A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens

Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela & Iivari Kunttu

> " *A single observation that is inconsistent with some generalization points to the falsehood of the generalization, and thereby 'points beyond itself'.* "
>
> Ian Hacking
> Philosopher of science

Drone imaging has been shown to have increasing value in monitoring and analysing different kinds of processes related to agriculture and forestry. In long-term monitoring and observation tasks, huge amounts of image data are produced and stored. Environmental drone image datasets may have value beyond the studies that produced the data. A collection of image datasets from multiple data producers can, for example, provide more diverse training input for a machine learning model for vegetation classification, compared with a single dataset limited in time and location. To ensure reproducible research, research data such as image datasets should be published in usable and undegraded form, with sufficient metadata. Timely storage in a stable research data repository is recommended, to avoid loss of data. This work presents research datasets of 2020 drone images acquired from agricultural and forestry research sites of Häme University of Applied Sciences, and from Hämeenlinna urban areas. Those images that do not contain personal data are made freely available under a Creative Commons Attribution license. For images containing personal data, such as images of private homes, privacy-preserving forms of data sharing may be possible in the future.

Introduction

The development of digitalization and measurement technologies in recent decades has enabled digital devices and sensors to produce huge amounts of data that has great potential in optimizing processes related to production chains or service production. In the field of bioeconomy, the main production processes are related to food and biomass production. Digitalization provides a wide variety of opportunities to support, manage, and monitor production based on data collected from the field.

Image-based data collection and analysis provides a huge potential to support these goals. Visual data collected from agricultural fields enables automated analysis tasks and can provide real-time information on production status. To acquire suitable visual information, basically two different alternatives exist:

remote sensing based on satellite imaging, and drone (that is, unmanned aerial vehicle, UAV) imaging. In this paper, the main application fields studied are agriculture, forestry, and private urban gardens, in all of which remote sensing imaging can have many different purposes.

Precision farming technologies aim to optimize the use of farming inputs both spatially and temporally for improved economic outcomes and reduced environmental impacts of farming. In precision farming, a field is considered as a heterogeneous entity with variable topography, soil properties, weed infestation, and yield potential, whereby management practices are tailored spatially and temporally (Finger et al., 2019). Precision farming thus strongly relies on site-specific sensing of variables that are essential for management decisions. Georeferencing techniques and spatial mapping are important elements in precision farming.

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

Spatial information can be collected with scanners mounted in tractors (Pallatino et al., 2019) or using satellite imagery (Segarra et al., 2020).

As well, drones have increasingly been used to collect data on several features relevant for precision agriculture (Tsouros et al., 2019). Compared with satellite-based remote sensing, drone technology can produce images with considerably higher spatial resolution in the centimeter range. Also, the temporal resolution of drone-based imagery can be decided by the user, which leads to flexibility in comparison with satellite data. Drones have also been used for research purposes in monitoring field experiments (Viljanen et al., 2018, Dehkordi et al., 2020). Non-destructive monitoring of vegetation is a major benefit for practical agronomy as well as for research use. Drone aerial imaging has been utilized in a wide range of agricultural applications.

Some of the most common applications of drone imaging in precision agriculture are weed mapping and management, vegetation growth monitoring and yield estimation, vegetation health monitoring, and irrigation management (Tsouros et al., 2019). Imaging has been used in monitoring many vegetation traits, for example, biomass amount (ten Harkel et al., 2020), nitrogen status (Caturegli et al., 2016), moisture and plant water stress status (Hoffmann et al., 2016), temperature (Sagan et al., 2019), and various vegetation indices (Viljanen et al., 2018). Deep learning-based prediction of crop yield from drone aerial images has also shown promising results (Nevavuori et al., 2019, Nevavuori et al., 2020).

In the field of forestry, drone-based photogrammetric methods can be used in several different ways. The methods used may provide general forest inventory data that focuses on common stand variables such as volume and height (Tuominen et al., 2017). Practical forest planning in Finland based on drone-collected photogrammetric data is rapidly advancing and is currently being piloted. Drones have proven especially useful in detection and inventory of various forest damage areas, such as windthrow areas (Mokros et al., 2017, Panagiotidis et al., 2019) and bark beetle outbreak areas (Näsi et al., 2015, Briechle et al., 2020). Drones have been successfully used in various forest fire suppression and prevention tasks for several years (Ollero et al., 2006, Akhloufi et al., 2020). Increasing

demand to safeguard forest biodiversity has also encouraged the use of photogrammetry-based methods. These methods have proven to be a useful inventory tool, when important structural factors such as keystone species like aspen (Viinikka et al., 2020), standing dead trees (Briechle et al., 2020), or coarse woody debris (Thiel et al., 2020) are located in a forest landscape.

In the area of private urban gardening, drone-based imaging may provide new approaches to monitor the effects of gardening practices on the vegetation and on carbon sequestration. In low-density housing areas, the surface coverage pattern is typically very diverse, consisting of numerous individual plots and gardens. Homeowners reshape and modify private domestic gardens based on personal preferences and individual gardening practices. The role and meaning of vegetation and gardening practices vary, resulting in plot-to-plot variations in carbon sequestration and evapotranspiration, which affects stormwater management and the degree of reduction in the urban heat island phenomenon. Approaches to sustainable urban development have put an increasing interest in low-density housing areas that cover large areas in cities. A single plot is not the main focus, but rather the entity they form together. This raises the challenge to find suitable methods for easily studying the on-going changes at multiple scales to provide data both on the quality and quantity of vegetation. Plot and block scale choices and elements define housing area scale attributes.

