig. 12. Different metrics for characterizing system timeliness calculated for the P1 and P3 algorithm configurations. P1 represents a system with low data lag, while P3 has high data and algorithm lag. The Sigmoid Accuracy metrics are introduced in this research and represent the lag and accuracy at the minimum time required to create an alert (Initial Delay) and the point that accuracy stabilizes (Level Off Point). Tang et al. (2019) proposed the Detection Rate (DR) metric, defined as the proportion of events detected at a specified lag. Here we converted DR to be the lag at which DR reached a detection threshold of 50% (DR^{-1}). Reiche et al., 2015 formalized two metrics of lag averages, the mean time lag for alert confirmation (MTL) and the mean time lag for initial alert detection (MTL_F). The metrics are plotted along the kernel density estimates of lags between T_{ref} and T (Confirmation Lag) and between T_{ref} and T_F (Initial Detection Lag) in A and C and along the lag-accuracy curve in B and D. The 95% confidence interval for MTL and MTL_F are calculated from the standard error of the mean, while for the Sigmoid Accuracy metrics they are the BCa bootstrap intervals.

example, using sample-based agreement the (lag, commission) Initial Delay coordinates of P4 and P3 were (26, 0.5, 146, 0), respectively, indicating that early alerts from P3 were slow but reliable while P4 was quicker but more likely to contain commission errors. These patterns persisted as accuracy saturated at the Level Off Points, which were (115, 0.44) for P4 and (228, 0.28) for P3.

Errors of omission were common for all algorithm configurations. Sentinel-1 is sensitive to changes in surface roughness and dielectric constant, unlike lower-frequency radar sensors that are better suited to detect changes in forest structure, or spectrometers that measure changes in reflectance. As a result, numerous studies have shown Sentinel-1 to be less accurate for detecting forest change, which may cause only slight changes in roughness if the subsequent land cover contains woody vegetation (Carrasco et al., 2019; Shimizu et al., 2019). Disturbance in our study region did not typically result in a dramatic change in land use (e.g., conversion to settlement or agriculture), and thus the change signal is weak in the Sentinel-1 data. Inherent noise in the data further reduces the change signal, which is amplified by the hilly topography of our study region.

The performance of change detection is also dependent on characteristics of the change event (e.g., patch size and disturbance intensity). Omission of change events over one-half hectare in size was 22% lower than events under one-half hectare using pixel-based agreement, and 74% lower under event-based agreement (Fig. 11). The minimum mapping unit used by Madagascar for REDD+ reporting is half a hectare, and P4 had (lag, omission, commission, F_1) coordinates of (107, 0.26, 0.26, 0.74) for events larger than one-half hectare, suggesting that our CCDC and Sentinel-1 monitoring system is capable of accurately detecting change events relevant to national reporting in 107 days. Although created using a different assessment design and therefore not directly comparable, omission and commission error reported in Madagascar's proposed reference level was 42% and 25%, respectively (Bureau National de Coordination REDD+, 2017). Furthermore, omission of stand-replacing disturbances was 28% lower than non-stand replacing (i.e., degradation events). A similar pattern was seen for center pixels, which had 17% less omission for pixel- and block-based agreement (sample location within a patch was not relevant for event-based agreement since any pixels within the patch counted towards detection of the event). Edge pixels had Level Off Coordinates of (106, 0.14) while center pixels were (107, 0.32), suggesting that patch edges were found in similar time but with less accuracy.

A thorough comparison to other timely forest monitoring systems was beyond the scope of this study. However, a simple comparison of our reference data to GLAD humid tropic alert system found GLAD to omit 79.5% of the stand-replacing disturbances in our reference dataset for patches of at least one Landsat pixel ($30m^2$). The (lag, omission) coordinates of P4 at the Level Off Point and when applying the same response design was (107, 0.76), slightly better than GLAD. While it is expected that a locally tuned system will achieve better performance than a global system, the high omission of the GLAD alerts demonstrates the difficulties in detecting change in our study region.

