Practical drone ecology: Data management to facilitate open, reproducible science in drone-based terrestrial ecology applications

Michael J. Koontz[1*], Victoria Scholl (?)[1], Anna Spiers (?)[1,2] Megan Cattau (?)[1,3], Joe McGlinchy (?)[1], Jennifer Balch[1]

[1]Earth Lab, University of Colorado-Boulder; Boulder, CO, USA

[2]Department of Ecology and Evolutionary Biology, University of Colorado-Boulder; Boulder, CO, USA

[3]Department of Human-Environment Systems, Boise State University, Boise, ID, USA

[*]*Correspondence:* `michael.koontz@colorado.edu`

Keywords: drone, sUAS, data levels, data management, open science

Date report generated: August 09, 2019

Abstract

"Small science" ecology labs are increasingly collecting "big data" from drone-based instruments [1]. There is currently a lack of a consensus framework on how to best manage these data from collection to publication. Here, we propose a guide for established "data levels" for ecological applications akin to those developed for NASA and USGS Earth observation data as well as practical suggestions for ensuring data privacy, data redundancy, and reproducible workflows using drone-derived data.

Introduction

For drone technology to revolutionize spatial ecology [2], it must move beyond the "peak of inflated expectations" on the Gartner Hype Cycle for Emerging Technologies [3].

Drones can be a means to democratize landscape and macroscale ecology due to their relative inexpensiveness, their ease of use, and their ability to meet the demanding primary need of these disciplines: fine-enough resolution data at a broad-enough spatial extent.

Thus, drone ecology is at a pivotal moment in its development wherein decisions made by current practitioners can determine whether the field will add to or subtract from the "research debt" of future practitioners [4].

We advocate an open science model by which new, potential practitioners are welcomed into the field with a clear-eyed view of how to best make use of the emerging technology in a practical sense.

In this brief **Communication**, we answer the call of [1] and propose a data organization and management scheme that will help guide users to making intentional decisions about drone-derived data to foster collaboration, data privacy, and reproducible science using data that may be many orders of magnitude larger than those typically collected by ecologists.

Data collection

- Multiple SD cards
- Multiple hard drives in the field (to ensure redundancy)
- Network attached storage
- Cloud backup
- Separate hard drive for local processing

Conceptual framework for varying levels of drone-derived data products

NASA developed the concept of data levels (https://earthdata.nasa.gov/collaborate/open-data-services-and-software/data-information-policy/data-levels) as a means of organizing the conceptual flow of a processing pipeline for Earth observation data (e.g., from satellites).

There are some parallels with drone-derived data, but some important differences that limit the ability to map the data typically collected for drone ecology onto these established levels [1].

We propose a scheme that fits the spirit of the NASA data levels but that more appropriately maps to typical Structure from Motion workflows for terrestrial ecology missions. This example is a plant science example (measuring and classifying trees), but we hope it has some capacity to generalize to other types of data collection for ecological studies.

Level 0

Raw data from sensors. Often images representing reflectance or emission of objects on the ground.

Level 1

Basic outputs from photogrammetric software.

Level 2

Radiometrically or geometrically corrected outputs from photogrammetric software. Some of these corrections can be done within the photogrammetry software itself. Some can be done outside the photogrammetry software using other software (e.g., `R` [5]).

Level 3

Domain-specific information extracted from Level 2 products.

Level 3a

Information extracted using **just** spectral information or **just** geometric information.

Level 3b

Information extracted that uses **both** spectral and geometric information.

Level 4

Aggregations of Level 3 products to regular grids. For instance, to match to other sensors or to account for the scale at which calibration was performed.

Figure 1

Discussion

Acknowledgements

We gratefully acknowledge funding from the USDA Forest Service Western Wildlands Environmental Threat Assessment Center (WWETAC) as well as the CU-Boulder Grand Challenge.

References

1. Wyngaard, J.; Barbieri, L.; Thomer, A.; Adams, J.; Sullivan, D.; Crosby, C.; Parr, C.; Klump, J.; Raj Shrestha, S.; Bell, T. Emergent Challenges for Science sUAS Data Management: Fairness through Community Engagement and Best Practices Development. *Remote Sensing* **2019**, *11*, 1797.

2. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment* **2013**, *11*, 138–146.

