ISPRS-SHY - OPEN DATA COLLECTOR FOR SUPPORTING GROUND TRUTH REMOTE SENSING ANALYSIS

F. Pirotti[1,2]*, M. Piragnolo[1,2], M. Yoshimura[3], J.P. Hernandez [4], M.A. Brovelli [5], B. Leblon [6], M. Yamashita[7]

[1] Department of Land, Environment, Agriculture and Forestry (TESAF), University of Padova, Legnaro, Italy – francesco.pirotti@unipd.it
[2] Interdepartmental Research Center of Geomatics (CIRGEO), University of Padova, Legnaro, Italy
[3] Center for Spatial Information Science, The University of Tokyo. Tokyo, Japan - gr4m-ysmr@asahi-net.or.jp
[4] Forestry and Nature Conservation Faculty, Universidad de Chile – jhernand@uchile.cl
[5] Department of Civil and Environmental Engineering, Politecnico di Milano, Italy – maria.brovelli@polimi.it
[6] Faculty of Forestry and Environmental Management University of New Brunswick, Canada - bleblon@unb.ca
[7] Division of Environmental and Agricultural Engineering Tokyo University of Agriculture and Technology, Fuchu, Tokyo, Japan - meguyama@cc.tuat.ac.jp

KEY WORDS: crowdsourcing, data sharing, open data, geographic information systems, webGIS.

ABSTRACT:

The 2021 Scientific Initiatives in ISPRS funded this project called ISRS-SHY from "*SHare mY ground truth*". It was intended as a collector of geographic data to support image analysis by sharing the necessary ground truth data needed for rigorous analysis. Regression and classification tasks that use remote sensing imagery necessarily require some control on the ground. The rationale behind this project is that often data on the ground is collected during projects, but is not valued by sharing across projects and teams globally. Internet has improved the way that data are shared, but there are still limitations related to discoverability of the data and its integrity. In other words, data are usually kept in local storage or, if in an accessible server, they are not documented and therefore they will not be picked up during search. In this initiative we created a portal using the Geonode environment to provide a hub for sharing data between research groups and openly to the community. The portal was then tested within the framework of three projects, with several participants each. The data that was uploaded and shared covered all types of geographic data formats and sizes. Further sharing was done in the context of teaching activities in higher education.

The results show the importance of creating easy means to find data and share it across stakeholders. Qualitative results are discussed, and future steps will focus on quantitative assessment of the portal's usage, e.g. number of registered users in time, number of visits, and other key performance indicators. The results of this project are to be considered also in light of the effort in the scientific community to make research data available, i.e. FAIR - Findability, Accessibility, Interoperability, and Reuse of digital assets.

1. INTRODUCTION

It is not a novel affirmation that geographic data are increasing in number, volume and variety, in some cases reaching a bigdata paradigm. Geodata from satellite imagery, laser scanners in mobile platforms and other combinations of vehicles and sensors can be collected at impressive rates. Accurate geodata collection is only part of the process; it is soon followed by several other aspects that are just as challenging: data organization, visualization, structuring and sharing across a network. Having geodata in a single digital support, sharing it through devices only when necessary, is not a practical solution anymore. In time, folders might get deleted, devices change, digital supports get damaged or lost, and precious data is lost; it is a matter of time. The objective of the proposed scientific initiative was to implement a solution using a web-based platform for collecting and sharing data from areas that get sampled with ground-truth information, mainly (but not limited to) natural and agricultural ecosystems (Share mY ground truth - SHY).

Sampling and measuring data in the field is costly and time-consuming (Mozzato et al., 2018; Pagliacci et al., 2020; Vaglio Laurin et al., 2016) and all too often data are used for specific projects and then are lost - or become very hard to find – some time after the end of project activities. Sampling data over areas covered with vegetation can be particularly complex (thick canopy/crop, wet areas, limited accessibility), but is necessary for many applications e.g. biomass estimation (Pirotti et al., 2014), pest monitoring, yield prediction, analysis of climate change effects. Of course, data are not only limited to the realm of agro-forestry and natural ecosystems, but can be scaled to urban scenarios. The Eureka project (Prataviera et al., 2021) is an example of 3D geospatial models that can be used for energy modelling that can be shared via open-source means, and would benefit from an interface for sharing its data sources and project results.