3.4. Alternative metrics of system timeliness

Alternative measures of system timeliness include the mean time lag (MTL), mean time lag of initial detection (MTL_F), and Detection Rate (DR) (Boschetti et al., 2010; Campagnolo et al., 2021; Reiche et al., 2021, 2018, 2015; Tang et al., 2019; Ye et al., 2021). Using our reference samples, we calculated each metric to understand how they differ from the Sigmoid Accuracy metrics. We calculated DR^{-1}, defined as the inverse of DR or the lag at which a specified proportion of change events (P_{DR}) are detected, to be in comparable units (lag in days) to the other metrics. Fig. 12 shows the Initial Delay, Level Off Point, MTL, MTL_F, and DR^{-1} ($P_{DR} = 50\%$) for P1 and P3. Each metric is overlaid on their corresponding lag distributions (Fig. 12A, C) and in (lag, F_1) coordinates on the lag-accuracy curve (Fig. 12B, D).

The Sigmoid Accuracy metrics are located on the tail ends of lag-accuracy space, while MTL is a measure of central tendency. Our two critiques of lag average metrics are that they are calculated independent of spatial accuracy, and they do not encapsulate the full range of lags observed in the reference data. As can be seen in Fig. 12, accuracy increases after MTL for 35 and 36 days in P1 and P3, respectively. MTL has traditionally been reported alongside map accuracy metrics like omission and commission error, but those metrics reflect the end of the lag-accuracy curve, not the point of MTL. For example, the MTL of P1 is 72 ($\pm$ 6.7) days, and the omission error without consideration of lag is 16%. However, the (lag, omission) coordinates of MTL is (72 days, 51%), which reflects significantly worse spatial accuracy at the time of average lag. This distinction highlights the ambiguity in reporting independent temporal and spatial performance metrics.

MTL_F is calculated here as the average time between T_{ref} at the first time the system alerts of a possible change (T_F). P1 has low algorithm lag (Fig. 8) and, consequently, minimal difference between MTL and MTL_F. However, P3 uses monthly composite data and requires six observations to confirm a change, and as a result MTL is 157 days later than MTL_F. Our lag metrics are based on the date of confirmed change according to CCDC (T), however T occurs after multiple observations of "possible" changes. Theoretically, the lag-accuracy curve and corresponding metrics could be calculated using T_F rather than T, which would result in a quicker Initial Delay.

DR^{-1} is defined as the lag in which 50% of the change events ($P_{DR} =$ 50%) are identified and is calculated using the entire reference dataset, not the subset of samples correctly identified. Therefore, DR^{-1} can potentially be located anywhere in the lag distributions and accuracy curves shown in Fig. 12. Unlike MTL and MTL_F, DR^{-1} can encapsulate the full range of lags in a reference dataset, albeit with different P_{DR} parameters. Unlike MTL and MTL_F, DR and DR^{-1} directly link accuracy to timeliness since they are calculated using the producer's accuracy for specified lag times. However, none of these metrics incorporate commission error as they relate to time, and therefore no single metric can fully represent spatial-temporal performance on their own.

Unlike the alternatives, the Sigmoid Accuracy metrics describe the full range of lag-accuracy space. However, Fig. 12 demonstrates that MTL and DR^{-1} may complement Initial Delay and the Level Off Point as standardized performance metrics. MTL provides a measure of central tendency for observed lags, while DR^{-1} can be parameterized according to user needs. We recommend future investigation into MTL and DR^{-1} as additional metrics to be included in Sigmoid Accuracy. To overcome ambiguity in reporting temporal and spatial accuracy independently, we recommend reporting MTL, MTL_F, DR, and DR^{-1} in (lag, accuracy) coordinate syntax.

3.5. Data source and algorithm

There are general characteristics of a forest monitoring system that make it appropriate for near-real time monitoring. The system should be able to: 1) dynamically update with new observations; 2) use an algorithm that can detect recent change, and; 3) be deployed on a system that can handle daily data ingestion and algorithm processing. For this reason, we piloted a system that uses CCDC with Sentinel-1 on GEE to demonstrate the concepts of timeliness.

CCDC operates online, which means that change is identified by moving forward through a time series of data. When a new observation is added to the time series, the model updates and the moving window of observations used to flag change is moved forward by one increment. In this manner, CCDC can detect change alerts with every newly acquired observation. Similarly, BFAST-Monitor uses an online change detection test based on the moving sum of residuals (MOSUM) the identifies change in a window at the end of a time series (Verbesselt et al., 2010). While not based on regression, the classification approaches used by RADD and GLAD can also update with new observations in an online manner (Hansen et al., 2016; Reiche et al., 2021).

