We thank the reviewer for the helpful comments and suggestions. In the our response to the reviewer comments, the Reviewer Comment is first reproduced in black, followed by our response in blue, and changes to the manuscript in *orange*.

General

This work uses lidar observations based on VAD and VADoe methods and compare the results with radiosonde observations. VAD technique is well known for many years. Optimal estimation (OE) is a new technique for wind analysis and claimed to be better than VAD alone. VADoe advantage is to have full covariance analysis at each level improve the wind profiles. It is claim that OE technique provide wind data at the levels where classical VAD not but uncertainty is very high. It is suggested that new method is better than old one and provided data without having any instrument hardware changes.

Major issues

Overall, what is claimed to be can be useful but with large uncertainty; this means the results can have very large uncertainty at the higher levels. If you compare the figs 5 and 6, as well as 4, you can see what is going at higher levels.

We agree with the reviewer's assessment that the VADoe results have large uncertainty at higher levels where traditional VAD do not provide any results. However, we disagree with the reviewer's assessment that higher uncertainty results are inherently not useful. VADoe provides the mostly likely wind conditions at those levels based on the (noisy) observations at that height winds at lower levels, where CDL makes accurate measurements, and climatology. VADoe retrieval provides uncertainty estimates for each output, and users can decide the uncertainty threshold based on their application requirement. This is an improvement compared to traditional VAD which do not provide any output for SNR below a certain threshold. Figure 7c and 8c clearly shows the benefit of VADoe as some very accurate measurements are discarded by VADtrad due to this hard SNR cutoff. In addition, VADoe more easily facilitates assimilation into NWP as it provides uncertainties, and averaging kernels needed for data assimilation.

Ln 29; Gultepe et al 2018 A review on aviation meteorology…. PAAG) can be provided here.

This reference has been added to the paper.

Ln37/60; seems these parags should be given under the method section. Intro is very short if these parags are taken to another section. Intro should be developed into a better summary of earlier obs/issues.

We respectfully disagree with the reviewer on moving this section to the method section. This section describes the CDL observation and associated issues, why the measurements are limited to lowest 1-2 km, and a need for a new method that could provide consistent vertical coverage. Thus, this section is better suited in the introduction.

Fig. 4; how can we say the retrieval (middle) provided better results compared to radiosonde; I see the strikes in the retrievals. VAD shows nothing. How can say that retrievals are better?

We do not claim that the VADoe retrieval provides better results compared to radiosondes. We use radiosondes as a reference to determine how well the VADoe performs at different levels. We also do not claim that VADoe provides better results than the traditional VAD method. What we do claim is that VADoe provides equivalent results as VADtrad where VADtrad provides an output. Where the VADtrad does not provide results, VADoe provides the statistically most likely results based on the (noisy) observations at that height, the higher-signal measurements at lower levels, and climatology. Thus, there are only benefits to using VADoe.

Fig 5; Compare VAD versus OE; OE shows much larger scatter of the data points compared to VAD ones. Then how can we say OE led to better results compared to VAD?

We respectfully disagree with the reviewer's assessment that VADoe shows much larger scatter compared to VADtrad. Figure 5c and 5f compares VADtrad and VADoe for u and v vectors respectively and the scatter in data is comparable. The larger standard deviation seen in Figure 6a-e is due to the larger number of data points included in the comparison. Please see Figure 6f for number of data points included in the comparison at each level.

We have added standard deviation information to Figure 5 depicting the scatter in data points.

Fig. 6; OE shows better results (a) but with large uncertainties (Fig. 5). Sd for both methods are bad at higher levels any way. Again issues exist at higher levels and OE technique provides bad results (higher error) but VAD provides no results. Based on this can we use OE results accurately? Probably not. Say if wind speed 2 m s-1 and error is 4 m s-1, then why we have to use this data?

We would characterize OE results at higher levels as results with higher uncertainty rather than "bad results". The beauty of the OE retrieval is that it provides propagated error for each result based on measurements and a priori error covariances. It also provides the Averaging Kernel matrix which provides information about vertical resolution as a function of height as well as the maximum height to which the retrieval is mostly independent of the a priori profile. So, users can decide what they want to consider as good or bad data depending upon application. This is currently not an option with traditional VAD retrieval.

Fig 5 and 6; how many data points are used in the analysis and how data is averaged? Please provide some info in the captions of Fig5 and 6.

We have added number of data points used in Fig 5 in the figure caption (N = 59,403 for trad vs sonde, and N = 139,582 for oe vs sonde). Radiosonde data were vertically averaged to lidar vertical grid and temporally matched to lidar (within 30 mins) for this analysis. This information is already included in the paper (lines 212-214).

Figure 6f shows number of data points at each level included in the intercomparison.

