Comment: *The authors measured net CO2 fluxes over polygonal tundra in the high arctic on Svalbard. They present an interesting and well written manuscript, but seem to interpret some aspects differently than this reviewer. I think it is important to exchange ideas and opinions between authors and reviewers and that it is positive to have new ideas presented even if not everyone agrees. However, my main points given below is that the language should reflect this in a somewhat clearer way so that the uninitiated reader does not misinterpret the universality of some statements.*
Having said this, I must admit that I learned a lot reading this manuscript and fully support its publication after careful revisions.
Besides an important methodological aspect of how to compute defensible annual flux sums, a key statement of the paper is that all the detailed image analyses starting with pictures taken in 1948 cannot confirm a rapid degradation of this polygonal tundra, but rather support the view that this landscape has been quite stable over the last seven decades.

Reply: We thank Professor Eugster for his thorough review of our manuscript and his helpful feedback. We carefully considered each of his comments, paying special attention to the potentially exaggerated universality of some of our statements.

Main critique

Comment: *1. Your introduction completely misguided me in the wrong direction as your paper starts with the phrase "Carbon-rich Arctic tundra soils are often covered with polygonal ground patterns created by sub-surface ice wedges." – (1) Your paper is not addressing carbon-rich Arctic tundra soils! (2) Absence and presence of polygonal ground patterns is not directly related to carbon-richness of the soil (see e.g. Davis 2001). In fact, the non-orthogonal polygonal tundra patterns are mostly found on homogenous silty or sandy grounds, whereas the carbon-rich surfaces in my experience mostly show orthogonal polygonal patterns that differ from your site.*
In fact, this all does not matter, it is simply a problematic first phrase (the one scientific writing is all about). Please rephrase and start your story in the direction where you actually go. In fact, only on page 12, line 8 my initial suspicion was resolved as you wrote "with its typically shallow organic horizon in the soil".

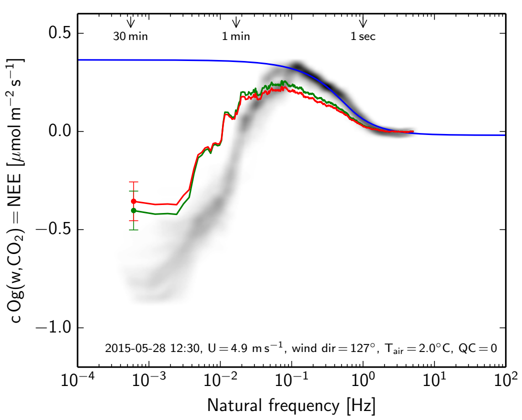
Reply: The top 100 cm of the soil in the EC footprint at the Adventdalen site contain about 30 kg SOC m-2 on average (personal communication with Peter Kuhry). In comparison with other permafrost-underlain well-developed soils, which can contain >100 kg SOC m-2 in the top 100 cm (see e.g. Hugelius et al. 2014), Adventdalen is not extremely carbon-rich. So we understand the confusion created by our first sentence and propose to resolve this misunderstanding by removing "carbon-rich" in the first sentence, so that it would read: "Arctic tundra is often covered with polygonal ground patterns created by sub-surface ice wedges."

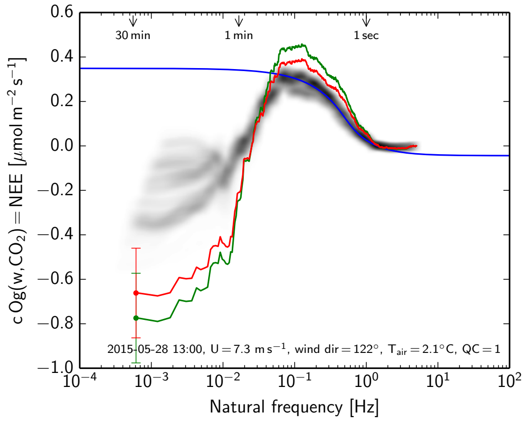
Comment: *2. Abstract. I was confused by the your flux numbers. In principle, a negative sink is a source (page 1, line 8), but as an expert I guessed that you use the negative sign for net uptake and thus a sink of minus something is still a sink (not a source). OK, my recommendation is to put the number in parentheses to avoid the interpretation that it is a source. But the most confusing statement follows in the last line of the abstract: the text in lines 6 to 8 reads like: conventional calculation gives −46 gC m−2, improved ogive optimization gives −82 gC m−2 which is a strengthening of net uptake, but your text on line 14 calls this "a weakening of the CO2 sink" . . . I assume you wanted the reader to read the abstract differently. Please reword and clarify. Maybe also define your sign convention in the abstract.*

