Review ESSD-2020-201

Interesting product. Night light (NTL) data used extensively in our science, either from DMSP (annual, only up to 2013) or NPP (authors correctly label this product as NPP-VIIRS): monthly (with some gaps) at higher spatial resolution but available only since 2012. Authors apply an image processing tool - auto-encoder - and several learning / comparison steps to produce a merged annual time series 2000 to 2018. Crossing the sensor divide, e.g. digital number (DN) from DMSP vs radiance values from NPP has proved a difficult barrier up to this point. These authors demonstrate a statistical image comparison tool to cross this gap, significantly different to prior approaches. Some users will prefer higher-res NPP products but for many purposes a verified homogenized longer time span could prove very useful combination. Taking this as an image processing challenge, and developing image-based statistical tools to bridge the DMSP / NPP gap, represents an important and useful approach! Authors have a prepared a good description but one that could benefit substantially from many clarifications and improvements.

Data accessible from a DataVerse site, as 19 separate annual images. My Mac stumbles a bit on the format designator (.tif) vs .tiff but I can open NTL images with GIS and image processing software. Compressed file = 50 MB, single annual image = 9 GB! Authors might add description of file structure and warning about image sizes in the Data Availability section. Authors might consider also making data for one of their figures (e.g. Figure 9 for this reviewer but editor might have different suggestion) available to save users from needing to do our own mass download and compilation?

Overall, well written with reasonable assessments of uncertainty (expressed as correlation coefficients) of various intermediate products and of final composite. Validation / verification mostly consists of comparison at various steps of modified images vs originals. Really not much else one can do with NTL; no comparable alternate sources. (Authors define and use better terms - extended, composited, official - but sometimes as verbs/actions and other times as descriptors/adverbs? Quick search and check should resolve any language uncertainties.) However, this statement (lines 363 and 364 in Limitations section): "For the further research, there are still concerns about the time-series of NTL data with better quality and longer time span." Paper cited in the sentence before (Li et al. 2019) focused only on NPP data, so how does their work impact this product? What - in terms of uncertainty - does that sentence portend? Better calibrations? Longer time-series? Corrections based on viewing angles would give better correlation coefficients? A future version that renders the present version obsolete? Please provide explanation and assurance. I suspect authors mean something like: further work e.g. on viewing angles might result in (slight?, substantial?) improvements in correlations. As written however the sentence leaves a reader / user wondering about product as described? A so-called 'harmonized DMSP' NTL data product already exists (Scientific Data, https://www.nature.com/articles/s41597-020-0510-y), discussed by these authors in Introduction at lines 66 to 73) but this reader, while impressed with this approach and this product, never found a quantitative statement of improvement over prior work?

Section 2, Data

Please provide exact data sources. Even if data came from Google Earth Engine (referenced earlier) authors must provide exact DOI, version number, date of access, etc. A data source table would work very well here. Something like Table 1 but with much more detail. Authors do a good job of describing (and labeling) their source products and their own subsequent modifications but user needs to see exact starting points. Why Colorado School of Mines? Whatever the source, assure open accessibility and provide exact details.

Did authors use Google Earth Engine to access data or as a computing service?

Line 145: Some prior reason for selecting those three cities?

Line 165: Figure 2 does not represent full processing scheme which includes access, pre-processing, AE CNN training, and post-processing. Figure 2 includes post-processing but omits substantial necessary pre-processing? Starts instead at AE step.

Line 190, Figure 4: informative perhaps as description of an image processing tool but less helpful in a data description of new homogenized NTL product? Authors / editors need to distinguish between focus on AE CNN tool and description of useful NTL data product.

Line 207: "Adam"?? Reader does not learn until line 228 that 'Adam' refers to an already published processing algorithm.

Line 231: 'epoch' as used here refers not to geologic time period but to an iteration quanta of the AE deep learning tool? At some point the loss term approaches some statistical noise lower limit? Authors mention (line 233) 4000 as their empirically-chosen iteration limit but this also has implications for computing resources? Compressed vertical scale in Fig 5 not helpful here.

