

# Chapter 22

## Introduction—State of the Art of Technology and Market Potential for Big Data in Forestry



Jukka Miettinen and Renne Tergujeff

**Abstract** Forest monitoring is undergoing rapid changes due to the growing data volumes, developing data processing technologies and increasing monitoring requirements. The DataBio forestry pilots set out to demonstrate how big data approaches can support the forestry sector to get full benefit of the evolving technologies and to meet the increasing monitoring requirements. In this introductory chapter, we describe underlying technical and market forces driving the forestry sector toward big data approaches, and give short overviews on the forestry pilots to be presented in the following chapters.

### 22.1 Evolving Technologies and Growing Data Volumes

The forestry sector has been one of the forerunners in processing and analysis of large datasets. Particularly, remote sensing-based forest monitoring has utilized large datasets in the form of digital imagery since the 1970s when the first Landsat satellite was launched [1]. Satellite-based inventory approaches have been integrated into large area forest inventory programs since the 1990s [2, 3]. But in many ways, the launch of the Google Earth Engine (<https://earthengine.google.com/>), a cloud-based platform for planetary-scale geospatial analysis, in 2010, and the first global multi-year tree cover clearance analysis produced on the platform by Hansen et al. in 2013 [4], can be seen as the start of the big data era in forest monitoring. The platform and the innovative tree cover clearance product very much showed the direction for future big data development in the forestry sector.

Since then, data volumes have continued to grow rapidly, and the availability of different types of datasets has improved, increasing potential use cases for forestry big data. While in 2014, there were only around 200 active Earth observation (EO) satellites in orbit, in 2019, there were nearly 700 of them [5]. Simultaneously with the increasing number of EO satellites, also the temporal and spatial resolutions have improved rapidly. The 10–30 m spatial resolution Landsat 8 and Sentinel-1

---

J. Miettinen (✉) · R. Tergujeff  
VTT Technical Research Centre of Finland, 02044 VTT, Espoo, Finland  
e-mail: [jukka.miettinen@vtt.fi](mailto:jukka.miettinen@vtt.fi)

and Sentinel-2 satellites are replacing coarse (250–1000 m) spatial resolution data in many large area forest monitoring applications, e.g., for burnt area [6], forest disturbance [7], and health [8] monitoring. The Copernicus Sentinel program alone (with its six satellites in orbit) produces over 12 TB of data per day [9]. On the commercial front, companies like Planet Labs (<https://www.planet.com/>) are able to scan the entire globe every day in 3–5 m spatial resolution, providing high-frequency data for forest monitoring in unprecedented spatial detail. Other companies concentrate on less than 1 m very high spatial resolution imagery, which can be used as reference data in big data forest monitoring approaches.

The increase in EO data volumes is combined with the escalation of drone surveillance (including hyperspectral cameras, etc.), and the continuing national monitoring campaigns with airborne optical and LiDAR sensors [10]. Furthermore, field measurements are increasingly taken with electric devices, increasing the speed and volume of data collected. Field measurement campaigns are supplemented by continuous data collection from machinery used in forest operations (e.g., location and measurement data from the cutting-heads of harvester machines [11]). And most recently, the launch of crowdsourcing-based data collection approaches allows fast and effective collection of large field observation datasets.

The most effective way for the forestry sector to utilize the great volumes of data produced by modern technology is through centralized storage and processing platforms. Since the early days of Google Earth Engine, numerous other online platforms have been set up. Nowadays, many online platforms operate in clusters that provide the resources to implement big data forest applications in an effective and innovative manner. Platforms like the Copernicus Data Access and Information Services (DIAS; <https://www.copernicus.eu/en/access-data/dias>) offer direct access to EO big data and processing capabilities, while other, often domain-specific platforms, such as the Forestry Thematic Exploitation Platform (Forestry TEP; <https://f-tep.com/>), additionally provide tools and services designed particularly for utilization of big data, e.g., in forestry. The platforms form a hierarchical offering, from data storages and processing platforms to nuanced application platforms and interactive user interfaces. This network of platforms allows creation of delivery pipelines that can maximize the benefits of big data in forestry, by providing users with timely datasets and analysis results that meet their specific information requirements.