3. Lee, J.-G.; Kang, M. Geospatial Big Data: Challenges and Opportunities. *Big Data Research* **2015**, *2*, 74–81.

4. Olah, C.; Carter, S. Research Debt. *Distill* **2017**, *2*, e5.

5. R Core Team *R: A language and environment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018;

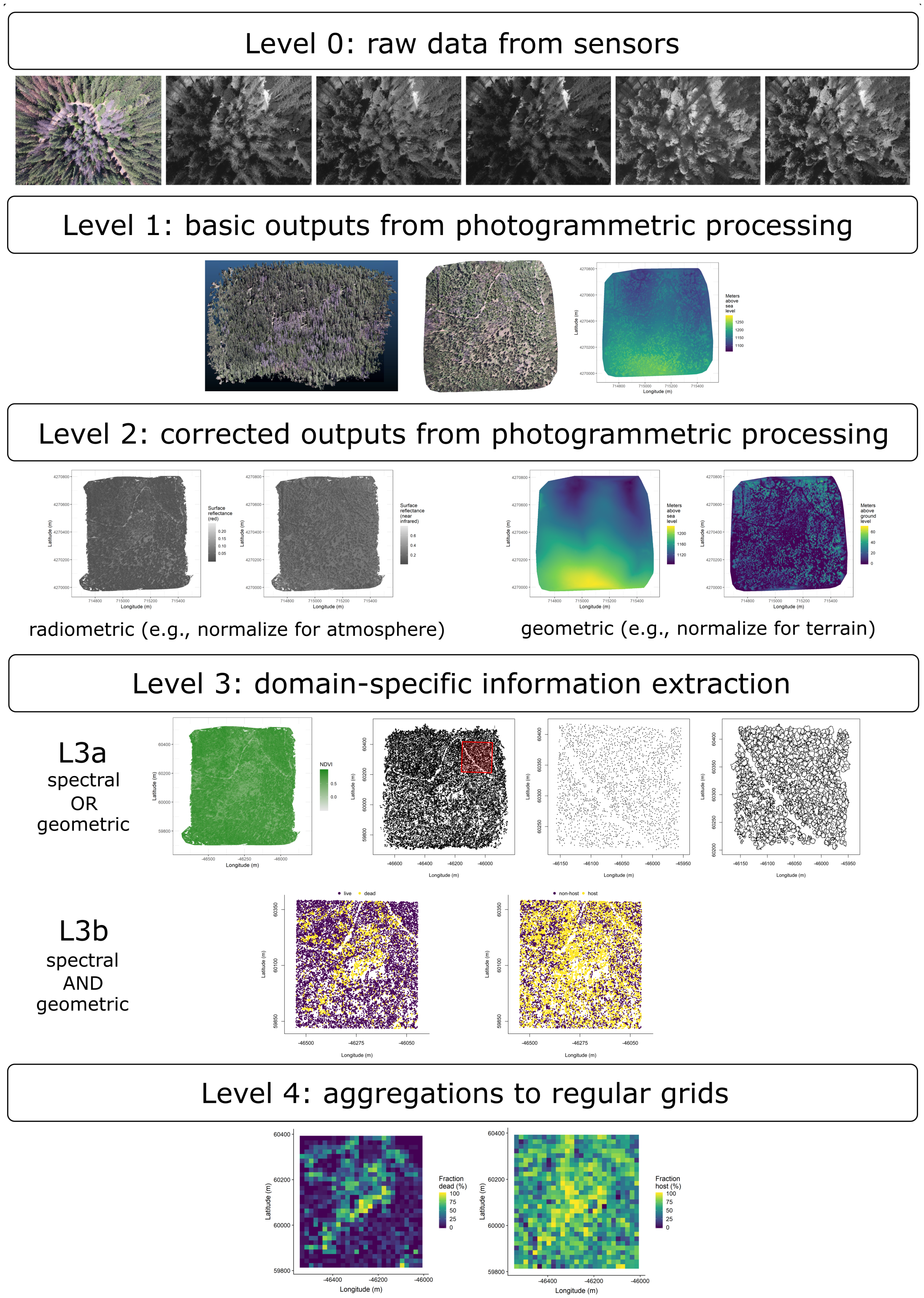

Figure 1: Proposed data levels for drone-derived data used in ecology applications.