This article describes vegetation monitoring-related aerial image acquisition by Häme University of Applied Sciences (HAMK) using drones, in 2020, both the processes and experiences gained. Apart from the image data, the main research in three areas is to be published separately. In total, approximately 200,000 image files, approximately 1 TB in size, are in the process of being published openly (see *Data Availability*). Particular features of the presented datasets are including the original image files, using a multispectral camera, and that one of the research sites, Mustiala biochar field, was imaged several times over the growing season with some near-simultaneous satellite imagery available from public sources. Based on a search by the authors using Google Dataset Search (https://datasetsearch.research.google.com/) at the time of writing, drone aerial image datasets are rapidly increasing in number, but are typically orders of

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

magnitude smaller than the datasets presented in this work and do not usually include the original image files, although they may be available upon request. In this article, we also discuss the benefits and challenges of publishing drone image datasets.

Methods

Drone aerial image datasets were collected from multiple sites in Kanta-Häme, in southern Finland. Three of the sites (Fig. 1) are presented in this work:

1. *Mustiala biochar field* (Fig. 6) is located at the Mustiala educational and research farm of Häme University of Applied Sciences, in Tammela (60˚49' N, 23˚45'24'' E). The biochar experiment consists of 10 adjacent plots, each of size 10 m × 100 m (1000 m²), and with a total area of 1 ha. Biochar soil amendment was applied on five of the plots at a rate of ca. 20 t/ha. The other five plots were control treatments without biochar amendment. Ground control points (GCPs) were placed covering the field in a roughly 100 m × 100 m die face-5 pattern (see Fig. 6). GCPs were 29 cm × 29 cm sized black-and-yellow 2x2 checkerboard-style cut-outs of A3 prints. They were georeferenced with the aid of a real-time kinematics (RTK) capable Trimble Geo 7X (H-Star) hand-held GPS receiver.

2. *Evo old forest* (Fig. 3) consists of seven separate stands all located in the Evo state forest. The stands, with a pooled area of 160 ha, are dominated by mature Norway spruce (Picea abies) with an age range of 80-120 years. All stands have a rather high amount of dead

standing trees, known to be important for biodiversity, for example, for cavity-nesting birds. The standing dead trees were catalogued in 2019-2020 to function as reference data for photogrammetric methods.

3. *Hämeenlinna private urban gardens* consist of approximately 5-10 domestic gardens in the sparsely populated urban areas of Hämeenlinna.

The drone used in all of the imaging missions was a DJI Matrice 210 RTK V2 quadcopter camera drone. The camera payload for each mission was selected (see Fig. 2) from the following cameras:

1. *DJI ZenMuse X5S FC6520* —a 3-axis gimbal-stabilized RGB camera with a 15 mm focal length lens, operated in sRGB JPEG still mode, software version 01.07.0044,

2. *DJI ZenMuse XT2* (radiometric) —a 3-axis gimbal-stabilized camera with a 13 mm focal length lens for the thermal sensor operated in radiometric JPEG mode, and an 8 mm focal length lens for the RGB sensor, software version 06.02.20, and

3. *Micasense Altum* —a radiometric multispectral camera with dedicated optics for each channel, separately timed and triggered from the other cameras, software version 1.3.6, with sunlight sensor DLS2.

Flights were planned in Dji Pilot software. Flight parameters were specified as best fit for the present environment, light and weather conditions, area size, and camera type. Flight parameters for Evo old forest

Figure 1. Research site locations (white dots) in Finland.

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

were the following: altitude 120 m above ground, side overlap 80%, frontal overlap 80%, speed 3-5 m/s, camera triggering mode: time interval. For the Mustiala biochar field, the following parameters were mostly used: altitude 80 m, side overlap 85%, frontal overlap 83%, and speed 2-3 m/s. The field was imaged multiple times over the growing season, using identical parameters. For Hämeenlinna private urban gardens, flight parameters were as follows: altitude 50 m, side overlap 85%, frontal overlap 85%, and speed 3 m/s. When cameras were used simultaneously, overlap was specified for the Micasense Altum, resulting in a higher overlap for other cameras. Flights were scheduled (Fig. 2) between 11.00 am and 15.00 pm (UTC+3), from May to August 2020.

The DJI gimbal cameras were nadir-pointing, that is, straight down, and synchronously triggered by the DJI Pilot software. The Micasense Altum was triggered by its own timer and pointed straight down in the drone's internal coordinate system. The Micasense Altum images of a Micasense calibrated reflectance panel (Fig. 5) were taken before or after each imaging mission, or both. A Geotrim Trimnet VRS virtual RTK station was used, while the DJI cameras received RTK GPS information from the drone. The Micasense Altum utilized its own GPS receiver.

