Comment on essd-2021-222
Vitezslav Moudry (Referee)

Referee comment on "EcoDes-DK15: high-resolution ecological descriptors of vegetation and terrain derived from Denmark's national airborne laser scanning data set" by Jakob J. Assmann et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-222-RC1, 2021

First of all, I would like to congratulate the authors for their excellent work. I’ve been hoping for such an initiative for some time. You have done an excellent job. Please find below minor comments that could improve your dataset/manuscript.

1 Introduction

I suggest to either simplify the introduction (by omitting some technical information) or provide better explanation, as in its current form it could be misleading (see my specific comments below).

Lines 43-45: “Some surfaces may scatter large parts of the light which may result in multiple returns of different strengths, vegetation is amongst those (Wagner et al., 2006).” I am not a physicist (neither native speaker), but not sure with the terminology (wording) in this sentence. What do you exactly mean be “may scatter large parts of the light”? I would probably use reflect rather than scatter? In vegetated areas, the laser beams are usually reflected from several layers of vegetation. The interaction of the laser beam with the canopy is then characterized by multiple returns from several depths of the vegetation. The first return typically comes from the vegetation canopy surface and the second and possibly other returns follow (intermediate returns from leaves and branches), with the last one ideally being a return from the terrain. In other words, LiDAR pulses can penetrate through gaps in the vegetation canopies and register multiple returns representing both above-ground objects and the terrain.

Lines 49-50: What do you mean by ALS methodology. That is quite vague. Do you mean data acquisition, or point clouds processing?
Lines 50 - 52: Maybe add one or two sentences about data filtering/classification. So those unfamiliar with LiDAR point clouds get the idea about how the point clouds are classified/filtered. You refer to classification later in the text. You write: “Depending on the properties of the LiDAR sensor, different surfaces will reflect the light with different strengths, allowing, for example, to separate returns from vegetation from those of bare-ground”, but returns intensity is not the main characteristic that allows to separate vegetation from ground. This may give to an unexperienced user feeling that laser pulse is able to recognize building from vegetation, which is not true (see Sithole and Vosselman, 2004; Moudrý et al. 2020). Most of the state-of-the-art filtering techniques uses the relative positions of the points (similar situation is when distinguishing buildings from vegetation). Use of additional parameters, such as those tested in Wager et al. 2006 is to my knowledge rather experimental.

Sithole, G. Vosselman, Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds, ISPRS J. Photogramm. Remote Sens. 59 (2004) 85–101,

Moudrý, V., Klápště, P., Fogl, M., Gdulová, K., Barták, V., & Urban, R. (2020). Assessment of LiDAR ground filtering algorithms for determining ground surface of non-natural terrain overgrown with forest and steppe vegetation. Measurement, 150, 107047.

Lines 52 – 53: “While early applications for ALS were focused on generating simple digital elevation models (DEMs) for city and landscape planning, as well as canopy height estimates for commercial forestry,...”. I thing that the references used: Bakx et al., 2019; Vo et al., 2016 are not relevant for landscape planning or forestry. You might consider Fogl and Moudrý (2016) for city planning and Nilsson et al. (2017) for forestry (but in fact these are quite recent rather than early applications).

Fogl, Michal, and Vítězslav Moudrý. "Influence of vegetation canopies on solar potential in urban environments." Applied Geography 66 (2016): 73-80.

Nilsson, M.; Nordkvist, K.; Jonzén, J.; Lindgren, N.; Axensten, P.; Wallerman, J.; Egberth, M.; Larsson, S.; Nilsson, L.; Eriksson, J.; et al. A Nationwide Forest Attribute Map of Sweden Predicted Using Airborne Laser Scanning Data and Field Data from the National Forest Inventory. Remote Sens. Environ. 2017, 194, 447–454.

Line 54: Exactly which recent advances do you have in mind? Sure, the technology matured and it is now easier to acquire point clouds of high density, especially for large areas, but not sure if we were limited in calculation of complex measures in the past? Besides, after this information you continue with terrain derived measures of ecological interest, such as slope, but these are not complex measures and can be calculated using point clouds of comparatively low density (especially if you use 10m resolution).
Lines 58 – 61: Not sure here if having sensor and point cloud characteristics in one sentence makes sense. I would simplify it and use solely the point cloud characteristics as it will depend on the used sensor. Besides, it is not only sensor (wavelength, footprint, etc.) but also surface reflectivity what matters. Not sure, that the review by Bakx et al. (2019) is the best citation here.

