The manuscript presents a 15 year time series from eddy covariance and meteorological data from the Tibetan Alpine Steppe. Due to the remote site and harsh environmental conditions the timeseries in unfortunately full with gaps and many needed maintenances have not been done. Additionally, the fetch was strongly disturbed by construction of buildings. Still the authors challenged these hard data conditions and tried to extract all that was possible out of these data.

First of all, I want to express my deeps respect for the thorough work the co-authors did to create this data set. Obviously, a lot of brains were put together to solve the many problems present in the original data set and time series.

I carefully read the manuscript, downloaded the data files and R-codes. It is a great effort in terms of reproducibility of the data which I very much appreciate. I want to highlight that I truly think these data are worth to being published. My impression is that maybe too much data massaging (with the best intentions I'm sure) was done so it looks like more than it is.

The manuscript is well written and clearly structured, the data are well documented and the provided code is in a good shape. Figures n the manuscript nicely illustrated.

Due to the points below, I recommend a major revision. I'm want to see the manuscript again and I'm very much looking forward to see your comments and suggestions. I'm sure you can tackle all the points and solve them.

Major points:

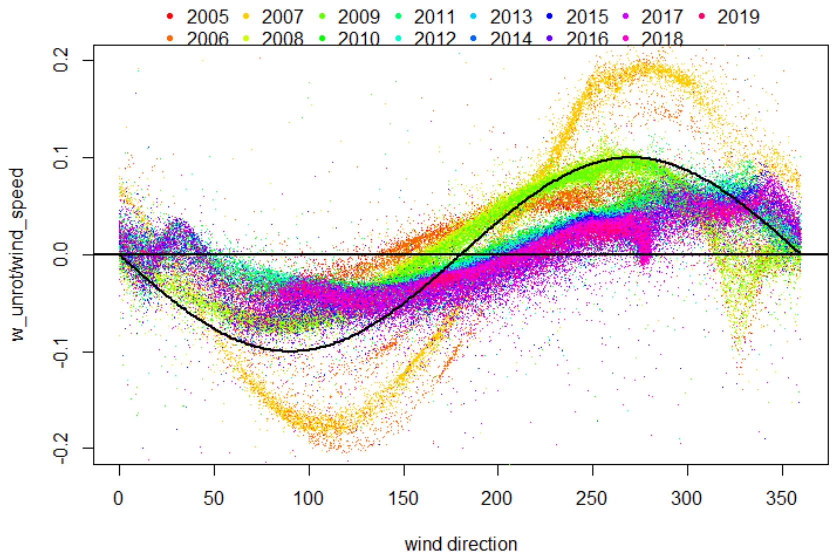
WPL and SSH correction:

After WPL and SSH you still have a diurnal cycle in the CO_2 data even during winter. Obviously, this pattern is not real but this is not at all discussed in the manuscript. This is most likely a WPL correction effect and not a physiological meaningful signal.

The buildings:

The wind disturbance due to the buildings is basically argued away even though the problem still remains. The easiest solution would be the removal of the wind direction 230 – 300 degree. Maybe account for the years and increase the angles based on the years when the respective buildings were constructed. I fully understand that you want to keep as many data as possible but the undisturbed wind field is not given at all if there are massive buildings so close to the tower. Additionally, the footprint calculation have basically no value as the assumptions of a homogeneous terrain and wind flow are strongly violated. It is a shame that the buildings were build there. The consequence is that you can't use the data and that must be faced. Further, I assume the buildings are creating heat and CO_2, greenhouse maybe even contribute to a CO_2 sink. All these influences can't be accounted for that is why it is required to remove these data.

This simple plot gives some nice indication when things changed and how they influenced the wind field of the sonic. A slightly tilted sonic in flat terrain would have a sine shape. Here you can see the obstacles and how they influence the wind field and when things changed. This could let you also think about the size of the sectors for the planar fit methods (just as an idea).



The wind coming from the back of the sonic in a set up as you have (CSAT) it should generally be removed due to flow distortion. That would be something like 350 – 10 degree.