For the figures used in this article, Agisoft Metashape Professional version 1.7.0 (Agisoft 2020a) was used to color-correct the Micasense Altum images and to generate an orthomosaic of the Mustiala biochar field using the workflow described in Agisoft (2020b), without the use of ground control points. For Figures 5 and 7, single Micasense Altum photos from the Mustiala biocarbon field mission dated 2020-05-22,

10:00–12:00 (UTC), were exported in calibrated form from Agisoft MetaShape, and their blue, green, and red spectral channels were stacked (aligned) in Adobe PhotoShop version 21.2.4 with distortion correction. The Micasense Altum radiance and reflectance images (Figs. 4-7) and the Sentinel-2 satellite image of Figure 6 were converted to the sRGB color space in GIMP version 2.10.18 by assigning an sRGB gamma=1 color profile, by adjusting brightness in the curves tool using a linear ramp crossing the origin, and by converting to sRGB color profile using a relative colorimetric rendering intent. Geographical illustrations were made in QGIS version 3.12.2 (QGIS Development Team 2020).

For a comparison with the aerial images, a Sentinel-2 satellite image (Fig. 6) of the Mustiala biochar field was manually selected and retrieved from the Copernicus Open Access Hub (Copernicus Sentinel Data 2020) for a cloud-free day that coincided with a drone imaging mission on 2020-05-22. For Figure 7, machine learning image segmentation of sRGB-color space images in 0.1 m / pixel resolution was done using the DroneDeploy Aerial Segmentation Benchmark U-Net model "keras baseline" run gg1z3jjr by Stacey Svetlichnaya (DroneDeploy 2019), using a tile size of 300 × 300 pixels.

Results

The GCP location data and most of the acquired aerial images are being made publicly available (see section *Data Availability*). Figure 3 shows the camera locations for all individual images taken during the Evo old forest imaging missions. Figure 4 shows a sample image from each camera from a Mustiala biochar field imaging mission on 2020-05-22. Before that flight, an image was taken of the calibrated reflectance panel (Fig. 5). Near-
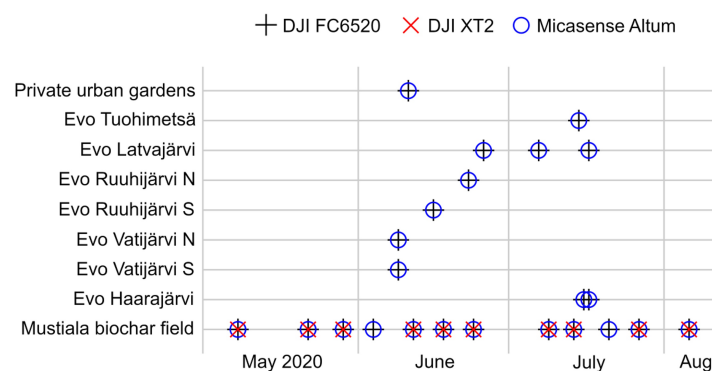

Figure 2. The 2020 imaging mission schedule for the research sites, with the camera payload indicated.

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*
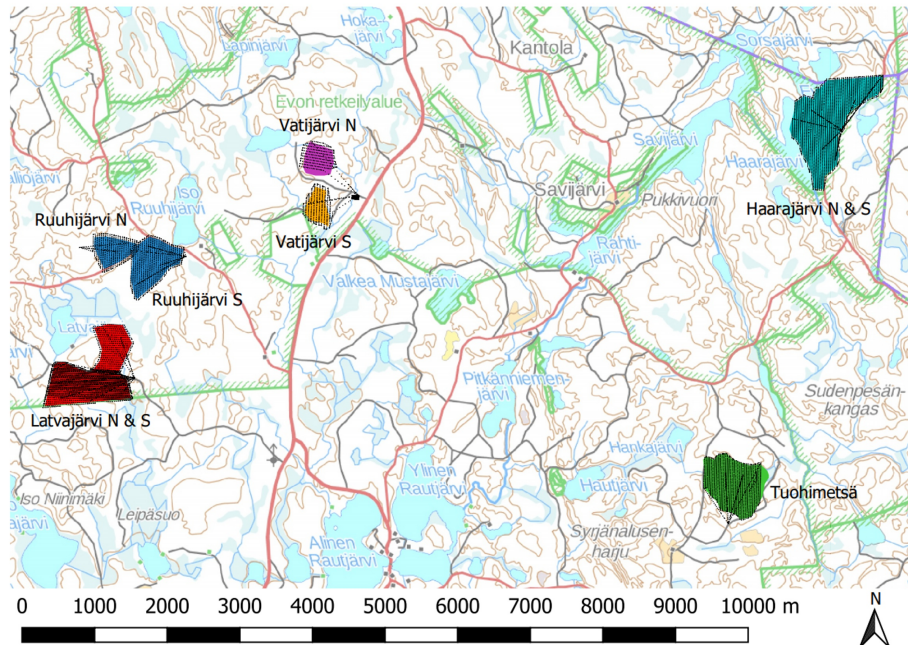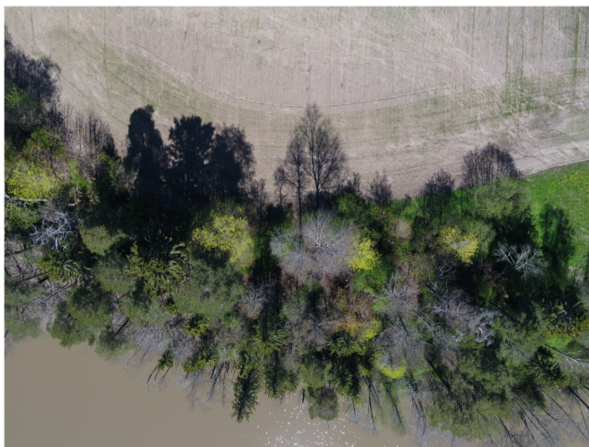

