We thank reviewer 2 for the time taken in reviewing our paper and for the very helpful comments, responses to which we detail below. The suggestions offered by R2 have helped us illustrate the main concepts behind the paper, improved the systematic methods of detecting signal shift, and sharpened our illustration of the method we propose for correcting the issue. We offer our responses to each of R2's comments in detail for the remainder of this letter.

I am reasonably confident that the authors of the present manuscript have correctly identified a change in spatial registration between the two image composites that they have analysed, although their method of doing so is (as they admit) somewhat crude. What remains of concern is the fact that they do not clearly enough explain how they choose their control points, nor discuss the fact that the shift between dates is smaller than the spatial resolution of the image data. Their observations that spatial errors have the ability to confound inferences drawn from geographical units (the third-level administrative divisions) that are not large compared with the size of the errors is undoubtedly correct. However, I would like to see some discussion of the match between the spatial scale of the Nawahi in the most densely populated areas and the spatial resolution of the OLS data – in other words, whether there was any real likelihood that the DJB analysis could have produced meaningful results even if the spatial shift had been corrected.

We thank R2 for bringing each of these points to our attention. We address most of the points in the comments below. For nawahi, we realize that we could have done a much better job of communicating their relative size to the individual pixels and pixel clusters. We add footnote 7 to illustrate the variability of the size of nawahi (the smallest is only 6 square km, which would be subject to spatial correlation), the largest is over 7000 square km in the Syrian desert.

In the case of smaller nawahi, their relatively small size makes it more difficult to pick up the variation in regression models without being swamped with spatial correlation. We briefly discuss this in the modified footnote 8 (formerly footnote 3) and in more detail in appendix 6 (formerly appendix 4). The assumption in regression models of independence of observations can be relaxed by clustering standard errors at a larger unit of spatial aggregation than the unit of analysis. Somewhat puzzlingly, DJB say they do this in their paper at some points, but not in others and most (but not all) the regression specifications in their replication materials do not cluster standard errors. Without clustering standard errors, spatial correlation is unaccounted for. With clustering, it is assumed that units within a specific level of aggregation are correlated -- it is not as precise as a spatial lag, but nonetheless sufficient for assuming a large amount of spatial autocorrelation if clustering is at a high enough level. In the article, we cluster at the governorate level in our model -- a high level of aggregation, as there are relatively few (13) governorates. DJB's results replicate (see model 1 in what is now Table 3 [formerly Table 2]). This suggests that their results hold up even when assuming a fairly high amount of spatial correlation across nighttime lights data in all of Syria. The issue is primarily with the shift, as that is what ultimately renders their results insignificant in both Table 3 (formerly Table 2) and in the models that do not use clustered standard errors in appendix 6 (formerly appendix 4).

32. The years 2007 and 2009 were chosen for the present study because these were also used in the work of De Juan and Blank (DJB) and the authors here wish particularly to show that the conclusions of the DJB study were incorrectly drawn from the data. This is valid, but the generality of the results presented here would be increased by analysing other years too. Would it be hard to add a few more? Also, the authors don't say whether they used the F15 or F16 data from 2007.

We thank R2 for the suggestion. We bridge this suggestion with the later suggestion regarding spatial cross-correlation to examine correlations between annual composites of nighttime lights observations from all of the F16 satellite (Table 1 in the new version of the paper and lines 63 through 66 as well as the new footnote 5). By evaluating interannual correlations of the full F16 satellite, we can now show that the 2009 map was anomalous with respect to not only 2007, but most of the remaining years. Our wider analysis in appendix 1 suggests there is a broader pattern of the first and last years of a DMSP-OLS satellite's coverage being less correlated to the remaining years. This is also noted in footnote 5.

We also add appendix 2, which replicates Figure 1 in the manuscript for other years captured by the F16 satellite and 2009. This allows us to more definitively see whether the shift from 2007 to 2009 is unique to those years or extends to other composites captured by the F16 satellite. We observe a similar shift from 2005, 2006 and 2008 to 2009. Our control points, generated from 2007 and 2009 observations, have similar effects in correcting the differences from 2005 to 2009, 2006 to 2009 and 2008 to 2009. We do not observe a difference between 2004 and 2009 data, further suggesting that the first and last composites from each satellite are different from those in the middle -- observed in correlations in appendix 1 in the F12, F14 and F15 satellites. We add a summary of the findings from appendix 2 to lines 97 -- 101.