CO2 concentration correction:

Here a novel data correction method is introduced. The correction seems reasonable but it has not been tested, nor have uncertainties and problems been investigated? There are quite some differences between Mauna Loa and e.g. Mt. Waliguan the closest station to NAMORS I saw at https://www.esrl.noaa.gov/gmd/. What would be the differences using one or the other for the flux calculations? Of course, one can run the analysis for 50 other atmospheric CO2 background stations and see how the fluxes change but we are still missing the truth at the site. If such a method is to be used it must be thoroughly evaluated and this has not happened. Even if this correction is valid for the flux calculation it is for sure not valid to sell the resulting CO2 concentrations as the measured concentrations. If the data have not been measured by the instrument the qc-flag must be 2 and not 0. I would also like to address one point which I guess Mr. Fratini can help with or at least validate. His paper from 2014 was done with a LI7200 (and a LI7000 as reference) which is an enclosed instrument using an inlet and in best case a filter that ensures that the inside of the sensor stays clean. As you described it you used a LI7500 open path sensor that besides the changes in the offset and the span is also highly affected by the dirt accumulating on the windows. But this effect cannot be simply calculated back, correct? If I remember correctly the LI7500 puts out the "automatic gain control" (AGC) as an indication how clean/dirty the windows are. And, there are recommendations to which AGC-value data should be used or discarded. If you have any change to get this value out of the raw data t would for sure help you to better QC the data.

The CO2 concentration data in the data file are now following on average Mauna Loa but can we assume this is correct? The half hourly data show a gigantic scatter in mixing ratios between 0 and 600 ppm. Throughout the measuring period there are values of 0 in CO2 concentration. This is interesting because when using the "qc_co2_flux_composite" filter and only select data when "qc_co2_flux_composite" is equal to zero there are many of these 0-concentration data left. In fact, there are 612 data point for which CO2 concentrations are below 300 ppm (including zero-values) or above 600ppm and fluxes seem to be of high quality. This means that the fluxes have been calculated from an average concentration of 0. Does that make any sense? I would say no. You might say who cares about 612 points in a data set of 241178 but it shows that the QC scheme is still including errors.

I'm honestly not convinced by this correction method especially because it was not developed for an instrument where changes in absorption might also arise from dirt on the windows. And because it was not tested and cannot be evaluated with the current data set. I'm sorry for being so negative about this correction but I hope I made my point clear and you share my point of view.

H2O concentration:

The H2O concentrations provided in the data-file are not the once from the LI7500 but the once from the biomet data, i.e. the temperature and relative humidity sensor. This might be okay for a normal data set where no issues are present with drifts, dirty windows, concentration etc. But here I would highly recommend to provide or at least look at the water vapor concentration of the LI7500. When you use EddyPro for processing I think the only way to get the true LI7500 H2O concentrations is when you run the processing without providing the biomet file. The point here is that you can't use the concentration as a quality criteria. You can actually see that by looking at the number of qc_co2_mixing_ratio_composite and qc_h20_mixing_ratio_composite. The number of bad data for qc_co2_mixing_ratio_composite (==2) is 12730 and for qc_h2o_mixing_ratio_composite (==2) is 203. If the concentration of the one is bad usually also the other one is bad. Especially when this is due to dirty windows, precipitation, snow frost, etc… I really encourage you to use the real LI7500 water vapor concentrations to select a criterion to remove bad data and also bad fluxes of h2o.

In principle I would recommend to provide the raw CO2 and H2O concentrations and the corrected once.

Minor comments

Please include the countries to which the southern and western part of the TP belongs. I guess Nepal, Pakistan, India and Bhutan.