Figure 3. Evo old forest research sites (colored areas) and camera GPS locations of individual drone images (black dots). Some Micasense Altum images were taken on the way to or from the takeoff and landing site due to lack of drone-camera communication. (Map: National Land Survey of Finland Topographic Database 01/2021.)

Figure 4. Uncalibrated sample drone images from Mustiala biochar field, captured with the drone flying at an altitude of 80 m from the ground. The Micasense Altum spectral channel images are (from left to right): blue, green, red, near infrared, red edge, and thermal. The DJI XT2 images were obtained on a separate flight and are visible (left) and thermal (right).

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

Figure 5. A calibrated reflectance panel (Micasense) placed on the Mustiala biochar field, imaged hand-held using the Micasense Altum multispectral camera.

Figure 6. The Mustiala biochar field on 2020-05-22. Left: A 4 cm resolution drone orthomosaic, an artificial view straight down, made of Micasense Altum multispectral camera images without the use of ground control points (GCPs), acquisition time 10:20 to 10:27 (UTC). The geolocation error of GCPs was at most 0.8 m. Right: A 10 m resolution Sentinel-2 satellite bottom-of-atmosphere corrected reflectance image (Copernicus Sentinel Data 2020), acquisition time 09:50 (UTC).

simultaneous drone and satellite imagery from that day are shown for comparison in Figure 6. A single image and an orthomosaic built from multiple images are presented for comparison in Figure 7, together with their machine learning image segmentation. For the Mustiala biochar field, the location of the GCPs were resolved with a horizontal and a vertical accuracy of 0.1 m, as reported in the shapefile from the hand-held RTK GPS receiver.

Discussion and Conclusions

We encountered many technical challenges during data collection. Interoperability problems between drone and camera systems from different manufacturers prevented camera triggering by the drone and tagging images with high-accuracy RTK GPS information. A possible time zone misconfiguration affected image time stamping. Although automatic flight logs were generated, they were

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*
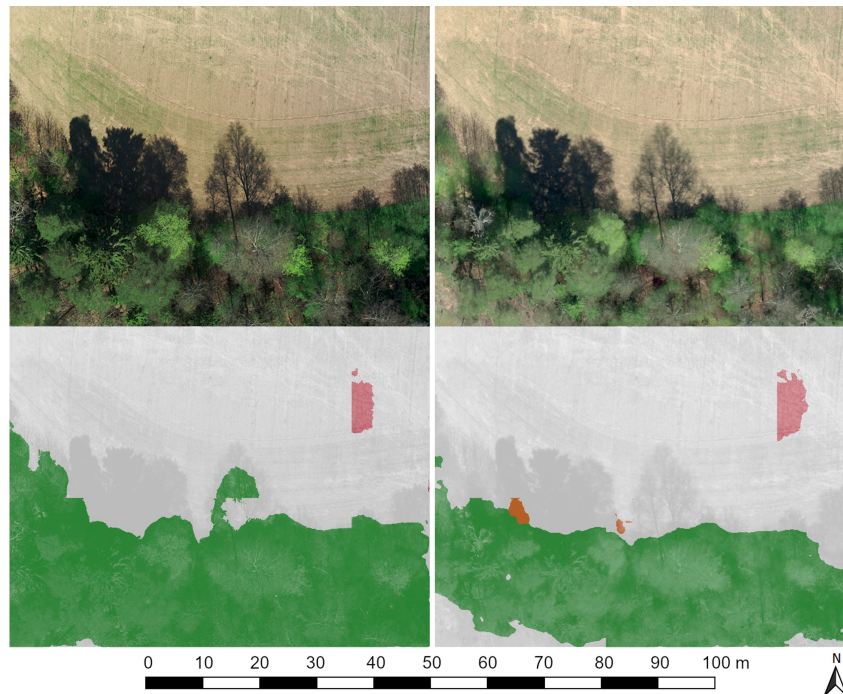

Figure 7. A single drone image (top left) aligned to a drone orthomosaic (top right), both using the Micasense Altum camera. The single image shows fine detail of complex objects better than the orthomosaic, at the cost of not being an orthoimage. Using an off-the-shelf U-Net machine learning model trained on orthomosaics (DroneDeploy 2019), the images were segmented into classes: ground (white), vegetation (green), building (red), water (orange), and cars and clutter (not found), overlaid on the single image (bottom left) and the orthomosaic (bottom right).

insufficient to answer all questions arising in later interpretation of the collected data, which resulted in increased work wrangling the data. It would be advisable for further work going forward to complement automatic logs with notes about the settings used and both operator intent and actions.