2 Source data and processing workflow overview

Line 104: Nord-Larsen et al (2017) -> Nord-Larsen et al. (2017).

Line 106: S is missing; ETRS89 UTM32N; No EPSG code for DVR90? Is it EPSG: 5799?

Lines 113 - 116: Not sure I completely understand this? Does it mean that complementary data (i.e. terrain or point cloud) are not available at all or just that tiles are somehow shifted? If this is the case, it would make sense to resample/re-tile them. It is 291 km squares which might be missing for potential users (0.7% of Denmark).

Lines 142 - 143: I suggest to write datasets or inputs instead of variables, which you use for the characteristics derived from the point cloud and DTM. Actually, you use term descriptors in the Title, Introduction, Conclusions, and some figures and tables while in chapters 2, 3, 4, and 5 you use term variables.

3 Data set description and known limitations

Line 149: S is missing; ETRS89 UTM32N

3.4. Elevation-model derived variables

Line 194: Kopecký et al (2020). -> Kopecký et al. (2020).

Aspect and Slope: Equally relevant would be to derive slope at 0.4 m grain size and than average it into the 10 m grain (although more computationally demanding). This way actually better preserves the original values at finest resolution, which might be important for some users. Maybe you want to mention it (see Grohmann, 2015 and Moudrý et al.
Similarly as you mention limitations of TWI, Grohmann, C. H. (2015). Effects of spatial resolution on slope and aspect derivation for regional-scale analysis. Computers & Geosciences, 77, 111-117.

Moudrý, V., Lecours, V., Malavasi, M., Misiuk, B., Gábor, L., Gdulová, K., Šimová, P. and Wild, J., 2019. Potential pitfalls in rescaling digital terrain model-derived attributes for ecological studies. Ecological Informatics, 54, p.100987.

Solar Radiation: Is it reasonable to calculate solar radiation in MJ/cm²/yr when you are using 10m resolution? You do not even have centimeter resolution in original DTM. I mean “average ecologist” can easily overlook this fact and considerably underestimate the solar potential in individual cells. Besides, if you calculate the solar potential as MJ/yr per cell, you will not have to use the conversion factor of 1000. In addition, I had to look to original text by (McCune and Keon, 2002) to realize that “ln” is natural logarithm (first I thought that it is a unit I am not familiar with) and that the data are not in arithmetic scale. So, if I am correct, to get the actual solar radiation in individual cell I have to multiply by 10 000 (cm -> 10m), divide by 1000 (conversion factor) and convert to arithmetic scale (if needed), right?

Besides, I strongly suggest that in a next version of this dataset you calculate solar radiation monthly and to use some better model for estimation (e.g. r.sun implemented in GRASS; Šúri & Hofierka, 2004). The model you used does not account for cloud cover, so you assume clear sky all the year. And it does not account to shading due to adjacent topography, which is a shame, considering how accurate terrain data you are using (although I realize that Denmark is relatively flat). Consequently, the first sentence of the paragraph “Incident solar radiation is a key parameter for plant growth and indicator for local microclimate.” is in my opinion irrelevant for your dataset as you do not consider local (microclimate) terrain conditions in your calculations. Maybe you should mention this in the text (or documentation). I would not count that users of your data will look at the original article you cite.

Šúri, M., & Hofierka, J. (2004). A new GIS-based solar radiation model and its application to photovoltaic assessments. Transactions in GIS, 8(2), 175-190.

Line 286: twi (also in Table 2); Line 287: TWI; be consistent.

3.5 Point-cloud derived variables

Lines 435 - 436: The way I understand the term „texture”, it is just other way (than
vegetation density) to describe vegetation structure. So why should I worry about it, it is a property that this variable (vegetation density) should describe. Or did I misunderstand what you mean? Besides, do you have any reference that would support your statement that different canopies scatter light differently, causing lower or higher number of returns? Is it that significant difference that it is worthy to mention?

**Lines 449-460:** Building proportion: Based on some random tile (which included only small buildings) I guess that what is classified as buildings are typically returns from the roofs (or are there also vertical walls represented)? Not sure, but maybe this should be also mentioned.