Reply: We acknowledge that these formulations could be misinterpreted and need clarification. So we propose to put these numbers in parentheses and define the sign convention at the first occurrence. The sentences 5 and 6 of the abstract would then read: "Non-local (low-frequency) flux contributions were especially pronounced during snowmelt and introduced a large bias of -46 gC m-2 to the annual CO2 budget in conventional methods (minus-sign indicating a higher uptake by the ecosystem). Our improved flux calculations with the ogive optimization method indicated that the site was a strong sink for CO2 in 2015 (-82 gC m-2) and due to differences in light-use efficiency, wetter areas with low-centered polygons sequestered 47% more CO2 than drier areas with flat-centered polygons."

Comment: *3. You strongly vote for the ogive optimization method. I am not perfectly in agreement with your argumentation, though. As I mentioned initially, it is good to lead the discussion, but some more critical assessment of this method is required, which should be reflected in revised wordings at several places.*
Your example in Fig. 2b clearly shows gravity waves seen with the bands of lenticular clouds. Under such conditions it is challenging to filter out the waves (which should not be considered fluxes). In principle, such conditions should fail any stationarity test and one could thus think of other methods to filter out such conditions.

Reply: Firstly, the picture shown in Fig. 2b was not meant to represent the exact same 30-min period used in Fig. 2a, but only approximately the time of year during snowmelt. Using the time stamp of this photo, we now derived the matching EC ogives, which show the same features as the time period used in the original manuscript:





Admittedly, the data quality (QC) is worse and the ogive density map does indicate a degree of non-stationarity. So Prof. Eugster rightly points out that the period shown in Fig. 2b is not ideal for EC flux measurements with any method. However, this gravity wave event seemed to be rather short-lived because no lenticular clouds could be seen on photos taken 3 hours earlier:

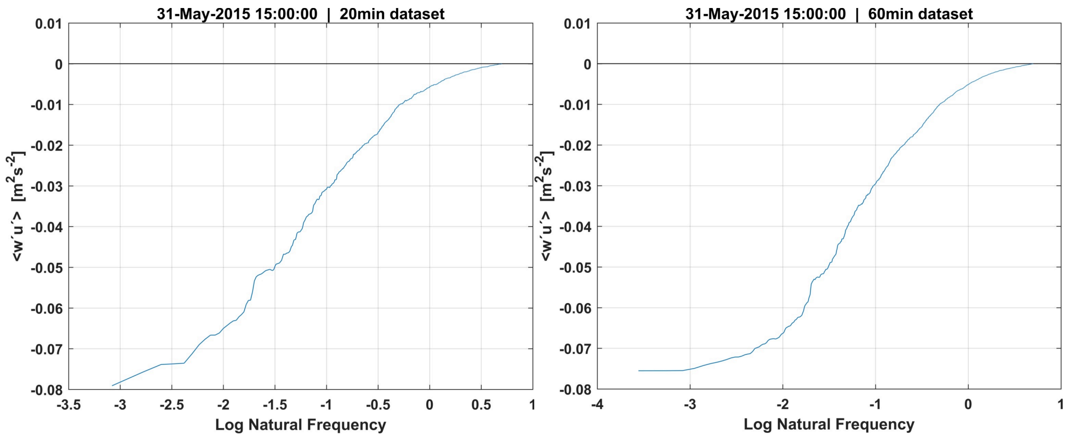
In fact, such lenticular clouds have otherwise never been observed during our site visits so the original picture shown in Fig. 2b is not very representative. To resolve this issue and minimize the potential for misunderstandings, we propose to exchange Fig. 2b in the revised manuscript with this picture, and indicate its time stamp in the figure caption (27 May 2015, 12:00 LT):