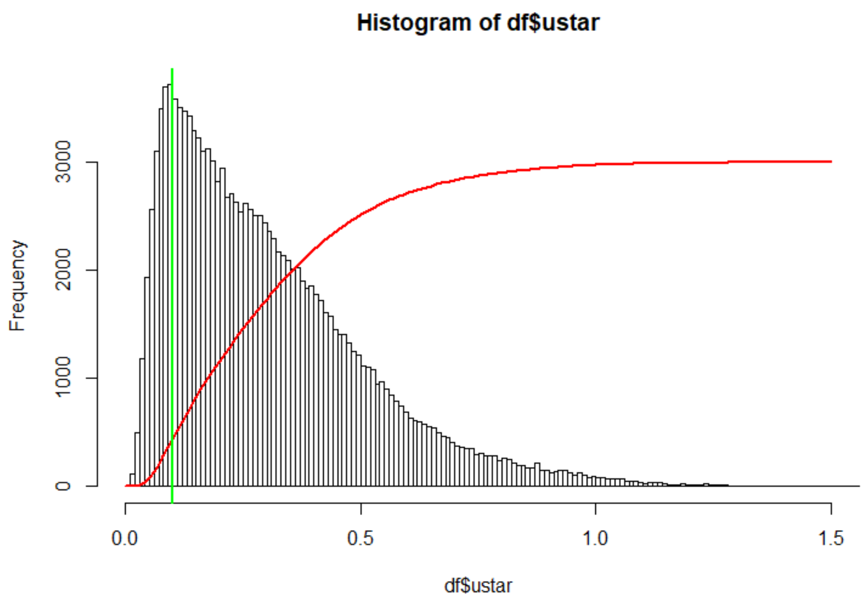
The sine-cosine model is not explained. Why not directly using a spline function or even a moving average? Did you use the flask samples or the continuous? The pattern of the CO2 concentrations is not really a sine or cosine.

I think the u* filtering should be applied. Just because there s wind does not mean there is no relationship. For grassland values around u* values of 0.1 m/s are not so uncommon and that accounts in your data set for 15% of all u* values. The red line in the plot shows the cumulated density function multiplied by 3000 to fit the scale. Green vertical line is at u*== 0.1 m/s.

Formular 7 and 8 you use mu and sigma which are usually the population mean and its standard deviation. I know you took it from the paper of Burba but there are plenty of other variables one can use.

For the uncertainty analysis of the WPL I have only a gut feeling that this is wrong but it would be good if you would get some statisticians input and explain why this is valid to do. In principle each value n the formula has an uncertainty e.g. Ta which propagates in Cp and rho. Sorry for not being more helpful on this one.

But generally, for the overall uncertainty I would rather take the NEE_fsd to calculate the uncertainty of the fluxes. It is including not only the random error but also temporal variability and spatial heterogeneity. There is a paper comparing these uncertainties with each other I think in a savanna (sorry I can't recall the author maybe worth a look).



The sentence in line 425 "The wind direction distributions of wind speed and TKE, as well as the analysis of cumulative footprints suggest that the several buildings which were constructed in close vicinity of the tower do exert some influence on the flow regime while not violating basic EC assumptions. Nevertheless, fluxes originating mainly from the disturbed areas should be excluded from further analyses as they may be compromised by human activities."

You have all indications that the flow was clearly disturbed and you still conclude that the assumptions of eddy covariance are met? How does that go together?

For the data description file some more info on the uncertainties would be great. The differences are not directly clear to the reader.

