This is a nice piece of work, and certainly suitable for publication after revision. Calibration of infrared Doppler lidars is certainly a challenge, the reasons for which are well articulated in the text. Rather than make use of the returns from a calibration target at an extended range, which is operationally difficult, the work here describes incorporating measurements from lidars operating in the visible region, where calibration is easier. Two additional data sources are incorporated: a Raman lidar to provide measurements of backscatter and extinction at 532 nm, and a sun photometer used to estimate the relationship between extinction at 532 nm and 2022 nm. The work is well-presented, with the results applicable for situations where all three instruments are present and where conditions are such that assumptions on temporal and spatial homogeneity can be applied.

The weakest part of the technique, in my opinion, is the computation of the ratio between the 532 and 2022 nm extinction coefficients using a single path-integrated value from the sun photometer. As noted in the text, and certainly for this case, the aerosol number density and composition varies over the total path observed by the sun photometer. Trying to find cases where one can separate the different regions or layers unambiguously will likely be quite difficult. Here the authors do the best they can, but the estimating the k values for the upper two layers is very much difficult because they are rarely present individually.

I might recommend a slight change in title. When I started to read the article I expected the focus to be on retrieval of backscatter and extinction coefficients in the infrared where the Doppler lidar operates. I have no problem with the focus on the visible – that's where satellite and many ground-based lidar operate, so augmenting that data set with airborne Doppler measurements is certainly worthwhile.

A few minor points:

Page 1937, line 29: Coherent detection is not inherently insensitive to spectrally broad signals – rather coherent Doppler lidars are designed to have a narrow bandwidth matched to spectrally narrow aerosol signals to increase sensitivity for that regime. One could design a coherent lidar for signals of any bandwidth.

Page 1938:, line 11: It seems to me that the concept of overlap is more generally applied to direct detection lidars, where the design of the system typically isn't diffraction limited. A more appropriate terminology might be "antenna efficiency", which more appropriately describes the loss of sensitivity, relative to ideal, resulting from backscattered signal/BPLO coherence mismatches at ranges in the near field of the telescope. This is the nomenclature used by Henderson et al, 2005. I note that later on in the text the authors use "heterodyne efficiency" as reflecting phase and amplitude mismatches between the signal and LO. In this nomenclature heterodyne efficiency would include antenna efficiency plus other effects (e.g., LO beam heterogeneities).

Page 1944, line 5: The sentence on subtracting the noise floor calls to mind that there is no mention of uncertainties resulting from the estimation process. Coherent lidar returns are characterized by speckle; with LO shot noise adding an additional noise term Although the high prf of the system probably minimizes the speckle noise term, a paragraph or two discussing uncertainties in estimation of power from coherent lidar returns, and why this is or is not a factor in these measurements, would be helpful, I think.

Page 1948, line 10: Again, I think overlap is better described as antenna efficiency.

Page 1952, line 18: As I understand it, the Raman returns are used as an HSRL to estimate extinction and backscatter. Perhaps some discussion on the effect of using 607 nm to characterize optical properties at 532 nm would be appropriate here.

Page 1949, Eq. 17: Is there an extra R term in this equation?

Page 1954, lines 15-30: The text states, without much substantiation, that " the contribution of the marine aerosol to the measured total AOD is lower than the contribution of the mixed layer" and later, that contribution of the dust layer to the total AOD is much larger than the contribution of the other two layers". Since this inability to separate the layers for this calculation is an inherent weakness of the technique, I think these statements require more discussion and substantiation.

Page 1955, line 22: It isn't clear to me why S is constant over each layer. Is there some averaging or filtering that I'm missing?

Page 1956, line 6: The bulge in the layers in Figure 13 at 00:10 is interesting – looks like some upward transport of the whole layer caused by convection. Maybe some comment here would be interesting.

Page 5.5, line 20: If the mixed layer was a combination of a marine and Saharan dust layer, one might expect the histogram to fall between that of these two layers, but in fact it is more skewed to the right than either of the other layers. Any thoughts on this?