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLIII-B5-2022-45-2022

SHY aimed at providing a user-friendly and guided interface for uploading spatial datasets and respective metadata, e.g. date, description, authors, license, links to references to literature. Both metadata and spatial information will be searchable to allow users to find data that might support their research.

2. MATERIALS AND METHODS

SHY has been installed in a remote machine with 128 MB of RAM and 16 available processors and a Linux Ubuntu 20.4 operating system. Implementation consisted on extending an open source project, Geonode. Geonode is an "Open Source Geospatial Content Management System - a web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructures (SDI)". It is based on Python and other open source projects, such as Geoserver and PostgreSQL for data management. The structure of the main system components, Geonode, Geoserver and PostgreSQL, is organized in different Docker containers for easier management and backup. For now, anyone can sign-up via normal registration or using social accounts from Linkedin or Facebook (Figure 1).

Contributors can choose to upload full datasets (Figure 2) or only the description and metadata of geodata shared via an external link or with a contact for requesting the full dataset. This option is ideal for very large datasets that might be problematic to upload. One example for the latter option that is already included in SHY is a UAV survey with more than 1000 images and the resulting point clouds with ~1 billion points; the actual data are stored respectively in a Google Drive folder and in an online viewer (Potree) that are searchable and linked in the portal. Users can either download the data directly or send an email to authors asking for the full dataset, in case only metadata and author reference is provided.

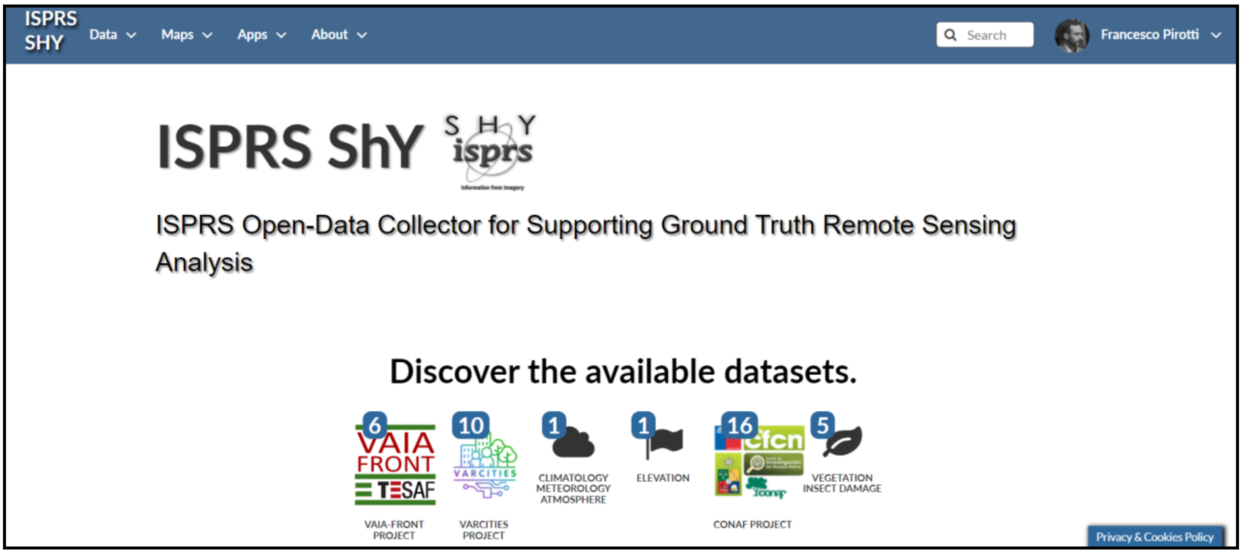

Figure 1. Front page of ISPRS-Shy data portal, with projects and datasets; accessible at https://isprs-shy.cirgeo.unipd.it

