The rise of big data and supporting technologies in keeping watch on the world's forests

R. TAYLOR, C. DAVIS, J. BRANDT, M. PARKER, T. STÄUBLE and Z. SAID

World Resources Institute, 10 G St. NE #800, Washington, DC 20002

Email: rod.taylor@wri.org, crystal.davis@wri.org, john.brandt@wri.org, meaghan.parker@wri.org, ts@heimatstudio.net, zuraidah.said@wri.org

SUMMARY

Technology-driven advances in the gathering, processing and delivery of big data are making it easier to monitor forests and make informed decisions over their use and management. This paper first describes how innovations in remote sensing and cloud computing are enabling generation of geospatial data more often, at lower cost and in more user-friendly formats. Second, it describes the evolution of systems and technologies to trace forest products, and agricultural commodities linked to deforestation, from source to final use. Third, it reviews the potential for emerging data mining technologies such as natural language processing, web scraping and computer vision to support forest policy analysis and augment geospatial data gathered through remote sensing. The paper gives examples of how these technologies are being used and may be used in the future to monitor and respond to deforestation, fire and natural disasters, improve governance by enabling faster and more comprehensive analysis of social networks, policies and regulations, and increase traceability and transparency within supply chains.

Keywords: forests, deforestation, geospatial, traceability, data-mining

L'essor des grandes données et des technologies les soutenant dans la surveillance des forêts du monde

R. TAYLOR, C. DAVIS, J. BRANDT, M. PARKER, T. STÄUBLE et Z. SAID

Les avancées poussées par la technologie dans le rassemblement, le traitement et la distribution des grandes données rendent la surveillance des forêts plus aisée, tout comme les prises de décision averties sur leur utilisation et leur gestion. Ce papier décrit tout d'abord comment les innovations dans le sensoriel à distance et dans l'informatique en nuage aident à créer des données géo-spatiales plus fréquemment, à moindre coût et dans des formats plus confortables à l'usage. De plus, il décrit l'évolution des systèmes et des technologies pouvant tracer les produits forestiers et les matières premières agricoles associées à la déforestation, de la source à leur utilisation finale. Il analyse ensuite le potentiel que détiennent les technologies émergeantes prospectrices de données telles que le traitement du langage naturel, le grattage web et la vision par ordinateur pour soutenir l'analyse de la politique forestière et augmenter les données géo-spatiales recueillies par télédétection. Ce papier donne des exemples de la manière dont ces technologies sont utilisées et comment elles pourraient être utilisées dans le futur pour gérer et répondre à la déforestation, les incendies et les catastrophes naturelles, pour améliorer la gestion en facilitant une analyse plus rapide et complète des réseaux sociaux, des politiques et des règles, et pour augmenter le traçage et la transparence au sein des chaînes d'approvisionnement.

El auge de los macrodatos y las tecnologías de apoyo para la vigilancia de los bosques del mundo

R. TAYLOR, C. DAVIS, J. BRANDT, M. PARKER, T. STÄUBLE y Z. SAID

Los avances tecnológicos en la recolección, el procesamiento y la transmisión de macrodatos están facilitando el monitoreo de los bosques y la adopción de decisiones informadas sobre su utilización y gestión. En este artículo se describe, en primer lugar, cómo las innovaciones en materia de teledetección y computación en la nube facilitan la generación de datos geoespaciales con mayor frecuencia, a menor costo y en formatos más fáciles de utilizar. En segundo lugar, se describe la evolución de los sistemas y tecnologías con los que dar seguimiento a los productos forestales y los productos agrícolas vinculados a la deforestación, desde su origen hasta su uso final. En tercer lugar, se examina el potencial de las nuevas tecnologías de minería de datos, como el procesamiento de lenguajes naturales, la extracción de datos de sitios web (*web scraping*) y la visión artificial, para apoyar el análisis de las políticas forestales y aumentar los datos geoespaciales recolectados por teledetección. El artículo proporciona ejemplos de la forma en que se están utilizando estas tecnologías y como podrían utilizarse en el futuro para monitorear y enfrentarse a la deforestación, los incendios y los desastres de amenazas naturales, mejorar la gobernanza mediante un análisis más rápido y completo de las redes sociales, las políticas y los reglamentos, y aumentar la transparencia y la capacidad de dar seguimiento en las cadenas de suministro.

INTRODUCTION

Big data, which involves computational methods that rely on the computing scale of cloud resources, and supporting technologies are increasingly being used to keep watch on the world's forests and enable better decision-making over their use and governance(Chen, Mao, & Liu, 2014). These technologies are being used to monitor the biophysical structure of forests, to ensure traceability and transparency within supply chains, and to analyze and improve forest policy and governance.

Within the field of forest monitoring, advances in remote sensing and cloud computing (the use of networks of remote servers hosted on the Internet to store, manage, and process data) are making it possible to monitor changes in forest cover and condition, as well as the extent of fires and the impacts of natural disasters more cost-effectively and more frequently than forest patrols and surveys are able to. In addition to remote sensing, drones and mobile technology are increasingly being used to monitor forests at local scales, often in combination with satellite and remote sensing data. Big data is also transforming traceability and transparency efforts within supply chains through sensor networks, genetic analysis, and smart labels. These technologies are used to track the chain of custody along the supply chain and to identify the taxonomy or geographic provenance of raw or processed materials in a product. Improved traceability in supply chains provides a means for business and other stakeholders to verify if the wood or agricultural ingredients in a product are responsibly sourced. Finally, big data methods are enabling researchers to analyze legislative and policy texts and social and news media data to improve policies and governance systems. This paper provides examples of how the latest technology is being used in these three application fields, the impact it is having, and what might be possible with further technological development on the horizon.

FOREST MONITORING

Historically, researchers documented the extent of and condition of forests – canopy cover, tree size, species, biodiversity, soil carbon content or seedling density – by boots-on-the ground surveys. At the national level, countries monitored their forests through site-based sample plots as part of a national forest inventory. These field-based efforts were resource-intensive and best suited to the scale of individual forest management units, protected areas, or a limited sample of plots across a country. Consequently, national inventories tended to be done infrequently, often with patchy coverage of remote forests that are not easily accessed by road or river. At a global scale, the FAO has conducted the Global Forest Resources Assessment at five-yearly intervals since 1948 to provide national statistics on forest cover extent and change (FAO 2015).These efforts rely primarily on statistics reported by countries and are thus dependent on the frequency and accuracy with which individual countries conduct their forest inventories or update forest-relevant statistics.

In recent years, advances in remote sensing and cloud computing have created a whole new array of options for large-scale forest monitoring and field work. These technologies have enabled better detection of forest change, more frequently, over larger areas, at less cost and with easier communication channels, such as the presentation of data in the form of geospatially-explicit maps that can be accessed online. At the same time, the advent of geographic positioning systems (GPS), and technology-enabled ground patrols and forest inventories, has allowed field staff to record more detailed coordinate points for their observations and upload those data into geographic information systems (GIS). This generates richer data in support of local forest management, as well as providing means to ground-truth and refine automated systems for interpretation, visualization and analysis of satellite data at global, continental or national scales.

Remote sensing first emerged with the use of cameras mounted on planes to take aerial photographs as early as World War I and has transitioned to a mix of airborne and satellite-borne imagery in recent decades. Airborne instruments – sensors attached to planes, or cameras mounted on drones – are still used today to capture detailed information about a specific forest area at higher resolution than can be achieved from space. Airborne light detection and ranging (LiDAR) sensors can capture detailed information about the physical structure of forests (Asner, 2009) at a resolution of one meter 1 m to up to ten meters, which is detailed enough to see individual tree crowns and map tree species distribution (Baldeck *et al*. 2015).

For satellite-based remote sensing, a major breakthrough occurred in 2008, when the U.S. Geological Survey opened all data from its Landsat satellite to the public for free (Wulder & Coops, 2014). permitting large-scale analyses through time back to 1972. Many previous mapping efforts had utilized freely available coarse resolution MODIS satellite data, which ranges from 250 to 1000 meters in resolution. Suddenly, Landsat offered 30-meter resolution data—almost 70 times better than MODIS—permitting much finer-scale monitoring, systematically and at global scale. Landsat became the "go-to" source of imagery for mapping forest extent and change.