Furthermore, CCDC has been demonstrated to be more sensitive to changes at the end of time series than other structural break detection algorithms (Bullock et al., 2020a). “Offline” algorithms that look at all the data at once to identify a break can be weaker indicators of change at the beginning or an end of a time series. Algorithms more likely to omit changes at the end of a time series will require more time to level off in accuracy and may also require more time before a first change alert. For these reasons, it is important to consider algorithm characteristics for applying them for near-real time monitoring.

To our knowledge, this study marks the first attempt to use CCDC with Sentinel-1 radar data. Sentinel-1 has numerous advantages for monitoring cloudy regions: it penetrates clouds, has a frequent (6 to 12-day) repeat time, and is freely available as a standardized data product. For our study region, there is an average of three times the frequency of observations as either Sentinel-2 or Landsat, meaning the data lag is three times lower (Fig. 4).

In previous research, CCDC was applied to daily MODIS data with a spatial resolution of 250 m (Tang et al., 2019). The use of MODIS data enabled timely alerts but at the expense of high omission of events smaller than a hectare. While MODIS data is acquired daily, it can often be at large view angles which necessitate data filtering. Tang et al. (2019) optimized timeliness at the expense of accuracy and spatial resolution. Similar issues arise when CCDC was applied to VIIRS data with a spatial resolution of 500 m (Tang et al., 2020). Moving forward, however, MODIS’s frequency of observation and timeliness might be matched at higher spatial resolution if effective fusion of data from multiple sensors can be achieved or with daily PlanetScope imagery.

Applications of forest monitoring may require the optimization of omission error, commission error, or timeliness. For example, Olofsson et al. (2020) demonstrated the importance of reducing omission error when using forest change maps for small area estimation using a sampling design. Other applications might require low commission error, as is the case for alert systems that are used to deploy costly field campaigns. Similarly, an approach may seek to optimize timeliness, as was the case in Tang et al. (2019). Other applications may optimize a combination of accuracy and timeliness, such as the GLAD alert system, which “provides forest disturbance updates as observed, with the objective of providing a conservative change detection system absent of errors of commission” (Global Land Analysis and Discovery, 2016; Hansen et al., 2016).

3.6. Applications for Sentinel-1 CCDC

A system’s timeliness determines its ability to achieve monitoring objectives, and a strength of our assessment framework is that it can be used to evaluate the ability of a system to achieve monitoring objectives. In practice, the error assessment process outlined here can be used by individuals that are choosing a system for their specific purpose.

Lag requirements are location and application specific, but comparison can be made to selected examples from the literature. For example, INSA (2000) evaluated detection requirements for fighting fires in Europe and concluded that there was a maximum of 15 minutes after a fire is started for an alert to be useful. Obviously, a system such as CCDC that has a minimum alert time of 22 (P1; pixel- and block-based agreement) days would not be useful for this context. Similarly, it would not be helpful in identifying the opportunistic illegal logging events described in Pendleton, 1998 that occur in one or a few days. However, it could potentially identify larger logging operations such as the one studied in Uhl and Vieira (1989) that operated for approximately three weeks.

3.7. Considerations

Our assessment framework focuses only on the first change detected at any given location. Repeat disturbances do occur within the short timeframes of interest under real-time systems (Schroeder et al., 2012). In the context of continuous monitoring, any location can theoretically alternate between being correct, omission error, and commission error over time. Though our online implementation of CCDC does not allow for retrospective removal of declared changes, our calculation of F_1 over time could accommodate systems where such revision occurs. However, the inclusion of sites with multiple real disturbances creates ambiguity in the assessment process. Because of this ambiguity and the fact that detection of the initial perturbation is often more important in real-time detection systems than precisely sequencing repeated disturbances, we recommend elimination or replacement of multiple-disturbance reference sites from the reference dataset. This measure, while improving comparability across systems and settings, may reduce applicability for situations where a second disturbance – for instance pile-burning after harvest – may be of primary interest.

We have proposed a generalized approach for calculating the Level Off Point that uses the point of maximum curvature to define the time that accuracy begins to stabilize. However, there may exist a better calculation for the Level Off Point that is not influenced by the Tolerance Window parameter and would therefore be more comparable across different assessment designs. Calculation of a curve asymptote for discrete data is an unresolved concept, and our choice to use the point of maximum curvature was chosen for simplicity. An alternative metric to using a curve knee is the (x, y) coordinates at the end of the Tolerance Window. However, this location would penalize systems that level off quickly. Our calculation of the Level Off Point is one generalizable solution but we suggest alternative approaches as a subject for future research.