## 22.2 Expanding Market

Forests are in focus nowadays perhaps more than ever before. Both political and market interest in bioeconomy, growing recognition of the importance of forests in climate change mitigation, and increasing requirements on forest management (e.g., in the field of sustainability) demand timely and affordable information on forest resources. Forestry stakeholders, like government entities, non-governmental organizations, private companies, and forest owners, are bound by a wide range of international and national strategies and legislation. For example, in Europe, forestry

stakeholders are not only affected by the European Forest Strategy [12], but also, e.g., by the Biodiversity Strategy [13] and Bioeconomy Strategy [14]. While the Forest Strategy provides a policy framework that coordinates and ensures coherence of forest-related policies, the Biodiversity Strategy aims to protect ecosystems (including forests) and biodiversity, and the Bioeconomy Strategy aims to serve as an umbrella for long-term sustainable development. These European wide strategies are reflected in national-level legislation in member states, requiring stakeholders to report and monitor increasing number of forestry indicators, ranging from harvest levels and reforestation status to biological diversity, carbon balance, forest health, and many more.

In most European countries, traditional methods for forest management are based on static management plans, created at the planting stage and reviewed after long periods (typically in five to ten years intervals). This type of management process does not meet the needs of modern requirements of manifold up-to-date information described above. Furthermore, in recent years, the management plans have become declarations of intentions, including objectives for multifunctional forests (non-wood products and services). However, the management system lacks effective monitoring methods that allow forest owners, managers, and regulators to validate the progress in achieving the target objectives set out in the management plan.

Forest owners, forestry operators, and companies using wood as raw material are also affected by various voluntary certification schemes like the Forest Stewardship Council (FSC; <https://fsc.org/en>) and the Program for the Endorsement of Forest Certification (PEFC; <https://www.pefc.org/>). They both aim to ensure sustainable forest management using a set of criteria ranging from sustainable wood production to biodiversity, forest health, and carbon balance. Independent auditors need access to a wide variety of forest variable information and change statistics to be able to verify that the certification standards have been followed correctly. Overall, the rising interest in forests and the widening range of forest management aspects included in strategies, legislation, and certification schemes are rapidly growing the market for timely forest information. The modern technology outlined in the previous sections can be used to establish operational monitoring systems providing transparent products helping to meet the increased monitoring and reporting requirements.

Big data can benefit both the provider and the customer side of the forest monitoring market. On the provider side, one of the main stakeholder groups in Europe consists of the Earth observation (EO) data, product, and service providers. According to the European Association of Remote Sensing Companies (EARSC), the EO service sector employed over 7000 people in 500 companies with over 900 million € revenue already in 2014 [15], with a strong growing trend. Forest monitoring is among the main focus areas of the European private EO sector. Although EO data cannot be used to monitor all of the variables required in modern forest management (e.g., plant and animal biodiversity in fine detail), it does provide the means to monitor key variables like the structural forest characteristics and forest health data, as demonstrated by the DataBio forestry pilots presented in the following chapters. In addition to the EO, forestry big data market benefits, e.g., consultants and forestry experts, IT specialists, and data analysis specialists.

The customer side of the market is likewise varied and expanding. The public sector has their monitoring requirements defined by national forest legislations, and non-governmental organizations want to monitor development on forest resources to support their activities. Companies directly involved in forestry activities need to have timely information on the forest resources, not only to support their own forest management, but also to provide data for certification purposes. Even companies that are not directly involved in forest management activities increasingly choose to get involved in the forest monitoring market due to the increasing legislation, certification, or consumer pressure. For example, food manufacturers, energy companies, and sellers of wood-based products (e.g., furniture) may have compulsory obligations or voluntary interest in forest monitoring. This trend is expected to grow in the future, as environmental issues are becoming an increasingly important part of business practices.

Overall, information on forest resources is nowadays needed frequently and in high spatial detail to meet the requirements of various reporting and regulative monitoring schemes. Moreover, the information is expected to be available in short notice and in easily reachable online platforms to allow direct integration of the data into the stakeholders' databases and operational analyses. These are hard demands, but forestry big data has a great potential to meet these demands, with appropriate network of online storage, processing, and distribution platforms. This is what the DataBio forestry pilots aimed to demonstrate.