Satellite imagery spatial resolution and availability continue to improve. In 2013, the European Commission and European Space Agency (ESA) decided to openly license data from the Sentinel satellites (European Space Agency, 2013), complementing Landsat with freely available, 10-meter data, as well as radar satellites that can see through cloud cover, smoke, and haze (Reiche *et al*. 2016). An increasing number of commercial satellite companies (e.g., Planet, TerraSar) offer high spatial resolution data (under 3 meters) that—while costly for large-scale systematic analyses—can be valuable for validation, calibration, and verification.

In the early days of satellite imagery analysis (starting in the 1980s), experts visually interpreted the images and delineated forest extent and deforestation by hand. For example, the annual deforestation monitoring system in Brazil (known as PRODES) still heavily relies on expert interpreters to manually inspect imagery to identify forest changes. By contrast, machine learning technologies rely on learning algorithms within computers that build mathematical models based on sample data, known as training data, and use this to interpret imagery without humans having to write explicit programs on how they perform such tasks (Bishop 2006). Since the early 2000s, innovative machine learning algorithms have facilitated automatic mapping of forest extent, changes, and values, producing results faster and more consistently than what can be done by human interpreters. Cloud computing platforms (e.g., FAO SEPAL, Google Earth Engine, Amazon Web Services) enable these algorithms to process large volumes of imagery cheaply. Google Earth Engine, for example, combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities.

These advances mean forests can be consistently characterized and systematically monitored over large geographic areas. The University of Maryland's ground-breaking high-resolution maps of annual tree cover change were the hallmark of a new era of global monitoring of forests from space (M.C. Hansen *et al*. 2013). Other pioneer products include the global and pantropical maps of above-ground woody biomass density from (Saatchi *et al*. 2011) and (Baccini *et al.* 2012). They also enable detection of change in near-real time (see Table 1).

As forest monitoring technology has evolved, so too has the demand to make the resulting information public. Once accessible only through paper maps (or not at all), forest monitoring data has become widely available through online geoportals and databases that are simple for non-experts to use. The launch of the Global Forest Watch platform in 2014 was notable in making the University of Maryland's global spatial forest monitoring data accessible to the public for free in easy-to-understand and dynamic maps, charts, and graphs.

Corporations and academics increasingly work in collaboration to improve forest monitoring methods and transfer expertise to government institutions. This has resulted in a dramatic improvement in national forest management capacity over the last decade. For example, the MapBiomas effort in Brazil involves leading researchers and technology companies working together to produce annual land use and land cover maps. Official government data produced via the TerraClass program of the Brazilian Space Agency and Agricultural Ministry only covers the Legal Amazon and is not published every year. MapBiomas uses an automated algorithm processed in the cloud to process satellite imagery and publish land cover maps each year for the entire country. The MapBiomas team includes members of the Brazilian government as expert reviewers and strives to transfer lessons learned to government institutions. Beyond Brazil, five other Latin American countries now operate near real time alerting systems, and globally, a dozen countries have adapted the global University of Maryland annual tree cover loss product to their national context. Many more countries, such as Suriname, through its National Forest Monitoring System, are using some form of satellite imagery analysis as part of their periodic national forest inventories and/or forest reference emission levels.

Many prospects for remote sensing monitoring systems with increased accuracy, spatial resolution, and temporal frequency are on the horizon. Higher resolution optical images will enable detection of fine-scale changes indicative of forest degradation rather than outright loss of tree cover (Fagan & DeFries 2009). Operational radar data from Sentinel-1 will enable detection of forest disturbances even through cloud cover. NASA's new spaceborne lidar instrument (GEDI), mounted on the International Space Station in early 2019, will map biomass and forest structure from space, enabling more sophisticated approaches for quantifying forest carbon. Expansion in cloud computing capacity will enable more imagery to be processed more quickly. More advanced machine learning algorithms, (e.g., neural networks) should enable more accurate monitoring, and possibly prediction, of forest change, though the applications of these methods have thus far been primarily limited to high-resolution imagery,

TABLE 1 *systems detecting near-real time forest change*

System	geographic coverage	spatial resolution	Update frequency
University of Maryland GLAD alerts – (Matthew C Hansen *et al*., 2016; Reiche, Hamunyela, Verbesselt, Hoekman, & Herold, 2018; Reiche, Verhoeven, *et al*., 2018).	30 degrees North to 30 degrees South	30x30 meters	Weekly
Real-Time System for Detection of Deforestation (DETER) (Shimabukuro, dos Santos, Formaggio, Duarte, & Rudorff, 2016).	Brazilian Amazon	250x250 meters	Monthly
Terra-I (Reymondin *et al*., 2012)	Whole of Latin America + tropics	250x250 meters	Weekly
Sistema de Alerta de Desmatamento (SAD) (De Souza, Hayashi, & Veríssimo, 2008).	Brazilian Amazon	250x250 meters	Monthly
Fire Information for Resource Management System (FIRMS) (Davies, Ilavajhala, Min Minnie Wong, & Justice, 2009).	Global	375x375 meters	Daily

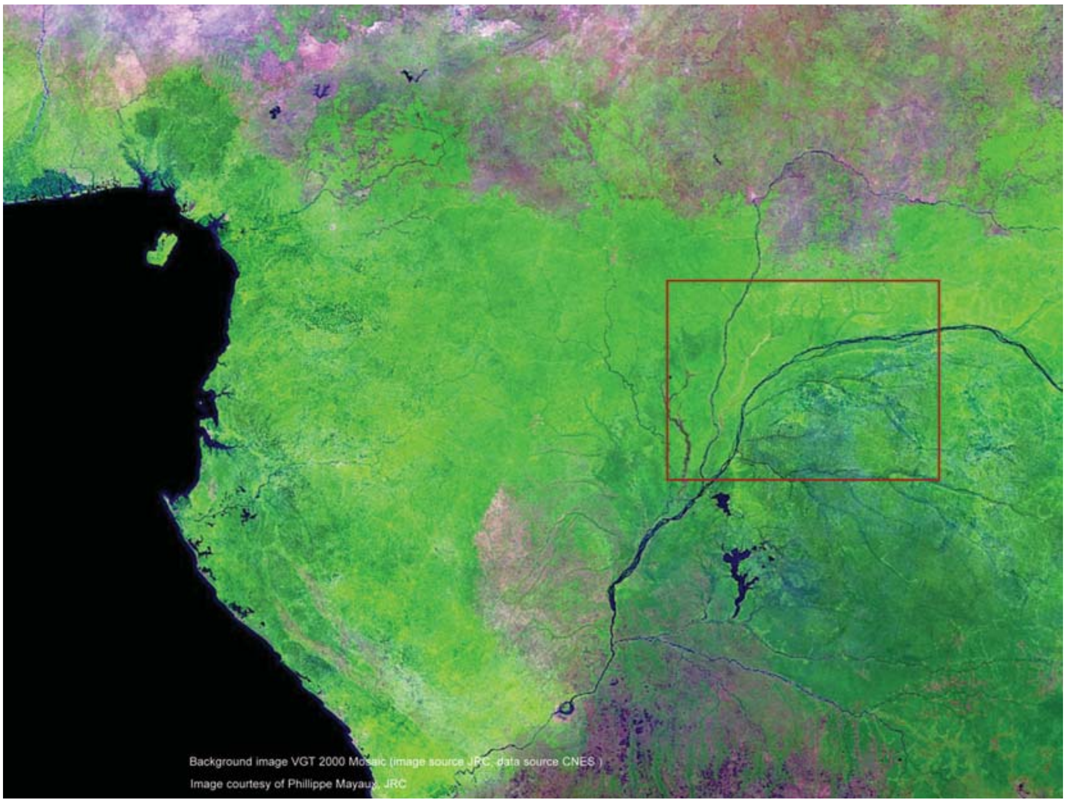

Image 1 - This cloud free mosaic of the Central African Forest Basin was assembled by the Joint Research Centre from daily images acquired by the European VGT sensor on board the SPOT satellite processed and distributed by the Flemish Technological Research Institute VITO.
The image shows the vast size of Central Africa's forests. It covers 2 million km^2 accounts for 22% of the World's humid tropical forests and contains the World's only habitats for the great apes. It is also home to around 40 million people.