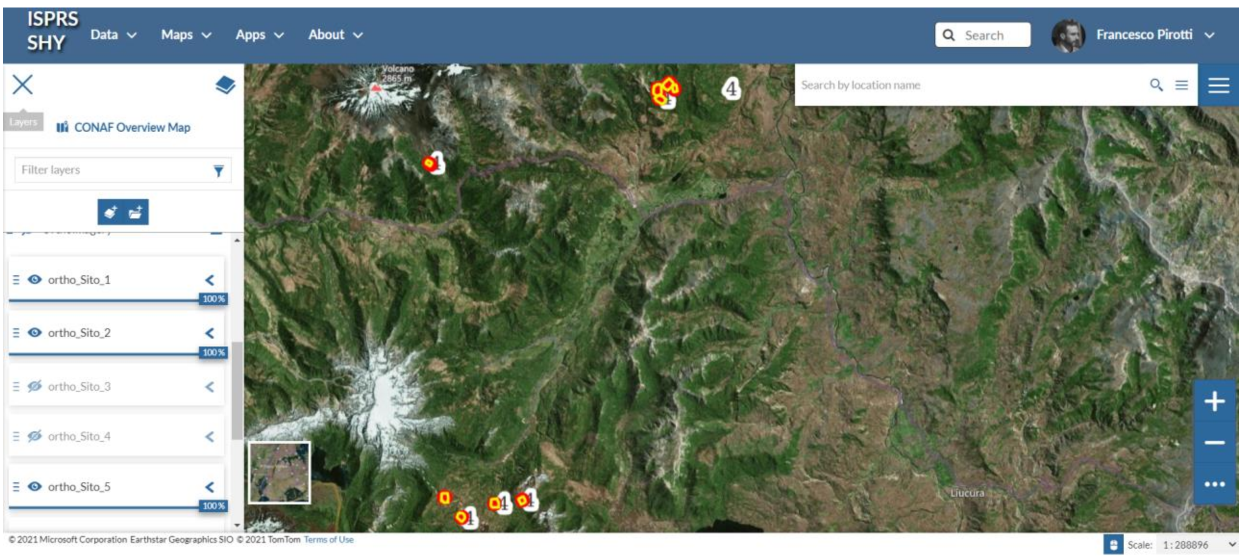

Figure 2. Study areas from one project with several high resolution orthoimages and ground truth plots.

3. RESULTS AND DISCUSSION

3.1 Pilot projects

The three projects that were used as pilots had different types of data and were used to test the portal. The first project is VARCITIES a European Union Horizon 2020 Research and Innovation programme. This programme includes data from IoT sensors, position data from smartphones and a 3D model from a laser scanning drone survey (Pirotti et al., 2022). The overall objective of the VARCITIES partners that use ISPRS-SHY is to study wellbeing of visitors of the historical garden of a cultural heritage site, Villa Revedin Bolasco. To this aim the sensors will collect microclimatic and air quality data at static points in the area and also via a mobile rover that will carry a set of sensors around the garden. The drone flight collected 3D point clouds from three types of surveys, two with lidar sensors (Riegl VUX-3 and VUX-120) and one with photogrammetry. This produced two point clouds consisting in 1 billion points, and one point cloud with 300 million points (figure 3).

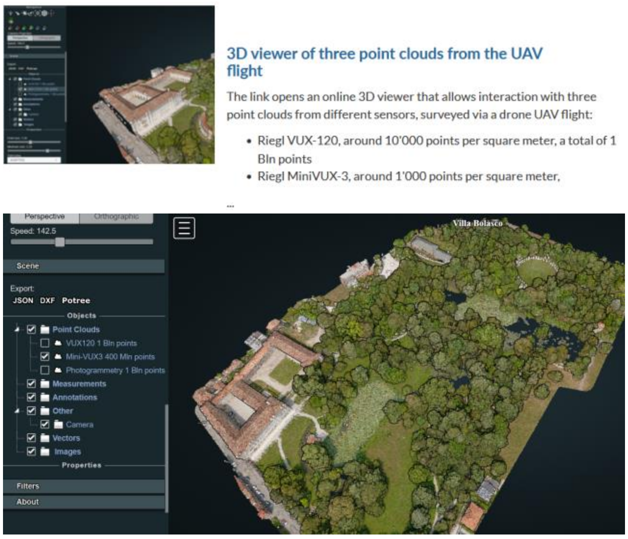

Figure 3. Large volumes of data from point clouds linked via metadata.