4. Conclusion

The speed and quality of forest monitoring capabilities will likely continue to improve as new sensors, improved fusion methods, and better distribution/processing systems become available. As monitoring options multiply, it is important to establish a common framework by which monitoring systems may be systematically compared and considered in light of operational needs. We propose focusing on two key moments of accuracy as a function of time: the point at which detection is first available and the point beyond which additional time returns minimal improvement in accuracy. Consideration of the coordinates for these points should allow users to choose an optimal system on the basis of accuracy and timeliness given local priorities and tolerances. Our demonstration using Sentinel-1 data and CCDC (the first such use of CCDC with radar data) showed that this system can help prioritize among monitoring systems differing only on the basis of one or two algorithm parameters. Moving forward, reporting of lag-sensitive accuracies (“Sigmoid Accuracy”) as standard validation metrics would facilitate objective selection and adoption of remote sensing methods in monitoring projects intended to respond quickly to ecosystem disturbance.

CRediT authorship contribution statement

Eric L. Bullock: Conceptualization, Methodology, Validation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Sean P. Healey:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Funding acquisition, Project administration. **Zhiqiang Yang:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Formal analysis. **Rasmus Houborg:** Resources, Data curation, Writing – review & editing. **Noel Gorelick:** Software, Resources. **Xiaojing Tang:** Conceptualization, Methodology, Writing – review & editing. **Carole Andrianirina:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by National Aeronautics and Space Administration grant 80HQTR21T0020. We are grateful for the Google Earth Engine and Python communities for the software used in this research, and for the European Space Agency, United States Geological Survey, National Aeronautics and Space Administration, and Planet for the satellite imagery. The authors report no conflicts of interest for this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2022.113043.