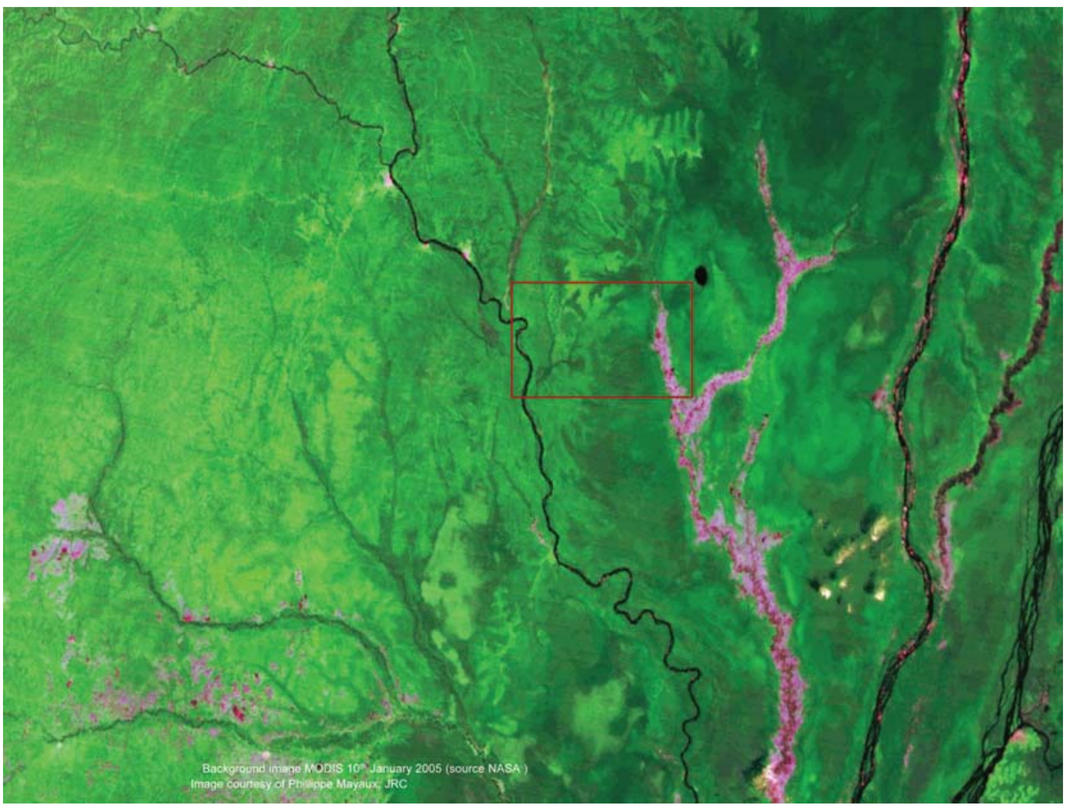

Image 2 - This detailed image (250 meter resolution) of the Central African Basin was acquired by the MODIS sensor on the US Terra satellite. The image shows the Sangha river (centre) and the Ubangi and Congo rivers to the right. The pink "river" is seasonally flooded grassland along the smaller Likouala river. There are clear signs of forest clearance and degradation around the towns such as Ouesso on the Sangha river and Mbandaka on the Congo. The impact of major roads such as the highway linking Ouesso with the coast can also begin to be seen at this resolution.

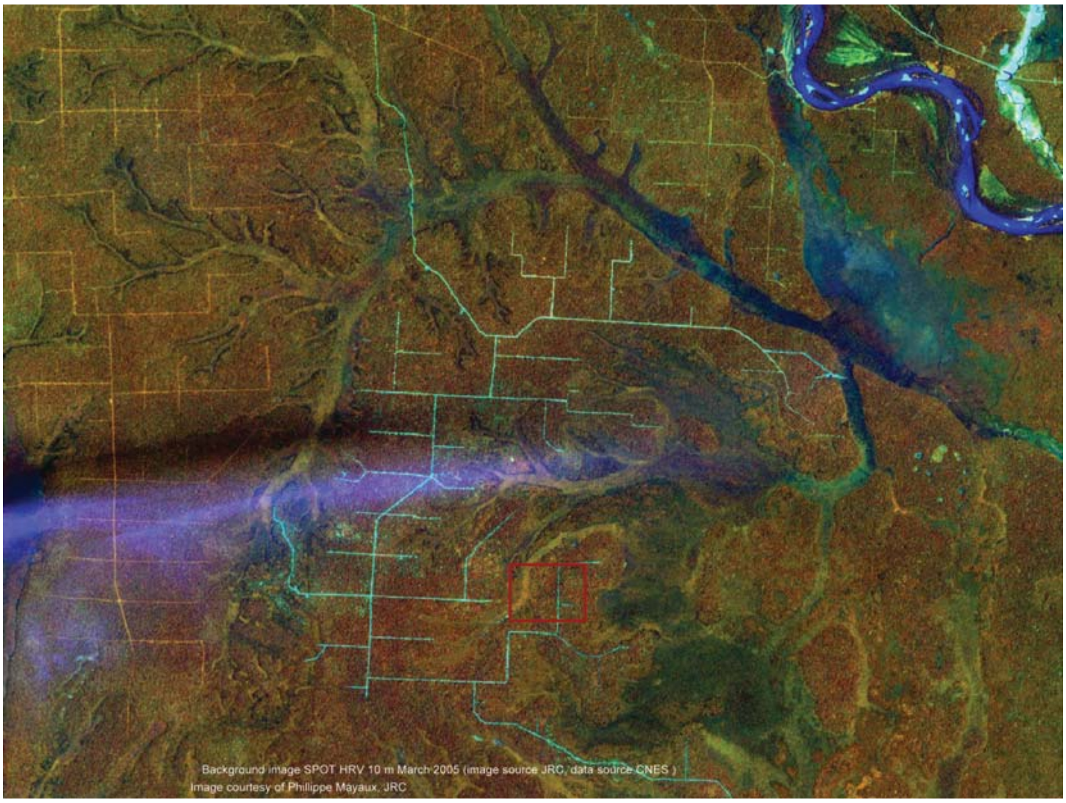

Image 3 - At resolutions of 10 meters, as with this image acquired by the European SPOT satellite, the true nature of the "undisturbed" forest begins to emerge. Both abandoned logging roads (orange) and new logging roads (blue) can be accurately mapped. Although no longer used for commercial timber exploitation the abandoned logging roads do provide access for poachers hunting for bush-meat including the great apes. The roads also provide access for less intensive, but no less destructive, timber extraction by illegal loggers.

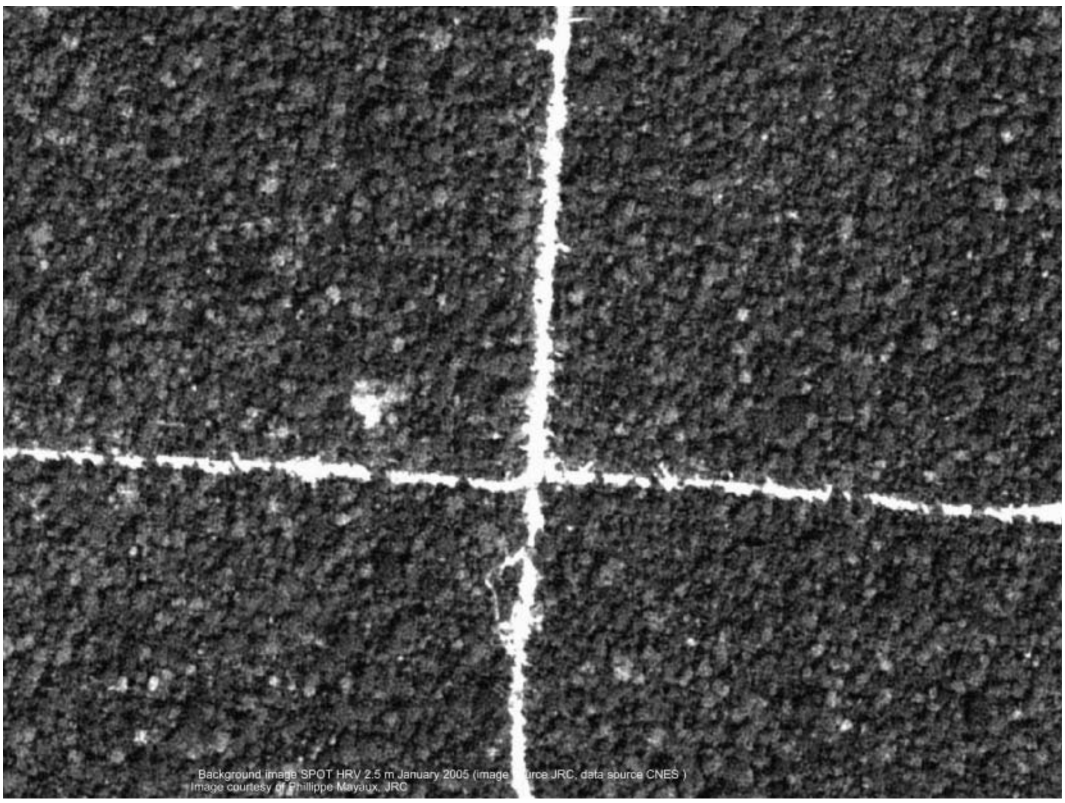

Image 4 - Using fine resolution imagery – in this case 2.5 metre resolution from the SPOT satellite, this image from the European Commission's Joint Research Facility enables measurement of the width of logging roads, and identifies the extraction of even individual trees (the white holes in the grey intact forest canopy).