I'm happy to follow up on this and to keep discussing

Find below my code for reading the data and making some plots. Nothing fancy but only fair to provide.

```
# check data for essd manuscript #


library(lubridate)
```

```
df_fromEPO<-function(FPath){
  df=read.csv(paste(FPath,sep =""), header=FALSE,skip=1,sep="\t",blank.lines.skip=FALSE,
fill=FALSE,strip.white=TRUE);
  cnames <- names(read.csv(FPath, nrows=1, skip=0, sep="\t"))
  names(df)<-cnames

  #convert date format with lubridate package
  df$date=as.character(df$date);
  df$time=as.character(df$time);
  as.character(df$rDate);
  df$rDateChr=paste(df$date,df$time);
  df$rDate=parse_date_time(df$rDateChr,"%y%m%d%H%M",tz="GMT");

  #delete missing values with error code -9999
  df[df==-9999]=NA

  return(df1)
}

FPath="D:/Reviews/2020ESSDLongTermECDataCO2H2OTibetanSteppe/NAMORS_EC_2005-2019.txt"

df=df_fromEPO(FPath)
plot(df$rDate,df$co2_flux,pch='.')
plot(df$rDate[df$NEE_fqc==0],df$NEE_f[df$NEE_fqc==0],pch='.')
points(df$rDate,df$co2_flux,pch='.',col='red')
plot(df$rDate,df$co2_mixing_ratio,pch='.')
plot(df$rDate,df$co2_mixing_ratio,pch='.',ylim=c(0,500))

IDX=which(df$rDate>="2017-05-01" & df$rDate<="2018-01-01")
plot(df$rDate[IDX],df$co2_mixing_ratio[IDX],pch='.',ylim=c(0,600))
```

```
IDX=which(df$rDate>="2017-08-31" & df$rDate<="2017-09-03")

plot(df$rDate[IDX],df$co2_mixing_ratio[IDX],pch=20,ylim=c(0,600))

plot(df$rDate[IDX],df$Ta[IDX],pch=20,ylim=c(273,290))


IDXQC=which(df$qc_co2_mixing_ratio_composite==0)

plot(df$rDate[IDXQC],df$co2_mixing_ratio[IDXQC],pch='.',ylim=c(0,600))


IDXQC=which(df$qc_co2_flux_composite==0)

plot(df$rDate[IDXQC],df$co2_mixing_ratio[IDXQC],pch='.',ylim=c(0,600))


length(which(df$qc_co2_flux_composite==0 & (df$co2_mixing_ratio<=300 |
df$co2_mixing_ratio>=600)))


plot(df$rDate,df$h2o_mixing_ratio,pch='.',ylim=c(0,20))


plot(df$wind_dir,df$w_unrot,pch='.')

abline(h=0,col='red')




plot(df$wind_dir,df$w_unrot/df$wind_speed,pch='.')



#check flow angle based on wind direction and year

df$YYYY=year(df$rDate)

DateUni=unique(df$YYYY)


RainCols=rainbow(n=length(DateUni))

for (i in DateUni){

  print(i)

  if (i == 2005){
```

```
plot(df$wind_dir[df$YYYY==i],(df$w_unrot[df$YYYY==i]/df$wind_speed[df$YYYY==i]),pch='.',col=RainCols[i-2004],ylim=c(-0.2,0.2),ylab='w_unrot/wind_speed',xlab='wind direction')
   }else{
points(df$wind_dir[df$YYYY==i],(df$w_unrot[df$YYYY==i]/df$wind_speed[df$YYYY==i]),pch='.',col=RainCols[i-2004])
   }
}
legend('top',bty='n',xpd=NA,inset=c(0.5,-0.14),horiz=F,ncol = 8,pch=20,col=RainCols,legend=DateUni)
abline(h=0,lwd=2)
SinVals=seq(0,2*pi,length.out = 361)
xvals=seq(0,360,1)
points(xvals,sin(SinVals)*-0.1,type='l',lwd=2)

#check u-star distributon
plot(df$rDate,df$ustar,pch='.')
hist(x = df$ustar,breaks = seq(0,5,0.01),xlim=c(0,1.5))
uStarEcdf=ecdf(x = df$ustar)
points(x=seq(0,1.5,0.01),uStarEcdf(v = seq(0,1.5,0.01))*3000,type='l',col='red',lwd=2)
abline(v=0.1,col='green',lwd=2)

uStarEcdf(0.1)*100 # percent of u*-values <= 0.1 m/s

plot(df$rDate,df$qc_co2_mixing_ratio_composite,pch='.')
points(df$rDate,df$qc_h2o_mixing_ratio_composite,pch='.',col='red')

length(which(df$qc_h2o_mixing_ratio_composite==2))
length(which(df$qc_co2_mixing_ratio_composite==2))
```