As mentioned, the large volume of data that point clouds represent can be loaded in the portal by linking a viewer, Potree (Schuetz, 2016), to the metadata that is loaded in ISPRS-SHY. With this solution, users can discover the data by using keywords or the spatial location, but the data itself is located in a third party server. This solution optimizes the data mining requirements with the space requirements of large volumes of data. It is of course obvious that uploading to the ISPRS-SHY portal several GB of point cloud data is not an effective solution, also considering that the Geonode infrastructure provides viewers for raster and vector data, but not for 3D point clouds. Therefore integration of a separated viewer (Potree) using the Geonode infrastructure to link the data viewer with the metadata, improves efficiency. Three-dimensional data are becoming more common in the geospatial community, and Geonode and other data sharing solutions would get much added value if viewers for this type of data can be added. Potree is a very versatile solution that might be embedded in Geonode. An upload system to deal also with 3D data formats, e.g. LAS/LAZ format (ASPRS, 2015) is feasible to implement to stream LAS format to Potree structure and automatically create the viewer. This might be a future development.

Another project is a national project that aimed at analysing the effects of an extreme event, the VAIA storm. VAIA damaged large parts of production forests in the north-east part of Italy (Piragnolo et al., 2021). This led to lying deadwood that will cause further risks related to fire (Marchi et al., 2018) and pathogens (e.g. bark beetle attacks).

The VAIA project not only has geographic data and maps loaded in the portal, but also links to non-spatial data such as newspaper articles (figure 4 top), and the proposal itself (figure 4 bottom).

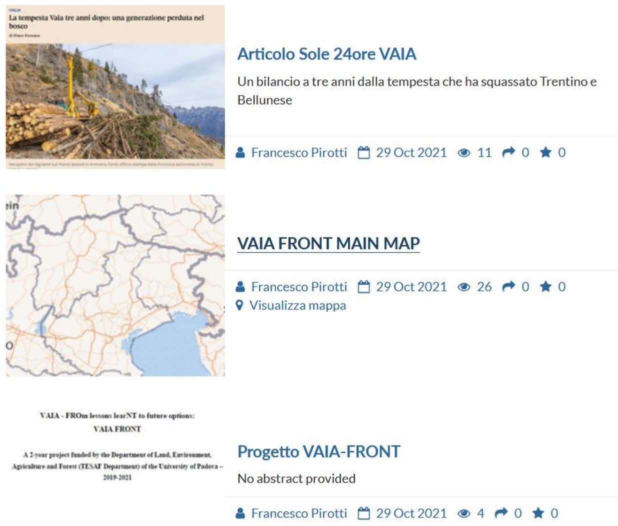

Figure 4. Three items from the VAIA project loaded in ISPRS-SHY – from top to bottom; top is a newspaper article, middle is the map, and bottom is the overall project proposal, which is available as linked webpage.

Remote sensing data and ground surveys have been stored in the ISPRS-SHY platform to cross-analyse information. Not only products from ad-hoc processing, but also parts of existing satellite-derived products, such as the Forest Cover dataset by (Hansen et al., 2013). This allows users to access data relative to the area and to the event, i.e. in this case the forest cover loss caused by the VAIA storm, and visualize them.

The third project is from the Chilean CONAF "Postulación Fondo de Investigación del Bosque Nativo" titled "Indicadores fenológicos y estructurales de alteración de hábitat en bosques de araucaria". In this project 12 stands of Araucania-Nothofagus forest were selected in southern Chile, which represented four alteration levels (Kutchartt et al., 2021; Pirotti et al., 2020). Drone flights acquired images that produced ortho-imagery and point clouds that were further analysed in the project. A total of 379 plots over the 12 stands collected ground data regarding tree species, and other tree parameters. Six users accessed ISPRS-SHY to load and share the project data (**Figure 5**). The data from this survey was saved in a table connected to the geolocations of the plots that were surveyed with a GNSS receiver and stored as vector points with a unique id for each plot. ISPRS-SHY allowed to view the plot position and the data relative to each plot (**Figure 6**).