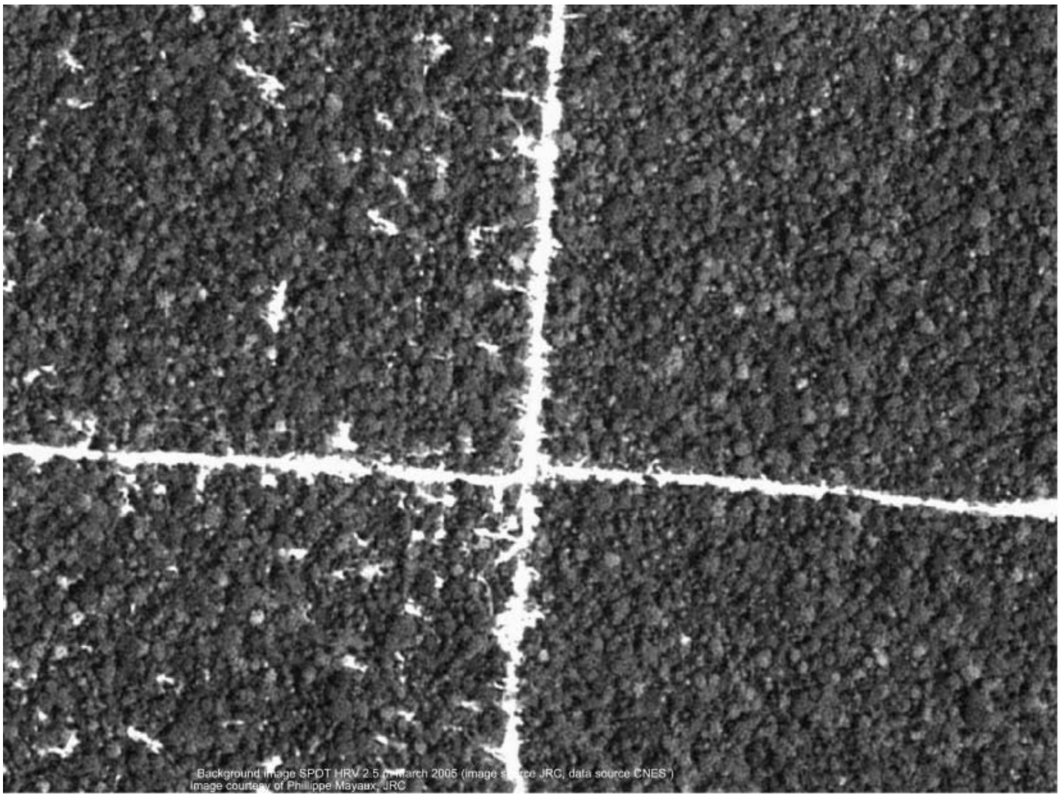

Image 5 - By using different images over time (this image was taken two months after the previous) an observer can determine overall rates of timber extraction, as well as locations – information which can help determine compliance with the terms under which any logging company has been granted a permit to work a given timber concession.

which is computationally infeasible at large geographic scales (Ma *et al*. 2019). Mobile data collection systems, such as Open Data Kit and CyberTracker, will help to unite ground-based perspectives with remote sensing data.

New data mining technologies also have huge potential to augment and automate the analyses of data collected through remote sensing, sensor networks or field measurements. Historically, analysis of such data depended on supervised techniques. These involved manual classification of geographic regions and plots to "train" algorithms to replicate human classification, or the prior formulation of hypotheses about what characteristics of a satellite image are indicative of a specific type of vegetation or disturbance. For example, identifying forest fires with remote sensing would typically require manual delineation of fire extent in thousands of images and empirical research to formulate the relationship between spectral signatures, rainfall patterns, slope, and other biophysical variables and fire disturbance. Data mining techniques can reduce the need for prior classification of data sets or prediction of causal relationships. For instance, one academic study applied the "fuzzy C-means clustering algorithm" to identify forested regions impacted by natural disasters or fires with 98.8% accuracy across varied geographic contexts without human-labelled training data (Singh & Singh 2018). Other academic studies have demonstrated the accuracy of data mining methods in modelling biomass and carbon storage relative to methods based on allometric equations (Carlos R Sanquetta *et al*. 2015, Carlos Roberto Sanquetta, Wojciechowski, Paula, Corte, & Rodrigues, 2013). Advances in clustering methodology may allow for faster and more accurate unsupervised classifications of remote sensing, sensor network, and biometric data. One such advancement is spectral clustering, which groups observations into clusters based on similarity metrics between low-dimensional mathematical representations of variability known as principal components. While spectral clustering consistently outperforms previous clustering approaches, it was not until 2018 that spectral clustering was computationally efficient enough to handle remote sensing data (Dhanachandra, Manglem, & Chanu, 2015; Shaham *et al*. 2018; Tung, Wong, & Clausi, 2010; Zhang & You, 2017). Spectral clustering is likely to replace traditional clustering methods in forest monitoring applications soon, increasing the accuracy of unsupervised approaches and further reducing data needs.

New data mining methods can also help reduce bias from seasonal, biometric, and cultural differences between geographic regions in remote sensing models, which are typically trained in geographies where training data is available and may then be applied to different geographies (Xie, Jean, Burke, Lobell, & Ermon, 2016). Conditional generative adversarial networks (cGANs) are a type of neural network that learns how to transform between domains of images, such as those taken in different seasons or regions. These have been used to generate ground-level views from satellite imagery (Deng, Zhu, & Newsam, 2018), identify road networks (Q. Shi, Liu, & Li, 2018), generate building footprint information from satellite imagery (Y. Shi, Li, & Zhu, 2019), and learn transformations between geographies (Kniaz, 2018). Unsupervised approaches to learning generalizable features from satellite imagery have also shown promise in mitigating geographic biases (Jean *et al*. 2018). Taken together, these advances in machine learning and data mining have significant potential to improve accuracy in forest monitoring tasks where training data is expensive to generate or where such data exists only in specific geographies.

In addition to satellites, wireless sensor networks are increasingly deployed in forests to record sounds, temperature or movement for purposes such as fire control, prevention of illegal logging and biodiversity monitoring. Integrated data mining techniques, where data mining methods such as clustering or anomaly detection are built into the sensors themselves, can improve efficiency, reduce energy use, and lower data throughput requirements in such networks (Czúni & Varga, 2014; Saoudi, Euler, Bounceur, & Kechadi, 2016). Geolocation and social media data within individual electronic devices such as mobile phones also have huge future potential to support accurate real time detection of wildfires, floods, earthquakes, wildlife migrations and the spread of invasive species (Daume, 2016; Middleton, Middleton, & Modafferi, 2014; Tanev, Zavarella, & Steinberger, 2017). Social media activity on wildfires, for example, is highly corelated to where and when fires occur, and can thus be used to provide early warnings of fire outbreaks (Boulton, Shotton, & Williams, 2015).

Applying forest safeguards in supply chains

Tracking the movement of materials through supply chains is often critical for quality control, safety and financial discipline along the chain. It is also useful in distinguishing products sourced illegally or implicated in deforestation from those that come from well-managed forests or farms. Governments can help promote the application of good chain of custody practice by integrating requirements for adequate product flow controls in regulations and compliance monitoring.

Supply chain traceability requires careful documentation of the path that product ingredients take as they move from the farm or forest to the end customer, including any mixing or transformation along the way. A traditional chain of custody system is literally a "paper trail" documenting the flow of a specific batch of materials along a supply chain. However, advances in information technology, internet access and connectivity, GPS tracking systems and product scanning devices, mean a modern chain of custody system can live mostly online.