Line 234, Figure 5: Same question as for line 190 above: does this relate to new image processing tool or to higher quality NTL product?

Line 235 and following, section 4.2 'Accuracy evaluation'. Confusion here about numbers and Figure 6. Fig 6a has ordinate and abscissa range of 300, suggesting that an individual pixel can have an intensity value up to e.g. 250 or 300? From text, Fig 6a should include 100k points? Perhaps give us an n(umber) value? Fig 6b should represent 400000 cities with max NTL intensity values of 600k (e.g. a range of 0 to 600k)? Integrated above some minimum intensity value over a NTL-determined or geopolitically-defined city area? The density scale (population density?) on the right refers only to 6b?

Line 253 Figure 7: same units as for Fig 6 but now for country. Authors discuss correlations (agree, good in all cases) but changing axis extents/scales indicate USA comparable to Brazil, both greater than South Africa which is greater (brighter) than China and Australia both of which brighter than Italy. Confirm from other evidence? Or, if we should not indulge in quick visual speculations, use a uniform axis scale on all panels. Again, knowing n(umber) could prove very useful.

Line 262 Figure 8: What city or cities. Same 400k cities used in Fig 6b? DN vs nanoWatt intensities so the axis scales have changed, correlation coefficients as before but now the solid red lines indicate data slopes not 1:1 lines. What do we make of the changing slopes. No mention from authors. We need n(umber) indication! 2004 tends to show outlying data points below the slope line, e.g. high intensities for low DN. 2006 and 2010 show opposite pattern: higher DN for lower intensities. Useful? Not useful? Perhaps slope lines have no value outside of specific chart but the apparent temporal variance of those lines 2000 to 2010 does not accord with Figure 10?

Line 272 Figure 9: Authors justify use of median, but from where do these data come? Text says "based on 5000 random validation points." Random pixels? Apparently so, with axis scales at 0 to 150. All with data greater than 0, e.g. density of 'low' or greater? Density axis (again, population density?) refers in this case to city, not pixel? What do I not understand here?

Line 278: the term "downtown" conveys neither geographic nor NTL density accuracy or specificity.

Figure 10 very useful, lots of information. Text discussion of Fig 10 implies greatest NTL intensity change occurs globally over 2010 to 2015? Highest longitudinal peak might occur in Middle East (not = Europe), not mentioned in text?

Line 316, Figure 13: Potentially useful figure but quite confusing. ID units relative, not km? No explanation for the circled discrepancy in 13c? Values represent averages (means, medians?) along track at some spatial resolution. Or pixel by pixel? Y-axis scales change with every city; use a common scale to allow city-by-city comparisons?

Line 319: "nearby the junction point" what does this mean? End of simulated and start of composited?

Line 323: world bank population data. We need a version number, DOI, accurate references, assurance of open access, … This data product should be referenced in improved Table 1, above?

Line 328 Figure 14. Ln:Ln plot gives little to no confidence. Linear trend of log-transformed data, is that even valid? Use an inset to show global data in separate non-log figure? Once reader gains confidence with global numbers, regional numbers will make more sense. Or separate into two panels, one for global and one for 7 countries. If this represents first use of population data, then what did density scales in Figs 6 and 9 represent? Some spatial clustering of high intensity pixels? What?

Line 330: here a reader learns about "temporal junction point" and can use Figure 15 to understand. Need similar clarity at line 319.

Lines 334 to 336: Do the time-dependent changes discussed here, strong from 2010 to 2014 but diminished from 2014 to 2018, accord with what the authors concluded earlier, e.g. about changes globally and regionally from 2010 to 2015 as in Fig 10?

Line 346 Figure 15. Same six countries used in Fig 7? Why these six? Y-axis scale varies in both minimum and maximum in almost every plot. Ln values? Not helpful! Again, one scale for global and a second uniform scale for all countries? These apparent offsets per country with time explain the different slopes in Figure 7?

Small language issues throughout (e.g. line 49 comparable not comparability, "worthy to note" rather than noteworthy in several places, etc.); one hopes / assumes Copernicus pre-publication efforts will clean these up?