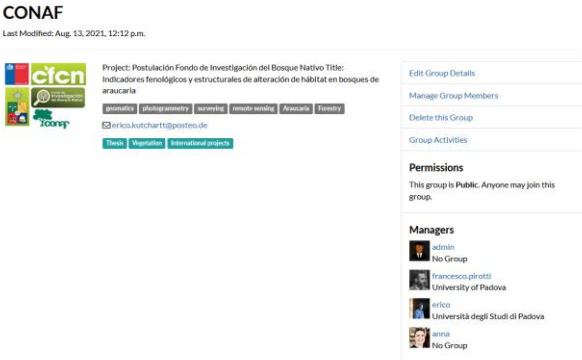

Figure 5. Example of project group users, with description of the project.

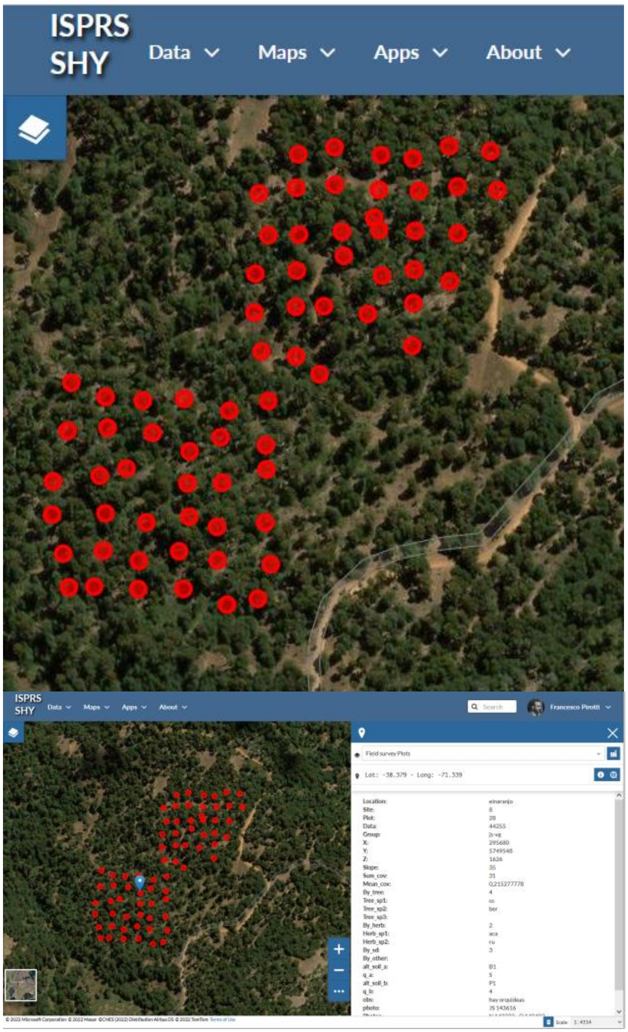

Figure 6. Plots from CONAF-financed project with relative data form ground samples.

3.2 Non-spatial datasets

Each dataset that is uploaded becomes a spatial layer or a document layer, depending on the type of dataset. Non-spatial datasets can be anything related to the project but without direct spatial information, e.g. newspaper articles, research papers. These types of data can be linked to spatial data and maps from the same project, so they can indirectly be assigned to a spatial object. Accessing the data is immediate as a link can be provided to users that want to download or simply view the data with the online tools. Depending on the data sharing policy that is given to each layer, it can be shared also as open access, without requiring the user to register to the platform, or to specific users.

The user profiling method allows to assign to users projects to that there is a many-to-many relationship between users and projects (Figure 7).

Figure 7. Single user with profile of relative projects and groups.

3.3 Geonode and Geoserver

Another aspect to be discussed is the integration between Geonode and Geoserver which provides several advantages, even if sometimes adds complexity to the system. As part of multi-component and modular open system (Brovelli et al., 2018; Coetzee et al., 2020), Geonode can interact with other open software. Geoserver is “an open source server for sharing geospatial data” using open standards. The advantage in this integration is that geospatial data can be organized via geoserver and be stored either in files or in a Postgresql database, thus improving interoperability. In particular, data can be provided also as Open Geospatial Consortium OGC services thanks to Geoserver. This means that Geonode can ingest data from the OGC service, e.g. through a Web Mapping Service (WMS), or, viceversa, third-party subjects interested in using the data can also use the WMS service.