We also now make it clear, throughout the paper, and on line 63 and the caption of Figure 1, in particular, that we are looking at changes within the F16 satellite when looking at the shift from 2007 to 2009.

67. Isn't the first step to obtain the OLS images? Where were they obtained from? Please give details. I downloaded the dataset F162009 (the only one calculated for 2009, as far as I am aware) from
https://eogdata.mines.edu/dmsp/downloadV4composites.html
to check the coordinates.

We now include a link to the NOAA website (largely identical to what R2 links to) in footnote 3.

74. 'suggesting the presence of a skew as a result of signal decay'. I do not think there is any evidence for a skew in the data, in the geometrical sense, just a shift. And I don't think the authors have presented any evidence for the shift being due to signal decay.

We apologize for the mistake in wording here. We have changed the passage, still on line 75, to: "further suggesting a shift." We did not mean to imply there was bias in the numerical measure of DNs.

Fig. 1 What are the red lines and the grey contours? What is the scale of these images?

We now add a scale to both Figure 1 and the newly added figure 2. We also explain the red (now green) lines in Figure 1 on lines 69 - 71. They are intended to better highlight shifted urban areas across the three maps on the top panel of figure 1. The grey lines are built-up areas (glossed on line 73 and footnote 6).

77 Manual technique. Although the consistency of the results in table 1 do suggest that the manual method is satisfactory, I have some reservations about it and would like to know more details about it

before I can agree that it is sufficiently rigorous. (1) It isn't clear how the control points' specific pixels, which have to do duty as spatial invariants, have been selected. For example, I examined the pixel in the 2009 image corresponding to the Tartous control point (last in table 1). Assuming I hit the right pixel, it has a DN of 22, which is not higher than all of its eight immediate neighbours, so since it is not a local maximum I am not clear how you could be sure to pick the same pixel in the image from another year. (2) Since the pixels in the image are 30 arcseconds ≈ 0.008 degrees in size, it is not immediately obvious that the coordinates of a pixel can be specified by manual interrogation to a precision of 0.001 degrees. That would require being able to differentiate between around 70 different positions within one pixel. (3) The actual spatial resolution of the OLS radiometer is at best (near nadir) 2.2 km, i.e. around 0.02 degrees, so an area of roughly 4 x 4 pixels is strongly correlated. In the light of these concerns, I would be much more convinced by a method of detecting the shift between image dates that used spatial cross-correlation or similar.

Replying to point 1: We now make it clear on lines 70 - 72 that the selection process focuses on pixels that are on the borders of cities and towns so that sharp distinctions in pixel value can be detected. We also make clear that rather than selecting individual pixels (as values can change over the course of one or two years), we choose intersections or borders between two or four pixels. The selections for two of the control points are illustrated on Figure 2 around the city of al-Qamishli.

Broadly, we believe that focusing on borders allows for more precision in terms of placement across the years. If the borders shift as well as the pixels, this becomes substantively relevant for our correction, rather than a product of error.

Replying to point 2: Our response here is similar to the first point, in that our pixel choice is usually the border between two pixels or the intersection between four. If the shift is across pixel values and not pixel position, then the shift should be in defined intervals. Any substantive shift in the position of the pixels themselves can then be incorporated into substantive shifts detected by control points as an intersection that does not appear in the same place would be detected by our control point placement.

Replying to point 3: We appreciate this suggestion as it has added an important component to the process of detecting signal shift that was previously missing. We now add raster correlations in both the article itself (new Table 1 and lines 63 - 66 for the F16 satellite) and the new appendix 1 (for the remaining satellites). The raster correlations that are utilized are moving window correlations of 5x5 pixel quadrants which should allow a capture beyond the 2.2km pixel quadrants that are minimally spatially autocorrelated. The correlations allow us to gain a rough idea of which composites within the same satellite differ from one another. While we do not include these in the paper, 7x7 and 9x9 windows show a similar order of correlated images from table 1 as 5x5, albeit as windows get larger, correlations get closer to 1 and it becomes more difficult to detect variation. We do not feel that we need to go past the moving window method to more advanced methods (i.e. Chen 2015). Chen (2015) in particular generates spatial lags to account for spatial autocorrelation between cities in China of varying distance to one another. The nighttime lights data itself does not necessarily require this sort of precision because of the limited window (up to 4.3 km according to Elvidge et. al. (1997)) in which the cells are autocorrelated.