Selecting the flight parameters turned into something of an "art in itself". Flight speed, altitude, side and frontal overlap, camera orientation, and triggering mode significantly influenced the image results. For example, flight altitude requires a compromise between image resolution and flight time. Lowering the altitude increases the flight time, which is accompanied by possible in-flight battery depletion, making further use of the images more complicated by more dynamic scenes.

For photometric applications, rapidly changing light conditions can be a problem even for short flights. As well, movement from the combination of windy weather and vegetation invalidates the assumption of a

static scene, made for example in a post-processing pipeline that produces a static 3-d point cloud, either as an intermediate step or as the final product. The planning of drone aerial imaging has been studied and reviewed by (Tmušić et al., 2020), which covers choices such as camera angle and flight pattern.

Much consideration should be given to the process of publication of research data from high-resolution aerial imaging missions. Aspects to consider include data storage and availability, software compatibility, the rights of data producers, the rights of possible data subjects, license agreements of processing software, and others.

Open publication of research data improves the reproducibility of science, and reduces barriers to participating in science and utilizing scientific data and results, especially for under-resourced and under-represented participants. For academic data producers, the growing recognition of *data as research output* (see San Francisco Declaration on Research Assessment

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

2012) may bring financial incentives to more widely publish research data. Incentives by science funders may be applied retroactively. A major practical reason to publish research data is the possibility that a dataset may have tremendous utility value outside the research project, far beyond the organization that produced it. Many possible later uses of data cannot be anticipated at the time of its creation and collection.

Not all research data that is openly published remains available. As an example, Khan et al. (2021) were only able to retrieve 94 out of 121 open-access medical ophthalmology imaging datasets. Storing research data in an established repository such as Zenodo (https://zenodo.org/) gives a level of guarantee of data longevity. It also allows obtaining a Digital Object Identifier (DOI) for sharing and citing the data. Upon a recent successful storage quota application by HAMK, the image datasets presented in this article are being stored and published in Fairdata services, funded by the Ministry of Education and Culture (Finland). Publication of research data in a repository effectively forces the storage of the data together with metadata describing the data, ownership of the data, and its usage license. The additional information resolves many ambiguities when using the data. Structured metadata in repositories ensures the dataset is indexed in research databases.

A repository may also allow incremental publication of data. If publishing the data takes place this way, already during its collection instead of at the end of a research project, then use of and citation of the data outside of the data producing organization can start much earlier. Accelerated publication also ensures that the data or information of concern is not lost during the project or when personnel leave the project. Academic data producers typically have an interest in priority publication of their research. Early publication of general-use research data, such as the image datasets presented in this article, is less likely to conflict with that interest, compared to early open publication of all data vital to the main publication.

Location data from an image or other information about a person or object that can be associated with a person, either directly or using additional information, is likely to be considered as "personal data" by the EU's General Data Protection Regulation (GDPR). Examples of such objects in aerial imagery include vehicles, land,

and buildings managed by a private person, or by their family. GDPR requires that before personal data can be handled for a purpose, the data subject's voluntary consent for that purpose must be obtained. A data subject also has the right to be forgotten, that is, to bindingly request that their personal data be erased without excessive delay. National laws may still allow legitimate scientific research that protects data privacy. Nevertheless, a data subject's rights may bar redistribution or open publication of personal data under an irrevocable license, such as any of the popular Creative Commons licenses, possibly even when a data subject proactively authors the data as free speech.

The image dataset concerning Hämeenlinna private urban gardens was collected with informed consent by the data subjects, the homeowners. It was deemed necessary to deposit the data in a fully closed manner, at least for the time being, while the legal landscape of personal data is still developing. The EU has proposed a Data Governance Act (COM/2020/767 final) following a European Strategy for Data (COM/2020/66 final). The Data Governance Act defines roles and mechanisms for altruistic data sharing in managed data ecosystems. Among other things, the act is intended to streamline handling of requests by authorized users to access data, on the condition that the data subject has consented to handling of the data for the requested purpose.

Returning to our case study, unlike the presented aerial images captured mostly 80 m above ground, images from a closer range (Fig. 5) would allow distinguishing not only trees, but also individual small plants. Likewise, it would be possible to identify and count the plants and analyze their physical characteristics (allometry) and health.

Drone image data is being increasingly utilized in machine learning, as exemplified by research cited above in the Introduction. The purpose of a machine learning model might be, for example, to segment an input image into different class labels (Fig. 7). Class label masks of drone imagery could be converted to lower-resolution ground truth class density data for interpreting satellite imagery. In *semi-supervised learning*, unlabeled images would also be included to help the model better capture the natural joint probability distribution of the images and segmentation. Such uses make general-use unlabeled data valuable as a research output, thereby complementing the existing

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

situation and future of labeled and unlabeled data. Machine learning also benefits from more diverse data, for example, from image datasets collected at an ecologically diverse set of locations. Big data in such cases can come from many small data.

Data producers may have reasons not to disclose their raw data. In such cases, other, somewhat futuristic forms of information sharing may be possible. In *federated learning*, data holders combine their efforts in a coordinated fashion to train a shared machine learning model, without communicating their original data. Alternatively, a *generative model* could be used to collect non-private artificial samples from the approximate distribution of the original data, while preserving the privacy of the original data points, typically measured by *differential privacy*. Depending on the amount of original data and required degree of privacy, generated samples may be of sufficient quality to be used similarly to the original data. For a privacy-preserving generative method suitable for images, see Chen et al. (2020).