3.4 Geostories

Webgis solutions are now becoming common also to transfer knowledge via “storytelling”. This becomes effective not only because data is provided via a web platform (Caradonna, 2017), but also because information can be assigned to space and to the territory. Therefore, the impact of extreme events like in the case of the VAIA-FRONT project (Figure 8) can convey what happened where with a very direct format. This can be done in the SHY platform, using the spatial data and the non-spatial documents linked to a specific project. This can also be used for public engagement in case of urban planning, where knowledge of the current state of a town and the future changes is of interest to the local residents (Damiano et al., 2015).

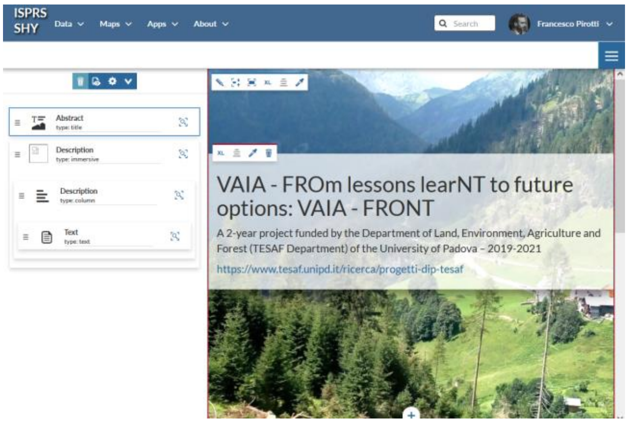

Figure 8. Example of a Geostory related to the VAIA-FRONT project.

4. CONCLUSIONS

Initial tests used four recent projects from the authors. Thirty-nine items are shared, consisting in geodata, articles and maps that aggregate geodata in layers in a single map (Figure 2). Twelve users so far have tested the platform. Some initial considerations are worth reporting. One is that registration via social media seems a trivial addition, but definitively increases the accessibility to the portal. Since sharing is an important aspect, faster access helps to attract users. Another consideration is that correct and detailed metadata completion allows better data mining. Often users tend to upload data without providing accurate information, making it difficult to search for the data as keywords are looked up from the inserted metadata. The title of the uploaded product is important as well, as it is the first text that is associated with the product.

The evolution of the ISPRS-SHY portal is to foster more collaboration between research groups by sharing more data like benchmarks (ISPRS and other societies) and projects. Some ongoing projects already involved in ISPRS-SHY (VARCITIES) will be pilot-cases for integration of further geodata. In particular in the near future the implementation of IoT gateways and devices for this project will be uploaded and data streamed and integrated via PostgreSQL. This area will be monitored over several years for air quality and climate information with data that has to be shared with other partners, and ISPRS-SHY will be the central node for this goal.

ACKNOWLEDGEMENTS

The creation of ISPRS-shy portal and this work was supported by the ISPRS Scientific Initiatives 2021. Data and testing were provided by Dept. TESAF, University of Padova in the framework of the project “Vaia FRONT” (CAVA_SID19_02), and by the project “Indicadores fenológicos y estructurales de alteración de habitat en bosques de araucaria” (016/2019), funded by the Fondo de Investigación del Bosque Nativo de la Corporación Nacional Forestal (CONAF) and by the Ministry of Agriculture of Chile.

REFERENCES

ASPRS, 2015. LASer (LAS) File Format Exchange Activities – ASPRS [WWW Document]. URL https://www.asprs.org/committee-general/laser-las-file-format-exchange-activities.html (accessed 2.13.19).

Brovelli, M.A., Minghini, M., Zamboni, G., 2018. New Generation Platforms for Exploration of Crowdsourced Geo-Data, in: Earth Observation Open Science and Innovation. Springer International Publishing, Cham, pp. 219–243. https://doi.org/10.1007/978-3-319-65633-5_9

Caradonna, G., 2017. WebGIS interattivo per il trattamento dei dati digitali del terreno (DTM) via web. Boll. della Soc. Ital. di Fotogramm. e Topogr. 28–31.

Coetzee, S., Ivánová, I., Mitasova, H., Brovelli, M.A., 2020. Open geospatial software and data: A review of the current state and a perspective into the future. ISPRS Int. J. Geo-Information 9, 90.