## 22.3 DataBio Forestry Pilots

The objective of the DataBio forestry pilots was to demonstrate how Big Data could boost the Forestry sector. The pilots, carried out in four countries (Belgium, Czech Republic, Finland, and Spain), were built around practical forestry cases. They validated the use of Big Data technologies and assessed how the expectations of user communities can be met. Overall, the pilots sought to demonstrate how big data approaches could be used to:

1. *Improve presentation and delivery of crowdsourced forest data* and introduce new functionalities on data distribution and analysis platforms. Crowdsourced data is among the newest types of data used in forestry. The best practices for data utilization are still very much in development. At its best, crowdsourced data provides an effective way to gather information, e.g., on forest damages after storm events. However, its usability may be affected by issues like bias or unreliability. Furthermore, new technical solutions are needed for both data collection as well as distribution of crowdsourced data. One of the DataBio pilots (Chap. 23) concentrated on crowdsourced data collection and utilization.
2. *Optimize the use of tree resources.* Detailed characterization of trees and forest characteristics is used to determine the optimal use of trees for a given output (e.g., pulp, paper, textile, and biofuels) in order to guarantee that supply

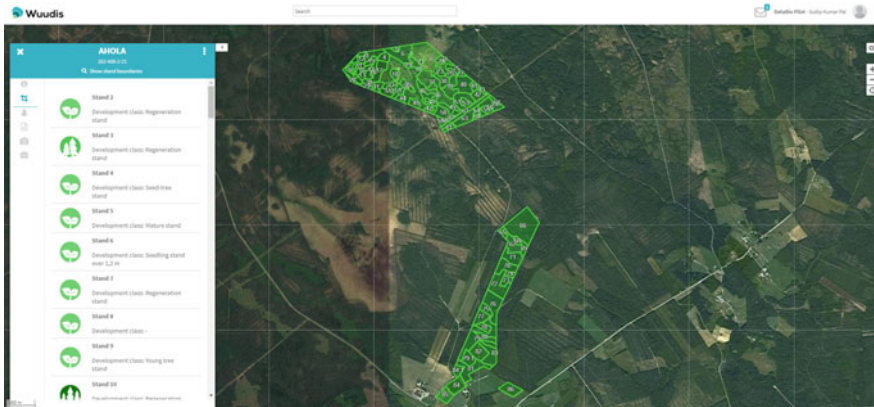
meets demand. To enable reliable optimization of forest management activities, information on forest structural variables (e.g., species, height, and stem number) need to be kept up-to-date. Outdated forest information is one of the major hindrances in forest management around the world. With the increased temporal and spatial resolution, forestry big data allow improved timeliness of accurate forest variable data provision, and thereby improved optimization of tree resources. Provision of up-to-date forest characteristics utilizing online platforms was looked into in one of the DataBio pilots presented in Chap. 24.

3. *Improve identification of forest health and damages* caused by biotic (such as pests and diseases) or abiotic (such as snow, wind, dryness, rains, and fires) agents using remote sensors. Biotic forest damages are expected to become increasingly common in the near future due to rising temperatures [16]. Similarly, the frequency of extreme weather events is expected to rise due to the climate change, increasing the risk for abiotic damages. Big data processing and analysis allows implementation of time series approaches that allow forest health and damage monitoring for large areas in high spatial and temporal detail. Two pilots dealing with forest health monitoring are presented in Chaps. 25 and 26.

An overarching idea in all of the DataBio forestry pilots was to develop integrated tools to support management planning that is based on online platform infrastructures. Several of the pilots were linked to the Wuudis platform (<https://www.wuudis.com/>), which was used as the central piece to develop and demonstrate usability of inter-platform approaches. The Wuudis Service developed by Wuudis Solutions Oy is a commercial service on the market for forest owners, timber buyers, and forestry service companies. It enables the management of forestry activities (e.g., thinning and harvesting) and forest resources (e.g., forest estate evaluation) through a single tool. It can be used to obtain real-time information about the forest and its timber resource, track executed silvicultural and harvest activities, and plan the needed forest management activities.

The Wuudis platform (Fig. 22.1) was linked with other platforms in the pilots to highlight the possibilities of inter-platform connections in big data processing pipelines. Most notably, satellite image processing and analysis capabilities of the EO Regions! (<https://www.eoregions.com/>) platform and the Forestry Thematic Exploitation Platform (Forestry TEP; <https://f-tep.com/>) were used to feed user specific forest variable information into the Wuudis system. The EO Regions! platform provides services, information, and products specially created for service providers in Wallonia and Europe, while the Forestry TEP platform enables commercial, research, and public sector users in the forestry sector to efficiently access satellite data-based processing services and tools for generating value-added forest information products. Via the Forestry TEP platform, the users can also create and share their own services, tools, and generated products.