The author of a dataset consisting of drone images may wish to limit data use. A photograph taken for scientific purposes by equipment under automatic control is unlikely to be considered as creative work and would not be protected by copyright as such. In the United States, a dataset can be copyrighted as a compilation if it is sufficiently original in selection, coordination, or arrangement, but the copyright of the dataset does not extend to any non-copyrightable data items (U.S. Copyright Office, 2021). Similarly, for EU-based authors, national implementations of the Database Directive (96/9/EC) enable copyright of a dataset as a creative collection. Separately from copyright, the directive enables sui generis protection of non-creative datasets based on substantial investment, with a 15-year term of protection (European Commission, 2018). In any case, if a separate contract is made between the dataset holder and its retriever, binding clauses in the contract may limit redistribution and use of the dataset by the retriever. For example, commercial use of a dataset by its retriever may be prohibited.

Legal aspects aside, when collecting and sharing data, care should be taken to ensure that the data will be delivered in a format that preserves sufficient quality, preferably in a format that is openly standardized, that is, not a proprietary, software-specific format.

Compression artifacts that arise from lossy image compression methods such as JPEG might interfere with radiometric analyses. On the other hand, compared to lossy compression, lossless image compression results in significantly larger image files, increasing the cost of storage and transfer. Storage space requirements of image datasets could be eased somewhat by re-encoding lossless TIFF files, using a more efficient lossless method than what is available from the camera. However, rewriting the image files might also detrimentally affect metadata or other extra data, reducing the usability of the files. Image file metadata could be restored by transferring it from the original file using tools such as ExifTool and Exiv2. Usability of any modified original files should be tested at least in the most likely processing pipelines available. Rewriting of image files may be necessary to mask out personal information.

For geographical image data, it is important to also publish metadata describing the radiometric quality and the location accuracy. Information about things that affect illumination, such as clouds and sun, as well as calibration images and light sensor data, can be important for future users. Processing pipelines should be described, and when possible, source code and operation environment or information likewise included. For more information on quality assurance data and other important metadata, see Aasen et al. (2018) and Tmušić et al. (2020).

The image datasets presented in this work consist mainly of unaltered raw images directly from cameras. Publishing the raw data without embargo also became an effective way to distribute the data within the HAMK organization, as well as outside it. Another rationale behind the decision to publish not only post-processed data products, but also raw primary data, was that data users might wish to apply their own processing pipelines to ensure uniform processing of all their input data. An example of a data product is an orthomosaic (Fig. 6), which is straightforward to use in various applications and valuable in providing a visual overview of data. As demonstrated in Figure 7, the visual clarity of an orthomosaic might not be quite as high as that of the source material, the individual images, with differences that affect labeling by a machine learning model. In the future, novel photogrammetry pipelines will likely result in data products of higher quality than what is achievable using today's tools – if the raw data is still available.

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

Data availability

The collected agricultural and forestry image datasets are in the process of being incrementally published under the Creative Commons by Attribution 4.0 International (CC-BY 4.0) open access license, and made available for download at (Häme University of Applied Sciences, 2021a https://doi.org/10.23729/895dfb4d-9a14-41cc-a624-727375275631 and 2021b https://doi.org/10.23729/d083d6ad-aa68-4826-8792-7a169e2d2dd9). The Hämeenlinna private urban gardens drone image dataset (Häme University of Applied Sciences 2121c) has restricted access due to personal data content. Other associated data such as calibration data and GCP coordinates are published with the datasets.

References

Aasen, H., Honkavaara, E., Lucieer, A., & Zarco-Tejada P.J. 2018. Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows. Remote Sensing. 10(7): 1091.
DOI: 10.3390/rs10071091

Agisoft 2020a. Agisoft MetaShape Professional (Version 1.7.0) (Software).
http://www.agisoft.com/downloads/installer/

Agisoft 2020b. MicaSense Altum processing workflow (including Reflectance Calibration) in Agisoft Metashape Professional.
https://agisoft.freshdesk.com/support/solutions/articles/31000148381-micasense-altum-processing-workflow-incl-reflectance-calibration-

Akhloufi, M., Castro, N., & Couturier, A. 2020. Unmanned Aerial Systems for Wildland and Forest Fires: Sensing, Perception, Cooperation and Assistance. arXiv:2004.13883 [cs.RO]. *Drones*, 5(1): 15.
DOI: 10.3390/drones5010015

Briechle, S., Krzystek, P., & Vosselman, G. 2020. Classification of tree species and standing dead trees by fusing UAV-based lidar data and multispectral imagery in the 3D deep neural network PointNet++. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2: 203-210.
DOI: doi.org/10.5194/isprs-annals-v-2-2020-203-2020

Carlini N., Tramer F., Wallace E., Jagielski M., Herbert-Voss A., Lee K., Roberts A., Brown T., Song D., Erlingsson U., Oprea A., & Raffel C. 2020. Extracting Training Data from Large Language Models.
DOI: arXiv:2012.07805 [cs.CR].