Labeling technologies in chain of custody systems facilitate rapid collection of large amounts of data that can be electronically time-stamped and cross-checked against records made at other checkpoints to detect and deter tampering (ITTO 2012). Labels containing nano-molecules or imprinted with bar codes can be scanned electronically. Others, such as RIFID labels, can be accessed using radio signals. Increasingly, data logging devices support data capture in the field for immediate or subsequent transfer to online databases. These devices can be handheld devices or integrated in existing machinery such as trucks and harvesting machines. Such technologies are more efficient than manual methods because they reduce the need for error-prone manual information transfer. Validation is also supported through the metadata automatically collected with each reporting event (e.g. who reported via the user-account, when the information was collected via the time-stamp, and where the information was collected via the GPS module in the device) (Baldwin, Markowitz, Koparova, Gerardu, & Zaelke, 2015).

Satellite-based GPS support traceability by enabling precise delineation of boundaries of forest management units and farms from which materials are sourced and tracking their transport to ports, processing and manufacturing facilities, and to final point of sale.

Increasingly, governments are deploying traceability technologies to augment regulation of the forest products trade. Countries currently operating or introducing mandatory public timber traceability systems with centralized reporting platforms include Indonesia, Brazil, Peru, Guatemala, Honduras, Colombia, Ecuador, Panama, Liberia and Ghana. However, while governments have the political power to make reporting to a traceability system a legal requirement, the scope of these systems is by default limited to the national border. Without an overarching, international system to cover the complex material flows from producer, via processing to consumer countries, the development of country by country mandatory traceability systems is unlikely to succeed in preventing products associated with illegal logging or

The Indonesian Timber Legality Assurance System, locally known as SVLK (*Sistem Verifikasi Legalitas Kayu*), is illustrative of recent developments in public sector supply chain control. Long before a tree is harvested, concession holders enter information in an online system on tree species, location and estimated timber volume. This generates a barcode that is attached to the tree. After felling, the same bar code is attached to its stump and logs. The barcode enables the logs to be tracked to the point of primary processing. Additional entries are made in the online system to track the timber through processing and to connect batches of processed products to export licenses. The system requires timber concession holders to directly enter tree harvesting data in the system with minimal government supervision. However, if the system detects excess harvesting it will lock the concession holder's account. Authorities can also monitor the system and take action if they find irregularities. Private conformity assessment bodies, authorized by the National Accreditation Agency, reconcile the data provided and, where necessary conducting a field visit, to verify the concession holder's legality certificate or issue a non-compliance report. While the system is currently focused on verifying the legal supply of timber, the Indonesian government has announced its intention to expand the scope to included performance assessment of concession holders and payment of non-tax revenues (MOEF 2018).

forest clearing from entering global markets. While the technology is available today, the institutionalization of a comprehensive global traceability system remains a transnational governance challenge for the future.

While mandatory traceability systems are intended to impede illegal logging and timber trade, they are vulnerable to manipulation through input of false data. They have even been described as "laundering machines" (EIA 2012, Greenpeace 2013, Kleinschmidt 2016, Nellemann 2012). If flawed documents, such as permits obtained fraudulently or allowable cuts not based on genuine forest inventories, can be registered in a traceability system, they effectively create "phantom" timber volumes that can be used to launder illegal wood. This problem is compounded when the traceability systems lack transparency and independent forest monitors cannot access the data in them. (JPIK 2018).

Where civil society can get access to data in government-run traceability systems, it can use the information to expose inconvenient truths, by cross referencing the information in the system with other sources. For example, in 2016 the BVRio Institute launched a due diligence and risk assessment system for Brazilian tropical timber trade. The system has a big data approach, drawing from public traceability systems, public registries of infractions and convictions, publicly available data on distribution and density of commercial species and spatial data from Global Forest Watch, the Brazilian Government and other NGOs. BVRio found that around 30% of 3,500 logging permits issued since 2006 from Para and Mato Grosso had questionable or unrealistic volumes (BVRio 2016).

Similar big data approaches are being used at international level to identify risk of deforestation in agricultural commodity supply chains. The Transparency for Sustainable Economies (TRASE) tool draws on production, trade, and customs data and modeling to trace commodity flows back to production landscapes while identifying the actors involved. It identifies individual companies that export, ship and import a given commodity and applies an enhanced form of material flow analysis to link them to specific production localities ("TRASE," n.d.). Initiatives like Chain Reaction Research also combine multiple data-types (deforestation alerts, chain-of-custody and trade data, corporate financial and governance data) to assess the exposure of individual companies to material financial risks within agricultural commodity chains (Graham, Thoumi, Drazen, & Seymour, 2018). The "Global Forest Watch Pro" application combines remote sensing data and cloud computing to help companies asses risk of tree cover loss occurring in the farms or supply sheds of the mills, silos, or slaughterhouses from which they source (Amaral & Lloyd, 2019).

The Open Timber Portal is another example of a transparency platform enabled by technology. The portal provides information about forest management practices and legal compliance in participating countries. It compiles information from three different sources: official concession boundaries and registered timber producers from the government; documents uploaded voluntarily by timber producers to demonstrate compliance; and observations by third party forest monitors ("Open Timber Portal," n.d.). The portal enables geospatial data, legal documents, and allegations of non-compliance from these diverse sources to be consolidated and presented in user-friendly formats. This transparent information sharing means all parties can upload data to challenge, verify or refute information claims made by others.

New forensic methodologies are being used to query claims around the origins or contents of agricultural, forest and wildlife products. For example, stable isotope analysis is commonly used to determine origin and subsequent legality of food products and more recently, timber (Camin *et al*. 2017, Dormontt *et al*. 2015). Likewise, genetic analyses have been successfully used to bolster prosecutions in illicit wildlife and timber court cases(Janjua, Fakhar-I-Abbas, William, Malik, & Mehr, 2017, Wasser *et al*. 2018). Newly applied wood identification tools are being scaled for use by both inspectors to screen suspect material in ports of entry and by scientists in the laboratory to generate prosecutorial evidence against entities accused of sourcing wood illegally.

Within the forest products industry, techniques such as chemical and genetic analysis can identify a timber species and its origin from elements present in a wood product (UNODC 2016). When a robust collection of physical reference samples has been gathered – coming from the natural range of a timber species – these techniques can validate or invalidate the declared species and origin claims on documentation. This provides authorities, buyers of the products, or activists with a means of testing suspect claims about the content of a product or its source. Wood identification technologies include:

- *Visual methods* – visual observation and analysis of the anatomical patterns in a wood product are used to identify the species. This ranges from simple visual inspections by a frontline official with the aid of a hand-held magnification lens, through to the use of sophisticated image capture devices and processing algorithms. The main constraint on these tools is human capital and the lack of developed image-based reference databases depicting the natural variations in wood structure within and across species.
- *Chemical methods* – Mass spectrometry is used to analyze the phytochemicals laid down in heartwood to distinguish between different species that look similar. Stable isotope ratio analysis probes variations in the presence of non-radioactive isotopes such as oxygen, hydrogen and nitrogen. The ratios between these isotopes in trees differ across landscapes depending on geology and weather patterns. Radiocarbon dating can be used to determine the age of timber samples and whether harvesting occurred after regulations protecting the species came into effect or after the species was listed under the Convention on international Trade in Endangered Species.
- *Genetic methods* – Genetic analysis through the use of techniques such as DNA barcoding, DNA fingerprinting and phylogenetics can be used to accurately

determine the species and/or harvest origin of wood products, with caveats. To date, genetic analysis of wood products is hindered by challenges in obtaining consistent, high-quality DNA from processed wood products. An additional barrier is the lack of genetic reference databases for commercial timber species and their harvest origins along with the high cost of developing such databases (Galpern, Manseau, & Wilson, 2012).

Some examples of how wood identification technologies are used include:

- If a log is falsely labelled as coming from country "a", these techniques can be used to prove the log was smuggled from country "b" which has a log export ban.
- To identify tropical hardwoods in charcoal; a WWF study found 61 percent of barbecues in Germany at risk with 42 percent of charcoal samples containing tropical woods (WWF Deutschland, 2018),
- To identify pulp from tropical hardwoods in books; a WWF study found 19 out of 51 German children's books produced in south east Asia contained pulp from tropical hardwoods (Peter Hirschberger, Jokiel, Plaep, & Zahnen, 2010).
- During the hunt for the people who illegally chopped down big leaf maple in Gifford Pinchot National Forest in Washington in 2015, investigators used genetic fingerprinting to match planks seized at a sawmill to the exact stumps in the forest from which the timber had come (Irwin, 2019).
- Stable isotope ratio analysis was used to show that Mongolian Oak purchased by a US hardwood-floor retailer was illegally sourced from the Russian Far East rather than legal stocks in China (Irwin, 2019).