Similarly, the Metsään.fi service (<https://www.metsaan.fi/>) was linked with Wuudis Service to enable the exchange of data in both directions, to expand the data resources and functionality of both services. Wuudis users benefit from the



**Fig. 22.1** Forestry estate borders and data transferred into Wuudis from the Metsään.fi service

open source data available in Metsään.fi, while users of Metsään.fi benefit from the additional functionalities available in Wuudis. The Metsään.fi service is provided by the governmental body Finnish Forest Center to make forest resource information available for citizens free of charge. The platform serves forest owners and Forestry service providers. Through the portal, forest owners in Finland can conduct business related to their forests at home from their own desktops. Metsään.fi connects forest estate owners with related third parties, including providers of Forestry services. This makes it easy to manage forestry work and to be in touch with forestry professionals.

In the following chapters of the book, four DataBio forestry pilots are presented. The presentations outline the pilot structure and highlight the main technical results. They also analyze the technological and market aspects of the usability and potential of big data in forestry. The chapters include:

**Chapter 23—Finnish Forest Data-Based Metsään.fi-services:** The best ways to utilize crowdsourced data in forestry are still very much in development. The pilot aimed to trial crowdsourced forest data presentation and new functionalities related to it. The launch of a new open forest data service, as well as related crowdsourcing services, was included in this pilot. Two areas for crowdsourcing solutions were implemented: (1) showing quality control data for young stand improvement and early tending for seedling stand, and (2) storm damage data. Other possible crowdsourced data, such as other forest damage than storm damage data, were also evaluated.

**Chapter 24—Forest Variable Estimation and Change Monitoring Solutions Based on Remote Sensing Big Data:** Lack of up-to-date information on forest structural characteristics commonly prevents optimal forest management in large parts of the world. The pilot aimed to demonstrate the feasibility of online platform-based forest inventory approaches. The pilot focused on developing the forest inventory system on the Wuudis platform, which is based on remote sensing data and field surveys. The pilot was started in Finland and Belgium, but later expanded into Spain. The goal was

to evaluate the usability of the technologies and processing methods of the project partners in different conditions varying from the Northern Boreal forests in Finland, through temperate forests in Belgium to the Galician forests in the Atlantic coastline in Spain. The pilot demonstrated inter-platform capabilities for comprehensive and near real-time quantitative assessment of forest cover over the interest areas.

Chapter 25—*Monitoring Forest Health: Big Data Applied to Diseases and Plagues Control*: Forest health monitoring is increasingly important due to the changing climate, and Big Data has the potential to offer means for effective large-scale forest health monitoring. The pilot set up the first version of a methodology and mathematical model based on remote sensing images (Sentinel-2 + Unmanned Aerial Vehicle) for the monitoring of health status of forests in the Iberian Peninsula. The work focused on the monitoring of *Quercus* forests affected by *Phytophthora cinnamomi* Rands and on the damage in eucalyptus plantations affected by *Gonipterus scutellatus*. After the definition of the big data algorithms and image processing techniques development, an EO-based system for monitoring the health of big forest areas was proposed, in order to enable public administrations to optimize their forest management resources.

Chapter 26—*Forest Damage Monitoring for the Bark Beetle*: Bark beetle outbreaks cause widespread ecological and economic damage in central Europe on a yearly basis, and are predicted to become even more severe in the near future. The pilot aimed to develop a new methodology for forest health assessment based on Copernicus satellite data (Sentinel-2). An approach was designed for assessment of forest health of the entire area of Czech Republic and other temperate forest regions in Europe, while reducing costs for field surveys. The method supports government officials by enabling effective identification of forest owners eligible for subsidies/tax relief. In addition, forest owners benefit from publicly available map server, where all forest health status maps are made available to allow pro-active management of forest properties.

After the individual pilot descriptions, a summary *Chap. 27—Conclusions and Outlook—Summary of Big Data in Forestry* will draw together the main findings of the DataBio project on the usability and potential.