Moreover, spatial cross-correlation is limited in that correlations could rise and fall due to actual changes in light emissions. Thus visual identification remains an important component in detecting signal shift. We now make this clear throughout section 2 of the article, with both the correlation of intrasatellite data and visual inspection as justifications for employing the control point method.

As we mention at the end of our article, we acknowledge the manual technique can be improved through an automated process. We clarify in lines 177 - 183 that an automated process would be difficult because it would need to detect systemic shifts in maps that are caused by signal shift, and not by substantive changes, and then correct for those shifts. We hope that our introduction of the manual process is a starting point for further automated developments. However, as with all automated processes, qualitative and manual checks are an important validation component and we believe that, even if automated processes are developed in the future, our technique will have value in validating those processes.

Table 1. Suggest replacing ‘X coordinate’ and ‘Y coordinate’ with ‘longitude’ and ‘latitude’ respectively, and replacing ‘aggregate’ with ‘mean’. The table could be more economically laid out, with just one line per place and still seven columns per line (place, 2007 longitude, 2009 longitude, 2007 latitude, 2009 latitude, long difference, lat difference). Would be good also to add the standard deviations of these differences, to be sure that they are in fact significant. Also it would be good to establish that the shifts are not spatially correlated (e.g. shifts predominantly in one direction in one part of Syria, another in a different part. My quick check, just by correlating shifts against latitude and longitude, was reasssuring. But it would be good to hear from the authors on this point.) By my calculations, the shifts are 0.0044 (0.0012) in longitude, and 0.0160 (0.0043) in latitude, both of which are significantly different from zero ($p<0.005$ for both).

We have taken the suggestions to revise Table 1 (now Table 2) extensively as R2 prescribes here. Each column has been rearranged as requested and we have changed the names as suggested. We have also added a line below the mean with standard deviations and a line below the standard deviations with T-values (the p-values were very small and we do not believe they would be as informative if displayed to three significant figures).

Regarding spatial correlation, we reiterate some of our responses to point (3) in the previous comment. The moving window correlation that we now employ exceeds the 4x4 quadrant where nighttime light pixels are likely to be highly spatially correlated. The technique shows a divergence of the 2009 (and 2004) composites from the 2005-08 composites. We are otherwise in agreement with the reviewer regarding correlations of the shifts themselves. In placing control points, we attempted to choose points to the south, north, east, and west of cities so that we were not capturing natural growth or loss or something of the sort. We include a note about the correlation of the shifts as part of Table 2 (formerly Table 1).

Table 2. Lacks enough detail to be understood without having the DJB paper in front of me. What are the numbers in the table?

We thank Reviewer 2 for this comment, the previous title was not sufficiently clear. We have modified the title of the table (now Table 3) to indicate the type of regression that is being

run (Logistic) and the dependent variable. We have also added to the note section in the table below to clarify which of the models (model 1) was the original run by DJB.

110 ‘their attempt to account for gas flares falls short of being a reliable technique.’ Please elaborate on this remark in the text.

We have modified this paragraph and added footnote 9 to clarify the technique used by De Juan and Bank (to remove all observations from any governorate that had gas flares -- 83 total subdistricts) could have been improved on by drawing on data from Elvidge et. al. (2009) and removing only the gas flares themselves and a 10km buffer around the flares. We go into detail about this process in appendix 4, where we also conduct further statistical analysis using the modified technique.

122-123. Equation (1) needs an ‘=’.

We have added a left-hand-side to the equation to rectify this.

<u>Citations</u>

Chen, Yanguang. "A new methodology of spatial cross-correlation analysis." PloS one 10, no. 5 (2015): e0126158.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, and Ethan R. Davis. "Mapping city lights with nighttime data from the DMSP Operational Linescan System." Photogrammetric Engineering and Remote Sensing 63, no. 6 (1997): 727-734.

Elvidge, Christopher D., Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. "A fifteen year record of global natural gas flaring derived from satellite data." Energies 2, no. 3 (2009): 595-622.