POLICY AND GOVERNANCE

Big data technologies can also be deployed to augment efforts to strengthen forest policies and governance. Policy reforms are usually complicated by: procedural challenges in ensuring that all stakeholder perspectives are voiced; the lack of ready means to detect social wrongs and impacts relative to biophysical conditions; the tendency for relevant regulations and functions to be spread across multiple line agencies or levels of government; and related potential for conflicts between laws or discrepancies between the letter of the law and administrative procedures as practiced. Text mining and natural language processing computational methods bring promises of scalable, fast-paced monitoring and analysis of such complexity within policy implementation and governance systems.

Text mining involves extracting underlying statistics from text such as word count and broad topics, while natural language processing details methods for analyzing the latent meaning and structure of text, such as actions, events, moods, and sentiment (Grimmer & Stewart, 2013). Text mining is currently applied in several sectors to prioritize policy agenda and streamline regulatory compliance. For instance, the World Bank applies text mining techniques to identifying policy priorities in presidential speeches to establish country-level drivers of long-term growth (Calvo-González, Eizmendi, & Reyes, 2018). Organizations such as the World Anti-Doping Agency also apply text mining algorithms to identify athletes who may be breaching doping regulations (Hong Bui 2018). The Oak Ridge National Laboratory uses text mining to identify drivers of clean energy innovation by analyzing investments and project finance documents (Lin *et al*. 2016). Text mining is also used to identify wildlife and environmental threats in oil and gas permits (Nasdaq, n.d.). With broad success across a variety of government and sectoral applications, these methodologies may also allow for faster, more efficient policy analysis and feedback during agenda setting, policy creation, and evaluation in the forestry sector.

Data mining methodologies have significant potential to improve monitoring and evaluation of the political and social economy around forests, which is an important but understudied aspect of forest monitoring (Mclain, Guariguata, Lawry, & Reed, 2019). They have similar potential to support forest governance monitoring, which encompasses the accountability, effectiveness, efficiency and fairness of policy and legal frameworks, decision making processes, and their implementation (FAO, 2011). Monitoring of indicators relevant to these issues has primarily relied on traditional survey methods, with researchers gathering data directly from field interviews and surveys (Jackson *et al.* 2004). However, advances in natural language processing and data mining are beginning to enable real-time, quantitative assessments of the impact of policy reforms and better understanding of contextual issues such as land tenure conflicts. Global news media coverage databases, such as the Global Database of Events, Languages, and Tone (GDELT) and the Integrated Crisis Early Warning System (ICEWS) provide detailed information about news events happening globally in real time. These data sources have recently been used to map social conflict (Sehgal, 2018), natural resource conflict (Wayland & Kuniholm, 2016), and political movements (Gao, Leetaru, Hu, Cioffi-Revilla, & Schrodt 2013). These data sources and methodologies bring significant potential to understand land driven conflict, social opinions in forest policy reforms, and shifts in government agenda through automated analysis of news media.

Data mining technologies can also be deployed to support social network analysis, to produce insights on the relationships that organizations and individuals have with each other, including the most powerful and important actors in a given social network. These "champions" can support the long-term success of forest conservation or restoration initiatives by facilitating information and knowledge transfer, influencing policy, and encouraging action (Paletto, Balest, Demeo, Giacovelli, & Grilli, 2016). Policy and legislative documents, including national and subnational plans and environmental policies, contain vast amounts of information relevant to forest monitoring that have yet to be tapped into. Data mining approaches can strengthen comparative policy analyses to inform policy-makers. (Cannon, Nakayama, Sasaki, & Rossiter, 2018) analyzed the rapid shifts of Turkey's Syria policies with text mining, finding reliable, valid, and generalizable results that greatly reduced the timeframe of policy analysis. (Ash, Chen, Delgado, Fierro, & Lin, 2018) found that machine learning models of judicial documents can accurately classify the impact of individual cases on policy. (Gilardi & Wüest, 2018) developed an end-to-end methodology for comparative policy analysis, finding that automated approaches to policy analyses increase transparency, facilitate replication, and allow for retroactive adjustments to and the scaling of existing analyses.

CONCLUSIONS

Forest stakeholders of all stripes are benefitting from faster computation of evermore data from earth observation, value chains and the data mining of texts and media. New technologies are enabling the transformation of this data into information that is more accessible, actionable and timely, making it harder to hide activities that harm forests or people living in and around them. Big data is shining a light on a diverse array of problems – illegal logging in remote frontiers, the willing purchase of commodities associated with deforestation, corrupt allocation of permits to log or clear forests, encroachment on the land of indigenous peoples without their consent, and official endorsement of implausible statistics.

Generation of data-driven insights is a necessary but insufficient condition for sound management of the Earth's forest assets. Quality information may fall on deaf ears because political will is lacking. It may not motivate remedial action due to fundamental flaws in governance. It may stay hidden in "black-box" government and corporate systems, denying access to civil-society watch-dogs or marginalized communities that it would otherwise benefit. These challenges are compounded by a confusing plethora of competing methodologies and data sources. This provides cover for lack of action and prevents comprehensive, transparent monitoring of progress towards global forest goals and corporate commitments.

That said, the diversity of forest data can also make forest sector actor more accountable. The multiple ways forest data can be generated – from high resolution satellite images, to mining of the "twittersphere", and genetic fingerprinting in a laboratory – ultimately make it harder to keep information hidden. This can manifest in a virtuous cycle that drives transparency. For example, the incentive for corrupt officials to obscure data on who is taking what volume of timber from a place will diminish if this can be discerned independently from satellite data and the mining of customs data. Similarly, the ability of an inspector or auditor to extract a kick-back by turning a bind-eye to a human rights violation, will diminish if that same violation is likely to be pinpointed through mining social media activity. If politicians are repeatedly queried on why their forests statistics tell a different story to data derived from independent geospatial data platforms, they may be motivated to upgrade their own forest monitoring systems. If companies that disclose very little about the sustainability of their supply chains are constantly facing down accusations of poor practice by campaigners, they might be moved to set ambitious sustainability targets and report openly and accurately on progress towards them.

While capacities and tools for forest monitoring will continue to improve, trade-offs will persist between the extent, resolution, precision, accuracy, and frequency of update of geospatial data (Fagan & DeFries, 2009). In developing forest monitoring systems, the key questions to ask are: what is the intended purpose of the system and what information is needed to fulfill that purpose? For example, a system to monitor national-level forest carbon changes for REDD+ will have different technical requirements than a system for quickly detecting illegal clearing within a national park. The purpose determines minimum requirements for: spatial resolution (what is the smallest object that can be distinguished): temporal resolution (how often does the data refresh); repeatability (can the methods be reproduced and compared across time to create a longitudinal record of changes); and affordability – lower cost systems are more likely to remain operational for large areas into the future (Davis & Peterson, 2016).

REFERENCES

AMARAL, L., and LLOYD, J. 2019. A new tool can help root out deforestation from complex supply chains. Retrieved July 15, 2019, from https://blog.globalforestwatch.org/commodities/a-new-tool-can-help-root-out-deforestation-from-complex-supply-chains

ASNER, G.P. 2009. Tropical forest carbon assessment: integrating satellite and airborne mapping approaches. *Environmental Research Letters* **4**(3): 034009. https://doi.org/10.1088/1748-9326/4/3/034009

BACCINI, A., GOETZ, S.J., WALKER, W.S., LAPORTE, N.T., SUN, M., SULLA-MENASHE, D., … HOUGHTON, R.A. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change* **2**(3): 182–185. https://doi.org/10.1038/nclimate1354

BALDECK, C.A., ASNER, G.P., MARTIN, R.E., ANDERSON, C.B., KNAPP, D.E., KELLNER, J.R., and WRIGHT, S.J. 2015. Operational Tree Species Mapping in a Diverse Tropical Forest with Airborne Imaging Spectroscopy. *PLOS ONE* **10**(7): e0118403. https://doi.org/10.1371/journal.pone.0118403

BALDWIN, G., MARKOWITZ, K., KOPAROVA, M., GERARDU, J., and ZAELKE, D. 2015. *Special Report on Next Generation Compliance*. Retrieved from http://inece.org/topics/next-gen-compliance/.

BISHOP, C.M. 2006. *Pattern Recognition and Machine Learning*. Berlin, Heidelberg: Springer-Verlag.