References

Abercrombie, S.P., Friedl, M.A., 2016. Improving the consistency of multitemporal land cover maps using a hidden Markov model. IEEE Trans. Geosci. Remote Sens. 54, 703–713. https://doi.org/10.1109/TGRS.2015.2463689.
Arévalo, P., Olofsson, P., Woodcock, C.E., 2020. Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series: a test methodology for REDD+ reporting. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2019.01.013.
BNC REDD, 2018. Définition d'un Système de classification de l'Utilisation et de l'Occupation des Terres (UOT) et la définition de la Forêt à Madagascar. Antananarivo, Madagascar.
Boschetti, L., Roy, D.P., Justice, C.O., Giglio, L., 2010. Global assessment of the temporal reporting accuracy and precision of the MODIS burned area product. Int. J. Wildland Fire 19, 705–709.
Bullock, E.L., Woodcock, C.E., Holden, C.E., 2020a. Improved change monitoring using an ensemble of time series algorithms. Remote Sens. Environ. 238 https://doi.org/10.1016/j.rse.2019.04.018.
Bullock, E.L., Woodcock, C.E., Olofsson, P., 2020b. Monitoring tropical forest degradation using spectral unmixing and Landsat time series analysis. Remote Sens. Environ. 238 https://doi.org/10.1016/j.rse.2018.11.011.
Bullock, E.L., Woodcock, C.E., Souza, C., Olofsson, P., 2020c. Satellite-based estimates reveal widespread forest degradation in the Amazon. Glob. Chang. Biol. 26, 2956–2969. https://doi.org/10.1111/gcb.15029.
Bureau National de Coordination REDD+, 2017. Forest Reference Emission Level of Madagascar for Reducing Emissions from Deforestation and Forest Degradation.
Campagnolo, M.L., Libonati, R., Rodrigues, J.A., Pereira, J.M.C., 2021. A comprehensive characterization of MODIS daily burned area mapping accuracy across fire sizes in tropical savannas. Remote Sens. Environ. 252, 112115.
Carrasco, L., O'Neil, A.W., Daniel Morton, R., Rowland, C.S., 2019. Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. Remote Sens. https://doi.org/10.3390/rs11030288.
Chinchor, N., 1992. MUC-4 Evaluation Metrics. https://doi.org/10.3115/1072064.1072067.
Cohen, Warren B., Yang, Zhiqiang, Kennedy, Robert, 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation. Remote Sens. Environ. 114 (12), 2911–2924. https://doi.org/10.1016/j.rse.2010.07.010.
Cohen, W.B., Yang, Z., Healey, S.P., Kennedy, R.E., Gorelick, N., 2018. A LandTrendr multispectral ensemble for forest disturbance detection. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.11.015.
Congalton, Russell G., Green, Kass, 2019. Assessing the Accuracy of Remotely Sensed Data Principles and Practices, Third edition. CRC Press.
Eckerstorfer, M., Vickers, H., Malnes, E., Grahn, J., 2019. Near-real time automatic snow avalanche activity monitoring system using sentinel-1 SAR data in Norway. Remote Sens. https://doi.org/10.3390/rs11232863.
Efron, B., 1987. Better bootstrap confidence intervals. J. Am. Stat. Assoc. 82, 171–185.
Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 54–75.
Finer, M., Novoa, S., Weisse, M.J., Petersen, R., Mascaro, J., Souto, T., Stearns, F., Martinez, R.G., 2018. Combating deforestation: from satellite to intervention. Science. https://doi.org/10.1126/science.aat1203.
Francini, S., McRoberts, R.E., Giannetti, F., Mencucci, M., Marchetti, M., Scarascia Mugnozza, G., Chirici, G., 2020. Near-real time forest change detection using PlanetScope imagery. Eur. J. Remote Sens. https://doi.org/10.1080/22797254.2020.1806734.
GFOI, 2020. Integration of Remote-Sensing and Ground-Based Observations for Estimation of Emissions and Removals of Greenhouse Gases in Forests. Methods Guid. from Glob. For. Obs. Initiat, p. 3.
Global Land Analysis & Discovery, 2016. GLAD Forest Alerts [WWW Document]. URL. https://glad.umd.edu/dataset/glad-forest-alerts (accessed 5.10.21).
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2017.06.031.
Hansen, M.C., Krylov, A., Tyukavina, A., Potapov, P.V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., Moore, R., 2016. Humid tropical forest disturbance alerts using Landsat data. Environ. Res. Lett. 11, 034008 https://doi.org/10.1088/1748-9326/11/3/034008.
Holmgren, P., Thuresson, T., 1998. Satellite remote sensing for forestry planning—a review. Scand. J. For. Res. https://doi.org/10.1080/02827589809382966.
Holmgren, P., Anton, P., Nillson, L., Nordahl, G., Wiklund, F., 1989. Datatillgänglighet—Hur ofta kan molnfria SPOT-bilder registreras.
Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., Vogelmann, J.E., 2010. An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. Remote Sens. Environ. 114, 183–198. https://doi.org/10.1016/j.rse.2009.08.017.