Caturegli, L., Corniglia, M., Gaetani, M., Grossi, N., Magni, S., Migliazzi, M., Angelini, L., Mazzoncini, M., Silvestri, N., Fontanelli, M. & Raffaelli, M. 2016. Unmanned aerial vehicle to estimate nitrogen status of turfgrasses. *Plos One*, 11(6): 0158268.
DOI: 10.1371/journal.pone.0158268

Chen, D., Orekondy, T., Fritz, M. 2020. Gs-wgan: A gradient-sanitized approach for learning differentially private generators.
DOI: arXiv:2006.08265 [cs.LG].

Dehkordi, R.H., Denis, A., Fouche, J., Burgeon, V., Cornelis, J.T., Tychon, B., Gomez, E.P., Meersmans, J. 2020. Remotely-sensed assessment of the impact of century-old biochar on chicory crop growth using high-resolution UAV-based imagery. *International Journal of Applied Earth Observation and Geoinformation*, 91: 102147.
DOI: /10.1016/j.jag.2020.102147

DroneDeploy. 2019. Aerial Segment by DroneDeploy. https://wandb.ai/dronedeploy/dronedeploy-aerial-segmentation/benchmark/.

European Commission. 2018. Commission staff working document: Evaluation of Directive 96/9/EC on the legal protection of databases.
https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection.

Finger, R., Swinton, S.M., El Benni, N., Walter, A. 2019. Precision Farming at the Nexus of Agricultural Production and the Environment. *Annual Review of Resource Economics*, 11(1): 313-335.
DOI: 10.1146/annurev-resource-100518-093929

ten Harkel, J., Bartholomeus, H., Kooistra, L. 2020. Biomass and Crop Height Estimation of Different Crops Using UAV-Based Lidar. *Remote Sensing*, 12(1): 17.
DOI: /10.3390/rs12010017

Hoffmann, H., Jensen, R., Thomsen, A., Nieto, H., Rasmussen, J., & Friborg, T. 2016. Crop water stress maps for an entire growing season from visible and thermal UAV imagery. *Biogeosciences*, 13(24): 6545-6563.
DOI: /10.5194/bg-13-6545-2016

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens   *Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu*

Häme University of Applied Sciences 2021a. Drone image dataset: Evo old forest 2020. https://doi.org/10.23729/895dfb4d-9a14-41cc-a624-727375275631.

Häme University of Applied Sciences 2021b. Drone image dataset: Mustiala biochar field 2020. https://doi.org/10.23729/d083d6ad-aa68-4826-8792-7a169e2d2dd9.

Häme University of Applied Sciences 2021c. Drone image dataset: Hämeenlinna private urban gardens 2020. https://doi.org/10.23729/0b8dba76-0e7f-48ad-9979-462ed9d49021.

Khan, S.M., Liu, X., Nath, S., Korot, E., Faes, L., Wagner, S.K., Keane, P.A., Sebire, N.J., Burton M.J, Denniston, A.K. 2020. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digital Health*, 3 (1): e51-e66.
DOI: 10.1016/s2589-7500(20)30240-5

Micasense 2020. RedEdge Camera Radiometric Calibration Model. https://support.micasense.com/hc/en-us/articles/115000351194-RedEdge-Camera-Radiometric-Calibration-Model

Mokros, M., Výbošt'ok, J., Merganič, J., Hollaus, M., Barton, I., Koreň, M., Tomaštík, J., Černava, J. 2017. Early Stage Forest Windthrow Estimation Based on Unmanned Aircraft System Imagery. *Forests*, 8(9): 306.
DOI: 10.3390/f8090306

Nevavuori, P., Narra, N., Lipping, T. 2019. Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture*, 163: 104859.

DOI: //doi.org/10.1016/j.compag.2019.104859

Nevavuori, P., Narra, N., Linna, P., & Lipping, T. 2020. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sensing*, 12(23): 4000.
DOI: 10.3390/rs12234000

Näsi, R., Honkavaara, E., Lyytikäinen-Saarenmaa, P., Blomqvist, M., Litkey, P., Hakala, T., Viljanen, N., Kantola, T., Tanhuanpää, T.-M., Holopainen, M. 2015. Using UAV-based photogrammetry and hyperspectral imaging for mapping bark beetle damage at tree-level. *Remote sensing*, 7(11): 15467-15493.
DOI: 10.3390/rs71115467

Ollero, A. and Merino, L., 2006. Unmanned aerial vehicles as tools for forest-fire fighting. *Forest Ecology and Management*, 234(1): S263.
DOI: 10.1016/j.foreco.2006.08.292

Panagiotidis, D., Abdollahnejad, A., Surový, P. and Ku elka, K., 2019. Detection of fallen logs from high-resolution UAV Images. *New Zealand Journal of Forestry Science*, 49.