BOULTON, C.A., SHOTTON, H., and WILLIAMS, H.T.P. 2015. Using Social Media to Detect and Locate Wildfires. *Tenth International AAAI Conference on Web and Social Media*, 178–186.

BVRIO. (2016). *Using Big Data to Detect Illegality in the Tropical Timber Sector A Case Study of BVRio Due Diligence and Risk Assessment System*. Retrieved from https://docspublicos.s3.amazonaws.com/madeira/BVRio-Big-data-to-detect-timber-illegality.pdf

CALVO-GONZÁLEZ, O., EIZMENDI, A., and REYES, G. (2018). *Winners Never Quit, Quitters Never Grow Using Text Mining to Measure Policy Volatility and Its Link with Long-Term Growth in Latin America*. Retrieved from http://econ.worldbank.org.

CAMIN, F., BONER, M., BONTEMPO, L., FAUHL-HASSEK, C., KELLY, S.D., RIEDL, J., AND ROSSMANN, A. 2017. Stable isotope techniques for verifying the declared geographical origin of food in legal cases. *Trends in Food Science & Technology 61*: 176–187. https://doi.org/https://doi.org/10.1016/j.tifs.2016.12.007

CHEN, M., MAO, S., and LIU, Y. 2014. Big Data: A Survey. *Mobile Networks and Applications* **19**(2): 171–209. https://doi.org/10.1007/s11036-013-0489-0

CZÚNI, L., and VARGA, P.Z. 2014. *Lightweight Acoustic Detection of Logging in Wireless Sensor Networks*. 120–125.

DAUME, S. 2016. Mining Twitter to monitor invasive alien species – An analytical framework and sample information topologies. *Ecological Informatics* **31**: 70–82. https://doi.org/10.1016/j.ecoinf.2015.11.014

DAVIES, D.K., ILAVAJHALA, S., MIN MINNIE WONG, and JUSTICE, C.O. 2009. Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. *IEEE Transactions on Geoscience and Remote Sensing* **47**(1): 72–79. https://doi.org/10.1109/TGRS.2008.2002076

DAVIS, C., and PETERSON, R. 2016. Tools for Monitoring Global Deforestation, Reference Module in Earth Systems and Environmental Sciences. *Elsevier*.

DE SOUZA, C.M., HAYASHI, S., and VERÍSSIMO, A. 2008. *Near real-time deforestation detection for enforcement of forest reserves in Mato Grosso Instituto do Homem e Meio Ambiente da Amazônia-Imazon*. Retrieved from http://www.obt.inpe.br/prodes/

DENG, X., ZHU, Y., and NEWSAM, S. 2018. What Is It Like Down There? Generating Dense Ground-Level Views and Image Features From Overhead Imagery Using Conditional Generative Adversarial Networks. *SIGSPATIAL*.

DHANACHANDRA, N., MANGLEM, K., and CHANU, Y.J. 2015. Image Segmentation Using K-means Clustering Algorithm and Subtractive Clustering Algorithm. *Procedia Computer Science* **54**: 764–771. https://doi.org/10.1016/j.procs.2015.06.090

DORMONTT, E.E., BONER, M., BRAUN, B., BREULMANN, G., DEGEN, B., ESPINOZA, E., … LOWE, A.J. 2015. Forensic timber identification: It's time to integrate disciplines to combat illegal logging. *Biological Conservation* **191**: 790–798. https://doi.org/https://doi.org/10.1016/j.biocon.2015.06.038

EIA.(2012. *The Laundering Machine: How Fraud and Corruption in Peru's Concession System are Destroying the Future of its Forests*. Retrieved from https://content.eia-global.org/posts/documents/000/000/501/original/The_Laundering_Machine_ENG.pdf?1475785276

EUROPEAN SPACE AGENCY. 2013. Free access to Copernicus Sentinel satellite data. Retrieved July 15, 2019, from https://earth.esa.int/web/guest/content/-/article/free-access-to-copernicus-sentinel-satellite-data

FAGAN, M., and DEFRIES, R. 2009. *Measurement and Monitoring of the World's Forests: A Review and Summary of Remote Sensing Technical Capability, 2009–2015*. Retrieved from https://media.rff.org/archive/files/sharepoint/WorkImages/Download/RFF-Rpt-Measurement and Monitoring_Final.pdf

FAO. 2011. *Framework for Assessing and Monitoring Forest Governance*. Retrieved from http://www.fao.org/climatechange/27526-0cc61ecc084048c7a9425f64942df70a8.pdf

FAO. 2015. *Forest Resources Assessment*.

GALPERN, P., MANSEAU, M., and WILSON, P. 2012. Grains of connectivity: analysis at multiple spatial scales in landscape genetics. *Molecular Ecology* **21**(16): 3996–4009. https://doi.org/10.1111/j.1365-294X.2012.05677.x

GAO, J., LEETARU, K.H., HU, J., CIOFFI-REVILLA, C., and SCHRODT, P. 2013. *Massive Media Event Data Analysis to Assess World-Wide Political Conflict and Instability*. https://doi.org/10.1007/978-3-642-37210-0_31

GRAHAM, P., THOUMI, G., DRAZEN, E., and SEYMOUR, F. 2018. *Mining Global Financial Data to Increase Transparency and Reduce Drivers of Deforestation*. Retrieved from https://wriorg.s3.amazonaws.com/s3fs-public/ending-tropical-deforestation-mining-global-financial-data.pdf

GREENPEACE. 2013. *The Amazon's Silent Crisis*. Retrieved from https://www.greenpeace.org/usa/wp-content/uploads/legacy/Global/usa/planet3/PDFs/Amazon5Ways.pdf

GRIMMER, J., and STEWART, B.M. 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* **21**(3): 267–297. https://doi.org/DOI: 10.1093/pan/mps028

HANSEN, M.C., POTAPOV, P.V., MOORE, R., HANCHER, M., TURUBANOVA, S.A., TYUKAVINA, A., … TOWNSHEND, J.R.G. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**(6160): 850–853. https://doi.org/10.1126/science.1244693

HANSEN, MATTHEW, C., KRYLOV, A., TYUKAVINA, A., POTAPOV, P.V, TURUBANOVA, S., ZUTTA, B., … MOORE, R. 2016. Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters* **11**(3): 034008. https://doi.org/10.1088/1748-9326/11/3/034008

HONG BUI. 2018. Artificial intelligence takes on doping. Retrieved July 11, 2019, from https://www.veritone.com/blog/artificial-intelligence-takes-doping/

IRWIN, A. 2019. Tree sleuths are using DNA tests and machine vision to crack timber crimes. *Nature* **568**(7750): 19–21. https://doi.org/10.1038/d41586-019-01035-7

ITTO. 2012. *Tracking Sustainability: Review of Electronic and Semi-Electronic Timber Tracking Technologies*. Retrieved from http://www.itto.int/direct/topics/topics_pdf_download/topics_id=3145&no=0&disp=inline

JACKSON, J.E., LEE, R.G., SOMMERS, P., JACKSON, J.E., LEE, R.G., SOMMERS, P., … LEE, R.G. 2004. Monitoring the Community Impacts of the Northwest

Forest Plan: An Alternativeto Social Indicators Insights and Applications Monitoring the Community Impacts of the Northwest Forest Plan: An Alternative to Social Indicators. *Society and Natural Resources* **17**: 223–233. https://doi.org/10.1080/08941920490270258

JANJUA, S., FAKHAR-I-ABBAS, WILLIAM, K., MALIK, I.U., and MEHR, J. 2017. DNA Mini-barcoding for wildlife trade control: a case study on identification of highly processed animal materials. *Mitochondrial DNA Part A* **28**(4): 544–546. https://doi.org/10.3109/24701394.2016.1155051

JEAN, N., AZZARI, G., WANG, S., LOBELL, D., SAMAR, A., and ERMON, S. 2018. *Tile2Vec: Unsupervised representation learning for spatially distributed data arXiv: 1805. 02855v2 [cs . CV] 30 May 2018*.

JPIK. 2018. *SVLK: A Process toward Accountable Governance*. Retrieved from www.jpik.or.id

KLEINSCHMIDT, D. 2016. No TitleIllegal Logging and Related Trade: Dimensions, Drivers, Impacts, and Responses. In *Global Scientific Rapid Response Assessment Report*.