Damiano, M., Dardanelli, G., Ferrara, S., 2015. Un SIT per il PPE del centro storico di Palermo: il caso studio del Mandamento Castellammare. Boll. della Soc. Ital. di Fotogramm. e Topogr. 4–9.

Hansen, M.C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Science (80-.). 342, 850–853. https://doi.org/10.1126/science.1244693

Kutchartt, E., Gayoso, J., Pirotti, F., Bucarey, A., Guerra, J., Hernández, J., Corvalán, P., Drápela, K., Olson, M., Zwanzig, M., 2021. Aboveground tree biomass of Araucaria araucana in southern Chile: measurements and multi-objective optimization of biomass models. iForest - Biogeosciences For. 14, 61–70. https://doi.org/10.3832/ifor3492-013

Marchi, N., Pirotti, F., Lingua, E., 2018. Airborne and Terrestrial Laser Scanning Data for the Assessment of Standing and Lying Deadwood: Current Situation and New Perspectives. Remote Sens. 10, 1356. https://doi.org/10.3390/rs10091356

Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E., Sartori, L., Mozzato, D., Gatto, P., Defrancesco, E., Bortolini, L., Pirotti, F., Pisani, E., Sartori, L., 2018. The Role of Factors Affecting the Adoption of Environmentally Friendly Farming Practices: Can Geographical Context and Time Explain the Differences Emerging from Literature? Sustainability 10, 3101. https://doi.org/10.3390/su10093101

Pagliacci, F., Defrancesco, E., Mozzato, D., Bortolini, L., Pezzuolo, A., Pirotti, F., Pisani, E., Gatto, P., 2020. Drivers of farmers’ adoption and continuation of climate-smart agricultural practices. A study from northeastern Italy. Sci. Total Environ. 710, 136345. https://doi.org/10.1016/j.scitotenv.2019.136345

Piragnolo, M., Pirotti, F., Zanrosso, C., Lingua, E., Grigolato, S., 2021. Responding to Large-Scale Forest Damage in an Alpine Environment with Remote Sensing, Machine Learning, and Web-GIS. Remote Sens. 13, 1541. https://doi.org/10.3390/rs13081541

Pirotti, F., Kutchartt, E., Csaplovics, E., 2020. Assessment of Volume and Above-Ground Biomass in Araucaria Forest Through Satellite Images, Comparing Different Methods in the South of Chile, in: 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS). IEEE, pp. 579–584. https://doi.org/10.1109/LAGIRS48042.2020.9165668

Pirotti, F., Laurin, G., Vettore, A., Masiero, A., Valentini, R., 2014. Small Footprint Full-Waveform Metrics Contribution to the Prediction of Biomass in Tropical Forests. Remote Sens. 6, 9576–9599. https://doi.org/10.3390/rs6109576

Pirotti, F., Piragnolo, M., Guarnieri, A., Boscaro, M., Cavalli, R., 2022. Analysis of Geospatial Behaviour of Visitors of Urban Gardens: Is Positioning via Smartphones a Valid Solution?, in: Borgogno-Mondino, E., Zamperlin, P. (Eds.), Geomatics and Geospatial Technologies. Springer International Publishing, Cham, pp. 351–365. https://doi.org/10.1007/978-3-030-94426-1_26

Prataviera, E., Romano, P., Carnieletto, L., Pirotti, F., Vivian, J., Zarrella, A., 2021. EUReCA: an open-source Urban Building Energy Modeling tool for the efficient evaluation of cities energy demand. Renew. Energy 1–17. https://doi.org/10.1016/j.renene.2021.03.144

Schuetz, M., 2016. Potree: Rendering Large Point Clouds in Web Browsers. Vienna University of Technology.

Vaglio Laurin, G., Hawthorne, W., Chiti, T., Di Paola, A., Cazzolla Gatti, R., Marconi, S., Noce, S., Grieco, E., Pirotti, F., Valentini, R., 2016. Does degradation from selective logging and illegal activities differently impact forest resources? A case study in Ghana. iForest - Biogeosciences For. e1–e9. https://doi.org/10.3832/ifor1779-008