DOI: 10.33494/nzjfs492019x26x

QGIS Development Team (2020). QGIS Geographic Information System. *Open Source Geospatial Foundation Project*. http://qgis.osgeo.org

Sagan, V., Maimaitijiang, M., Sidike, P., Eblimit, K., Peterson, K.T., Hartling, S., Esposito, F., Khanal, K., Newcomb, M., Pauli, D., Ward, R., 2019. UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sensing*, 11(3): 330.
DOI: 10.3390/rs11030330

San Francisco Declaration on Research Assessment. 2012. https://sfdora.org/read/

Segarra, J., Buchaillot, M.L., Araus, J.L., Kefauver, S.C. 2020. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy*, 10(5): 641.
DOI: 10.3390/agronomy10050641

Thiel, C., Mueller, M.M., Epple, L., Thau, C., Hese, S., Voltersen, M., & Henkel, A., 2020. UAS Imagery-Based Mapping of Coarse Wood Debris in a Natural Deciduous Forest in Central Germany (Hainich National Park). *Remote Sensing*, 12(20): 3293.
DOI: 10.3390/rs12203293

Tmušić, G., Manfreda, S., Aasen H., James, M.R., Gonçalves, G., Ben-Dor, E., Brook, A., Polinova, M., Arranz, J.J., Mészáros, J., Zhuang, R., Johansen, K., Malbeteau, Y., de Lima, M.I.P., Davids, C., Herban, S., and McCabe, M.F. 2020. Current Practices in UAS-based Environmental Monitoring. *Remote Sens*ing, 12(6): 1001.
DOI: https://doi.org/10.3390/rs12061001

Tsouros, D.C., Bibi, S., Sarigiannidis, P.G. 2019. A review on UAV-based applications for precision agriculture. *Information*, 10(11): 349.

Tuominen, S., Balazs, A., Honkavaara, E., Pölönen, I., Saari, H., Hakala, T. and Viljanen, N., 2017. Hyperspectral UAV-imagery and photogrammetric canopy height model in estimating forest stand variables. *Silva Fennica*, 51(5).
Doi: 10.14214/sf.7721

U.S. Copyright Office. 2020. Compendium of U.S. Copyright Office Practices § 101 (3d ed.), Chapter 500 & Chapter 600. https://copyright.gov/comp3/

Viinikka, A., Hurskainen, P., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Mäyrä, J., Poikolainen, L., Vihervaara, P., Kumpula, T. 2020. Detecting European Aspen (Populus tremula L.) in Boreal Forests Using Airborne Hyperspectral and Airborne Laser Scanning Data. *Remote Sensing*, 12(16): 2610.
DOI: 10.3390/rs12162610

A Year Acquiring and Publishing Drone Aerial Images in Research on Agriculture, Forestry, and Private Urban Gardens

Olli Niemitalo, Eero Koskinen, Jari Hyväluoma, Outi Tahvonen, Esa Lientola, Henrik Lindberg, Olli Koskela, Iivari Kunttu

Viljanen, N., Honkavaara, E., Näsi, R., Hakala, T., Niemeläinen, O., & Kaivosoja, J. 2018. A Novel Machine Learning Method for Estimating Biomass of Grass Swards Using a Photogrammetric Canopy Height Model, Images and Vegetation Indices Captured by a Drone. *Agriculture*, 8(5): 70.
DOI: 10.3390/agriculture8050070

About the Authors

Olli Niemitalo currently works as a data scientist at HAMK. He has a passion in open research and in development of data management practices. Olli has a BSc in biochemistry from Oulu University and a professional background in digital signal processing.

Eero Koskinen is in the last year of his forestry studies at HAMK. Eero has worked in forestry his whole life. During his studies at Evo he got hooked on learning more about GIS. The GIS and drone courses at Evo enabled him to get an internship at the HAMK Evo campus, during which he gathered and processed drone data, including the image datasets presented in this article.

Jari Hyväluoma currently works as a Principal Research Scientist at Häme University of Applied Sciences. He received a PhD in Applied Physics from the University of Jyväskylä in 2006. His research interests include soil structure, soil hydraulic properties, and organic soil amendments.

Henrik Lindberg (M.Sc., Forestry) is a senior lecturer in HAMK whose field is forest ecology and silviculture. In his research activities, he has focused especially on nature management, forest biodiversity, and forest restoration.

Esa Lientola (Master of Natural Resources, Forestry) is a senior lecturer in forestry at HAMK, who specializes in remote sensing, forest planning, and GIS-applications of forestry. In recent years, he has concentrated particularly on developing the practical use of drones for the study of natural resources.

Olli Koskela is currently working as a research manager at Häme University of Applied Sciences with a data science team. His research areas include many bioeconomic processes, such as dairy production, feed quality management, and soil maintenance. He holds a Master of Science in applied mathematics from Helsinki University, and is currently finalizing his PhD thesis in the field of biomedical engineering at Tampere University.

Iivari Kunttu holds a PhD in Information Technology from the Tampere University of Technology (TUT, 2005), a PhD in Economics (Management) from the University of Vaasa, Finland (2017), and M.A. in Education from Tampere University, Finland (2020). Currently he acts as Leading Research Scientist in Häme University of Applied Sciences. From 2012 to 2017, he held an Assistant Professor position in the Department of Management at the University of Vaasa. He has also held several R&D Manager and R&D Process Development specialist positions in the Nokia Corporation, and Project Manager positions in TUT. His current research interests include R&D and innovation management, data analysis, business development, as well as digital services. His works have been published in such international journals as *Pattern Recognition Letters, Machine Vision Applications, Optical Engineering, Journal of Telemedicine and Telecare, Annals of Long-term Care, Technovation, Industry and Innovation*, and *Technology Innovation Management Review.*