KNIAZ, V.V. 2018. Conditional GANs for semantic segmentation of multispectral satellite images. *Proc. SPIE 10789*, (October 2018). https://doi.org/10.1117/12.2325601

LIN, J., CHINTHAVALI, S., STAHL, C.D., STAHL, C., LEE, S., and SHANKAR, M. 2016. Ecosystem discovery: Measuring clean energy innovation ecosystems through knowledge discovery and mapping techniques. *The Electricity Journal* **29**(8): 64–75. https://doi.org/10.1016/j.tej.2016.09.012

MA, L., LIU, Y., ZHANG, X., YE, Y., YIN, G., and JOHNSON, B.A. 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing* **152**: 166–177. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2019.04.015

MCLAIN, R., GUARIGUATA, M., LAWRY, S., and REED, J. 2019. Integrating tenure and governance into assessments of forest landscape restoration opportunities. In *CIFOR Infobrief*. https://doi.org/10.17528/cifor/007132

MIDDLETON, S.E., MIDDLETON, L., and MODAFFERI, S. 2014. Real-Time Crisis Mapping of Natural Disasters Using Social Media. *IEEE Intelligent Systems* **29**(April): 9–17. https://doi.org/10.1109/MIS.2013.126

MOEF. 2018. *The State of Indonesias Forests 2018* (S. Nurbaya and S. Awang, Eds.). Retrieved from http://perpustakaan.bappenas.go.id/lontar/file?file=digital/191959-%5B_Konten_%5D-Konten E2337.pdf

NASDAQ. (n.d.). How Artificial Intelligence Is Taking Over Oil And Gas – Nasdaq.com. Retrieved July 11, 2019, from https://www.nasdaq.com/article/how-artificial-intelligence-is-taking-over-oil-and-gas-cm1005575

NELLEMANN, C. 2012. *Green carbon, black trade: Illegal logging, tax fraud, and laundering in the worlds tropical forests: A rapid response assessment*. Retrieved from www.grida.no

OPEN TIMBER PORTAL. (n.d.). Retrieved July 15, 2019, from https://opentimberportal.org/

PALETTO, A., BALEST, J., DEMEO, I., GIACOVELLI, G., and GRILLI, G. 2016. Power of Forest Stakeholders in the Participatory Decision Making Process: A Case Study in Northern Italy. *Acta Silvatica et Lignaria Hungarica*, **12**(1): 9–22. https://doi.org/10.1515/aslh-2016-0002

PETER HIRSCHBERGER, JOKIEL, D., PLAEP, C., and ZAHNEN, J. 2010. *Tropical Forest Destruction for Children's Books*.

REICHE, J., HAMUNYELA, E., VERBESSELT, J., HOEKMAN, D., and 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 Sensing of Environment* **204**: 147–161. https://doi.org/10.1016/J.RSE.2017.10.034

REICHE, J., LUCAS, R., MITCHELL, A.L., VERBESSELT, J., HOEKMAN, D.H., HAARPAINTNER, J., … HEROLD, M. 2016. Combining satellite data for better tropical forest monitoring. *Nature Climate Change* **6**(2): 120–122. https://doi.org/10.1038/nclimate2919

REICHE, J., VERHOEVEN, R., VERBESSELT, J., HAMUNYELA, E., WIELAARD, N., HEROLD, M., … HEROLD, M. 2018. Characterizing Tropical Forest Cover Loss Using Dense Sentinel-1 Data and Active Fire Alerts. *Remote Sensing* **10**(5): 777. https://doi.org/10.3390/rs10050777

REYMONDIN, L., JARVIS, A., PEREZ-URIBE, A., TOUVAL, J., ARGOTE, K., REBETEZ, J., … MULLIGAN, M. 2012. *Terra-i: A methodology for near real-time monitoring of habitat change at continental scales using MODIS-NDVI and TRMM*.

SAATCHI, S.S., HARRIS, N.L., BROWN, S., LEFSKY, M., MITCHARD, E.T.A., SALAS, W., … MOREL, A. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America* **108**(24): 9899–9904. https://doi.org/10.1073/pnas.1019576108

SANQUETTA, CARLOS R, WOJCIECHOWSKI, J., CORTE, A.P.D., BEHLING, A., NETTO, S.P., RODRIGUES, A.L., and SANQUETTA, M.N.I. 2015. Comparison of data mining and allometric model in estimation of tree biomass. *BMC Bioinformatics* **16**(247): 1–9. https://doi.org/10.1186/s12859-015-0662-5

SANQUETTA, CARLOS ROBERTO, WOJCIECHOWSKI, J., PAULA, A., CORTE, D., and RODRIGUES, A.L. 2013. On the use of data mining for estimating carbon storage in the trees. *Carbon Balance and Management* **8**(6): 1–9. https://doi.org/10.1186/1750-0680-8-6

SAOUDI, M., EULER, R., BOUNCEUR, A., and KECHADI, T. 2016. *Data Mining Techniques Applied to Wireless Sensor Networks for Early Forest Data Mining Techniques Applied to Wireless Sensor Networks for Early Forest Fire Detection*. https://doi.org/10.1145/2896387.2900323

SEHGAL, G. 2018. Spatio-temporal networks of social conflicts: analysis and modeling. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 740–743. https://doi.org/10.1109/ASONAM.2018.8508266

SHAHAM, U., STANTON, K., LI, H., NADLER, B., BASRI, R., and KLUGER, Y. 2018. SpectralNet: Spectral Clustering using Deep Neural Networks. *ICLR*. Retrieved from http://arxiv.org/abs/1801.01587

SHI, Q., LIU, X., and LI, X. 2018. Road Detection From Remote Sensing Images by Generative Adversarial Networks. *IEEE Special Section on Advanced Data Analytics for Large-Scale Complex Data Environments* **6**: 25486–25494. https://doi.org/10.1109/ACCESS.2017.2773142

SHI, Y., LI, Q., and ZHU, X.X. 2019. Building Footprint Generation Using Improved Generative Adversarial Networks. *IEEE Geoscience and Remote Sensing Letters* **16**(4): 603–607. https://doi.org/10.1109/LGRS.2018.2878486

SHIMABUKURO, Y., DOS SANTOS, J., FORMAGGIO, A., DUARTE, V., and RUDORFF, B. 2016. The Brazilian Amazon Monitoring Program: PRODES and DETER Projects. In *Global Forest Monitoring from Earth Observation* (pp. 166–183). https://doi.org/10.1201/b13040-13

SINGH, A., and SINGH, K.K. 2018. Unsupervised change detection in remote sensing images using fusion of spectral and statistical indices. *Egyptian Journal of Remote Sensing and Space Science* **21**(3): 345–351. https://doi.org/10.1016/j.ejrs.2018.01.006

TANEV, H., ZAVARELLA, V., and STEINBERGER, J. 2017. Monitoring disaster impact: detecting micro-events and eyewitness reports in mainstream and social media. *Proceedings of the 14th ISCRAM Conference*, 592–602.

TRASE. (n.d.). Retrieved July 15, 2019, from https://trase.earth/about/how-does-trase-work?

TUNG, F., WONG, A., and CLAUSI, D.A. 2010. Enabling scalable spectral clustering for image segmentation. *Pattern Recognition* **43**(12): 4069–4076. https://doi.org/10.1016/j.patcog.2010.06.015

UNODC. 2016. *Best practice guide for forensic timber identification*. Retrieved from https://www.unodc.org/documents/Wildlife/Guide_Timber.pdf

WASSER, S.K., TORKELSON, A., WINTERS, M., HOREAUX, Y., TUCKER, S., OTIENDE, M.Y., … WEIR, B.S. 2018. Combating transnational organized crime by linking multiple large ivory seizures to the same dealer. *Science Advances* **4**(9): eaat0625. https://doi.org/10.1126/sciadv.aat0625

WAYLAND, J., and KUNIHOLM, M. 2016. Legacies of conflict and natural resource resistance in Guatemala. *The Extractive Industries and Society* **3**(2): 395–403. https://doi.org/10.1016/j.exis.2016.03.001

WULDER, M.A., and COOPS, N.C. 2014. Satellites: Make Earth observations open access. *Nature*, *513*(7516), 30–31. https://doi.org/10.1038/513030a

WWF DEUTSCHLAND. 2018. *Tropical Wood in Germany's Charcoal*.

XIE, M., JEAN, N., BURKE, M., LOBELL, D., and ERMON, S. 2016. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, 3929–3935. Retrieved from http://arxiv.org/abs/1510.00098

ZHANG, L., and YOU, J. 2017. A Spectral Clustering Based Method for Hyperspectral Urban Image. *2017 Joint Urban Remote Sensing Event (JURSE)*, 1–3. https://doi.org/10.1109/JURSE.2017.7924602