**Acknowledgements** The authors would like to thank Seppo Huurinainen from Wuudis Solutions Oy, for leading the forestry pilots for most of the duration of the DataBio project, and everybody who contributed to the work in the forestry pilots during the project.

## References

1. Lauer, D. T., Morain, S. A., & Salomonson, V. V. (1997). The landsat program: Its origins, evolution, and impacts. *Photogrammetric Engineering and Remote Sensing*, 63, 831–838.
2. Tomppo, E. (1996). Multi-source national forest inventory of Finland. In: Päivinen, R., Vanclay, J., Miina, S. (Eds.), *New thrusts in forest inventory*. Proceedings of the Subject Group S4.02-00,

- Forest Resource Inventory and Monitoring, and Subject Group S4.12-00, Remote Sensing Technology, Vol I. IUFRO XX World Congress, Tampere, Finland, August 6–12, 1995. European Forest Institute, Joensuu, Finland.
3. Food and Agriculture Organization. (1996). Forest resources assessment 1990: Survey of tropical forest cover and study of change processes. FAO Forestry Paper 130. Food and Agriculture Organization (FAO), Rome.
  4. Hansen, M. C., Potapov, P. V., Moore, R., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342, 850–853.
  5. The GeoSpatial. (2019). How many earth observation satellites are there in space right now? <https://www.thegeospatial.in/earth-observation-satellites-in-space>. Accessed on 26 Sept 2020.
  6. Chuvieco, E., Mouillot, F., van der Werf, G. R., et al. (2019). Historical background and current developments for mapping burned area from satellite earth observation. *Remote Sensing of Environment*, 225, 45–46.
  7. Cohen, W. B., Yang, Z., Healey, S. P., et al. (2018). A LandTrendr multispectral ensemble for forest disturbance detection. *Remote Sensing Environment*, 205, 131–140.
  8. Hirschmugl, M., Gallaun, H., Dees, M., et al. (2017). Methods for mapping forest disturbance and degradation from optical earth observation data: A review. *Current Forestry Reports*, 3, 32–45.
  9. Copernicus. (2018). Develop your Copernicus-based business with the Copernicus incubation programme. <https://www.copernicus.eu/en/develop-your-copernicus-based-business-copernicus-incubation-programme>. Accessed on 26 Sept 2020.
  10. National Land Survey of Finland. (2020). Laser scanning data 2008–2019: <https://www.maanmittauslaitos.fi/en/maps-and-spatial-data/expert-users/product-descriptions/laser-scanning-data>. Accessed on 26 Sept 2020.
  11. Maltamo, M., Hauglin, M., Næsset, E., et al. (2019). Estimating stand level stem diameter distribution utilizing harvester data and airborne laser scanning. *Silva Fennica*. <https://doi.org/10.14214/sf.10075>.
  12. European Commission. (2013). A new EU forest strategy: For forests and the forest-based sector. [https://eur-lex.europa.eu/resource.html?uri=cellar:21b27c38-21fb-11e3-8d1c-01aa75ed71a1.0022.01/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:21b27c38-21fb-11e3-8d1c-01aa75ed71a1.0022.01/DOC_1&format=PDF). Accessed on 26 Sept 2020.
  13. European Commission. (2011). The EU biodiversity strategy to 2020. <https://ec.europa.eu/environment/nature/info/pubs/docs/brochures/2020%20Biod%20brochure%20final%20lowres.pdf>. Accessed on 26 Sept 2020.
  14. European Commission. (2018). A sustainable bioeconomy for Europe: Strengthening the connection between economy, society and the environment—Updated bioeconomy strategy. [https://ec.europa.eu/research/bioeconomy/pdf/ec\\_bioeconomy\\_strategy\\_2018.pdf](https://ec.europa.eu/research/bioeconomy/pdf/ec_bioeconomy_strategy_2018.pdf). Accessed on 26 Sept 2020.
  15. European Association of Remote Sensing Companies. (2016). Creating a European marketplace for earth observation services. <https://earsrsc.org/wp-content/uploads/2020/08/EARSC-Marketplace-Alliance-PressRelease-V1.pdf>. Accessed on 26 Sept 2020.
  16. Seidl, R., Schelhaas, M.-J., Rammer, W., et al. (2014). Increasing forest disturbances in Europe and their impact on carbon storage. *Nature Climate Change*, 4, 806–810.



**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