Figure 5: Scatter plots of the u component (top row) and v component (bottom row) of wind for radiosonde vs. VADtrad (left column, N= 59,403), radiosonde vs. VADoe (center column, N=139,582), and VADtrad vs. VADoe (right column).

Please do a Discussion section before the Conclusions. Then discuss the issues mentioned above.

We have added a discussion section as suggested.

The comparisons in the previous sections show that VADoe provides identical results as VADtrad where VADtrad results are valid. At these levels, where most if not all of the information are coming from measurements, VADoe is mathematically equivalent to VADtrad. At lower SNR levels (or higher altitudes), where VADtrad results are not available, VADoe results compare favorably with radiosonde measurements. VADoe at those levels are statistically most likely output based on the (noisy) observations at those levels, higher quality (precision) measurements at lower levels and climatology. The VADoe retrieval provides well characterized uncertainty for each profile, and the corresponding averaging kernels allow the determination of both the vertical resolutions as a function of height and the maximum height to which the retrieval is mostly independent of the a priori profile. Thus, the retrieval errors and averaging kernels could be used to determine data that are suitable for a given application.

One of the biggest challenges of setting up the VADoe retrieval is appropriately scaling the CDL radial velocity measurement error at low SNRs to provide stable retrievals. The CDL radial velocity measurement error is limited by the measurement bandwidth. For example the measurement bandwidth for the ARM SGP CDL used here is +/- 19 m/s. This maximum measurement error is smaller than the a priori error (standard deviation). This becomes even smaller when you consider multiple radial velocities from different azimuths that are included in the retrieval. If the measurement errors are not inflated appropriately, measurements will always be weighted heavily compared to a priori, and results in unstable retrievals. Thus, measurement error at low SNR levels needs to be appropriately scaled accounting for number of azimuths and elevation angles included in the retrieval and magnitude of the a priori error.

Successful implementation of VADoe retrieval requires knowledge of the a priori mean profile and covariance. We used radiosonde measurements to create monthly mean profile and covariance. However, radiosonde measurement sites are limited which limits the applicability of the VADoe retrieval presented here to locations close of radiosonde sites. Future work should include testing using a priori from other sources such as AMDAR and NWPs. This would make VADoe retrieval more widely applicable, and also use of higher time resolution a priori.

Conclusions:

• please provide a description of VAD technique, and what assumptions used?

Description of the VAD technique and assumptions used are provided in the section 2.1 Traditional VAD Method (VADtrad). We have also added equation used in VAD.

• What equation is used in VAD? What was the vertical air motion at the surface? Zero? Right? Was it correct?

We have added the equation used in VAD. Please see new Eq. (1).

Vertical air motion at the surface is not measured by ground based CDL. Typically CDLs have a blind zone near the instrument that is twice the laser pulse length.

• …..with correlations of 0.998 and 0.999 between the VADtrad and VADoe for u and v, respectively?Is this correct? If they are 100% correlated, it means no difference between 2 methods, then why we need OE technique? This cant be correct because OE provided large uncertainty compared to VAD when compared to radiosonde. Why is that?

Yes, the correlations between VADtrad and VADoe for u and v are 0.998 and 0.999 respectively (see figure 5c and 5f), and there is no difference between the two methods where VADtrad provides the results. What this excellent agreement shows is that there is no disadvantage in using the VADoe compared to VADtrad. The key point of this work is that VADtrad does not provide results at higher levels where the measurement SNR is below a certain threshold. The VADoe provides additional information at levels where VADtrad does not provide any results without any loss of information where VADtrad is valid.

Larger standard deviation for VADoe relative to VADtrad when compared to radiosonde that is seen in Figure 6 is due to the larger number of data points included in the comparison for VADoe. Figure 6f shows number of data points included in the comparison at each level. For example, there are twice as many data points included in the VADoe comparison compared to VADtrad.

• Results represent what? 1 day or 6 months? Also why not provide a few extreme cases?

The results presented in the paper are for the entire 2019 calendar year. This information is already included in the paper. We have also added this information to the abstract and conclusion.

It would be very difficult to assess the performance of VADoe retrieval under extreme cases because the reference sonde profiles themselves might not be representative. Thus, we did not highlight extreme cases. But we also did not exclude any extreme cases. One could also consider all data at low SNR levels are extreme cases.

This method was tested using a yearlong CDL measurements in ARM SGP Central Facility in 2019

• Ln353/354 is this correct? Please provide the conclusions with bullets..

Yes, it is correct. Figure 4-8 shows results at higher altitudes where SNR is too small and radial velocities are not reliable. As mentioned previously, VADoe results at those levels are statistically most likely results based on the noisy observations at those levels, higher quality measurements at lower levels and climatology. Demonstrating the application of the VADoe product is beyond the scope of this paper.