INSA, 2000. FUEGO Instrument Design, Prototype, Construction and Validation. Madrid, Spain.
Instituto Nacional de Pesquisas Espaciais, 2000. Deforestation estimates for the Brazilian Amazon. In: Deforestation Estimates in the Brazilian Amazon.
Justice, C.O., Giglio, L., Korontzi, S., Owens, J., Morisette, J.T., Roy, D., Descloitres, J., Alleaume, S., Petitcolin, F., Kaufman, Y., 2002. The MODIS fire products. Remote Sens. Environ. https://doi.org/10.1016/S0034-4257(02)00076-7.
Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr - temporal segmentation algorithms. Remote Sens. Environ. 114, 2897–2910. https://doi.org/10.1016/j.rse.2010.07.008.
Li, Z., Cihlar, J., Nadon, S., 2000. Satellite-based detection of Canadian Boreal forest fires: development and application of the algorithm. Int. J. Remote Sens. https://doi.org/10.1080/01431160050144956.
Li, Zhangqing, Kaufman, Yoram J., Ichoku, Charles, Fraser, Robert, Trishchenko, Alex, Giglio, Louis, Jin, Ji-Zhong, Yu, Xinwen, 2001. A review of AVHRR-based active fire detection algorithms: principles, limitations, and recommendations. SPB Academic Publishing, pp. 199–225.
Mardian, J., Berg, A., Daneshfar, B., 2021. Evaluating the temporal accuracy of grassland to cropland change detection using multitemporal image analysis. Remote Sens. Environ. 255, 112292.
Martinis, S., Plank, S., Ćwik, K., 2018. The use of Sentinel-1 time-series data to improve flood monitoring in arid areas. Remote Sens. https://doi.org/10.3390/rs10040583.
Matson, M., Stephens, G., Robinson, J., 1987. Fire detection using data from the NOAA-N satellites. Int. J. Remote Sens. 8 (7), 961–970. https://doi.org/10.1080/01431168708954740.
Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sens. Environ. 129, 122–131. https://doi.org/10.1016/j.rse.2012.10.031.
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.
Olofsson, P., Arévalo, P., Espejo, A.B., Green, C., Lindquist, E., McRoberts, R.E., Sanz, M.J., 2020. Mitigating the effects of omission errors on area and area change estimates. Remote Sens. Environ. 236, 111492 https://doi.org/10.1016/j.rse.2019.111492.
Pasquarella, V.J., Bradley, B.A., Woodcock, C.E., 2017. Near-real-time monitoring of insect defoliation using Landsat time series. Forests 8. https://doi.org/10.3390/f8080275.
Pendleton, M.R., 1998. Taking the forest: the shared meaning of tree theft. Soc. Nat. Resour. https://doi.org/10.1080/08941929809381060.
Pérez-Sánchez, D., Montes, M., Cardona-Almeida, C., Vargas-Marín, L.A., Enríquez-Acevedo, T., Suarez, A., 2021. Keeping people in the loop: socioeconomic valuation of dry forest ecosystem services in the Colombian Caribbean region. J. Arid Environ. https://doi.org/10.1016/j.jaridenv.2021.104446.
Potapov, P., Siddiqui, B.N., Iqbal, Z., Aziz, T., Zzaman, B., Islam, A., Pickens, A., Talero, Y., Tyukavina, A., Turubanova, S., Hansen, M.C., 2017. Comprehensive monitoring of Bangladesh tree cover inside and outside of forests, 2000-2014. Environ. Res. Lett. 12, 2000–2014. https://doi.org/10.1088/1748-9326/aa84bb.
Reiche, J., Verbesselt, J., Hoekman, D., Herold, M., 2015. Fusing Landsat and SAR time series to detect deforestation in the tropics. Remote Sens. Environ. 156, 276–293. https://doi.org/10.1016/j.rse.2014.10.001.
Reiche, J., Hamunyela, E., Verbesselt, J., Hoekman, D.H., Herold, M., 2018. Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. Remote Sens. Environ. 0–1. https://doi.org/10.1016/j.rse.2017.10.034. In review.
Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.-E., Odongo-Braun, C., Vollrath, A., Weisse, M.J., Stolle, F., Pickens, A., Donchyts, G., Clinton, N., Gorelick, N., Herold, M., 2021. Forest disturbance alerts for the Congo Basin using Sentinel-1. Environ. Res. Lett. https://doi.org/10.1088/1748-9326/abd0a8.
Reymondin, L., Jarvis, A., Pérez-Uribe, A., Touval, J., Argote, K., Coca, A., Rebetez, J., Guevara, E., Mulligan, M., 2012. A Methodology for Near Real-Time Monitoring of Habitat Change at Continental Scales using MODIS-NDVI and TRMM. https://doi.org/10.13140/RG.2.2.15618.99520.
Salvador, S., Chan, P., 2004. Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms. In: Proceedings - International Conference on Tools with Artificial Intelligence. ICTAI. https://doi.org/10.1109/ICTAI.2004.50.
SalvaTerra, 2017. Analyse des moteurs de déforestation et de dégradation dans les écorégions des forêts humides de l'Est et des forêts sèches de l'Ouest de Madagascar. Paris, France.

Satopää, V., Albrecht, J., Irwin, D., Raghavan, B., 2011. Finding a "kneedle" in a haystack: detecting knee points in system behavior. Proc. Int. Conf. Distrib. Comput. Syst, 166–171. https://doi.org/10.1109/ICDCSW.2011.20.

Sayler, K., Zanter, K., 2020. Land Change Monitoring, Assessment, and Projection (LCMAP) Continuous Change Detectionand Classification (CCDC) Algorithm Description Document (ADD). Sioux Falls, SD.

Schroeder, T.A., Wulder, M.A., Healey, S.P., Moisen, G.G., 2012. Detecting post-fire salvage logging from Landsat change maps and national fire survey data. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2011.10.031.

Shimabukuro, Y., Duarte, V., Anderson, L., Valeriano, D., Arai, E., Freitas, R., Rudorff, B. F., Moreira, M., 2007. Near real time detection of deforestation in the Brazilian Amazon using MODIS imagery. Ambient. e Agua-An Interdiscip. J. Appl. Sci. 1, 37–47. https://doi.org/10.4136/ambi-agua.4.

Shimabukuro, Y.E., dos Santos, J.R., Formaggio, A.R., Duarte, V., Rudorff, B.F.T., 2012. The Brazilian Amazon monitoring program: PRODES and DETER projects. Glob. For. Monit. Earth Obs. 354.

Shimizu, K., Ota, T., Mizoue, N., 2019. Detecting forest changes using dense Landsat 8 and Sentinel-1 time series data in tropical seasonal forests. Remote Sens. https://doi.org/10.3390/rs11161899.

Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. Int. J. Remote Sens. 35, 4923–4939. https://doi.org/10.1080/01431161.2014.930207.

Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2011.06.007.

Tang, X., Bullock, E.L., Olofsson, P., Estel, S., Woodcock, C.E., 2019. Near real-time monitoring of tropical forest disturbance: new algorithms and assessment framework. Remote Sens. Environ. 224, 202–218. https://doi.org/10.1016/j.rse.2019.02.003.

Tang, X., Bullock, E.L., Olofsson, P., Woodcock, C.E., 2020. Can VIIRS continue the legacy of MODIS for near real-time monitoring of tropical forest disturbance? Remote Sens. Environ. https://doi.org/10.1016/j.rse.2020.112024.

Twele, A., Cao, W., Plank, S., Martinis, S., 2016. Sentinel-1-based flood mapping: a fully automated processing chain. Int. J. Remote Sens. https://doi.org/10.1080/01431161.2016.1192304.

Tyukavina, A., Stehman, S.V., Potapov, P.V., Turubanova, S.A., Baccini, A., Goetz, S.J., Laporte, N.T., Houghton, R.A., Hansen, M.C., 2013. National-scale estimation of gross forest aboveground carbon loss: a case study of the Democratic Republic of the Congo. Environ. Res. Lett. 8 https://doi.org/10.1088/1748-9326/8/4/044039.

Uhl, C., Vieira, I.C.G., 1989. Ecological impacts of selective logging in the Brazilian Amazon: a case study from the Paragominas Region of the State of Para. Biotropica. https://doi.org/10.2307/2388700.

van Rijsbergen, C.J., 1979. Information Retrieval. Butterworths, London.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115. https://doi.org/10.1016/j.rse.2009.08.014.

Vollrath, Andreas, Mullissa, Adugna, Reiche, Johannes, 2020. Angular-based radiometric slope correction for Sentinel-1 on google earth engine. Remote Sens. 12 (11) https://doi.org/10.3390/rs12111867.

Woodcock, C.E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wynne, R., 2008. Free access to Landsat imagery. Science 320, 1011–1012. https://doi.org/10.1126/science.320.5879.1011a.

Woodcock, C.E., Loveland, T.R., Herold, M., Bauer, M.E., 2020. Transitioning from change detection to monitoring with remote sensing: a paradigm shift. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2019.111558.

Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., Hermosilla, T., 2018. Land cover 2.0. Int. J. Remote Sens. 39, 4254–4284. https://doi.org/10.1080/01431161.2018.1452075.

Ye, S., Rogan, J., Zhu, Z., Eastman, J.R., 2021. A near-real-time approach for monitoring forest disturbance using Landsat time series: stochastic continuous change detection. Remote Sens. Environ. 252, 112167.

Yommy, A.S., Liu, R., Wu, A.S., 2015. SAR image despeckling using refined lee filter. In: Proceedings - 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics. https://doi.org/10.1109/IHMSC.2015.236. IHMSC 2015.

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sens. Environ. 144, 152–171. https://doi.org/10.1016/j.rse.2014.01.011.

Zhu, Z., Woodcock, C.E., Olofsson, P., 2012. Continuous monitoring of forest disturbance using all available Landsat imagery. Remote Sens. Environ. 122, 75–91. https://doi.org/10.1016/j.rse.2011.10.030.

Zhu, Z., Zhang, J., Yang, Z., Aljaddani, A.H., Cohen, W.B., Qiu, S., Zhou, C., 2020. Continuous monitoring of land disturbance based on Landsat time series. Remote Sens. Environ. 238, 111116 https://doi.org/10.1016/j.rse.2019.03.009.