Comment: *In the example given in Fig. 2a you basically truncate the turbulence cospectrum at 1/25 Hz, thus arguing that 25 seconds of measurements is enough to determine a half-hourly flux. This is in stark contrast to other concepts such as large eddy simulations where the generally accepted knowledge is used that it is the larger eddies—not the small ones—that are relevant for the turbulent fluxes between the surface and the atmosphere.*

Reply: The shortest dataset evaluated by the ogive optimization software is 10 min and the longest one is 60 min, thus covering a wider range of turbulence scales than conventional fixed-averaging methods. The example in Fig. 2a shows that all relevant flux contributions are carried by turbulence with a scale shorter than 25 sec. While that is interesting, we don't think it is particularly uncommon for EC measurements to have quite small flux contributions below this frequency. However, we definitely don't want to say that 25 sec are in general enough to determine a 30 min flux. To clarify this point, we propose to add the following sentence to the "Results" section describing Fig. 2a: "The ogive optimization model indicates that all relevant flux contributions are carried by turbulence with a scale shorter than about 25 sec in this example (which does, however, not mean that 25 sec are in general enough to determine a 30 min flux)."

Comment: *Long ago I had to deal with a similar issue with my first measurements over lakes (Eugster et al. 2003) and there I used the direction of the momentum flux as a filter criterion. However, some software compute the momentum flux in a way that loses the directional sign so that it is unclear in which direction the momentum flux actually pointed. In principle the momentum flux (averaged over 30 minutes) should point towards the ground surface, but my experience is that in cases as you show in Fig. 2 there might be an upward momentum flux in the high frequencies which would question your interpretation that these high frequencies are better to estimate the local CO2 flux. This only holds if your momentum flux in the high frequency range is clearly downwards. To accept your interpretation I would need to see the cospectrum or ogive of the horizontal windspeed time trace and the vertical windspeed time trace. The horizontal direction must be aligned with the flow so that* $v' = 0$ *and* $u > 0$ *ms−1. Otherwise, if the turbulent momentum flux is in the wrong direction then your argument that the corresponding CO2 flux must be from the local surface would be incorrect.*
Maybe you have also measured a wind profile? If the peak wind speed near the surface is below your eddy covariance (EC) measurement height, then this would be a condition where the momentum flux measured by EC is upwards, not towards the ground. I must admit that filtering with momentum flux direction is very rigorous and in many cases may be overly picky, but I hope I could explain you why I am not really of the same opinion as you (page 13, lines 12–18): if momentum flux is upwards, then your EC system sees the inversion interface between the cold air on the surface and the warm air aloft (which is present if you have clouds as those shown in Fig. 2), not the ground interface. You may overcome this with a more critical rewording; your text on lines 13–14 does not really provide a realistic "speculation".

Reply: The ogive optimization software checks that the momentum flux is negative (towards the ground surface) in the mid-frequency range as one of the quality checks. While this test is described in the original publication of the method, we acknowledge that it would be relevant to briefly mention this test in the "Methods/Data processing" section of our manuscript. We therefore propose to add this sentence: "Ogive optimization furthermore only accepts periods with a negative momentum flux (i.e. directed toward the ground surface) in the mid-frequency range (10^-1.5 Hz)."
We don't have vertical wind profile measurements, but we specifically checked the momentum flux ogives for the period given in Fig. 2a:



These ogives indicate a well-behaved and downwards-directed momentum flux, so we are confident that our arguments hold.
Regarding our speculation of a vertical CO2 layering (page 13, line 13ff), we fail to see why it would be so unrealistic, and since it's clearly indicated as a speculation we cannot really express it more carefully. Still, the possible stability layering suggested by Prof. Eugster could also be at work, so we propose to broaden this speculation a little bit to also mention possible layers of different atmospheric stability, so that the revised sentence would read: "One might speculate that the systematic occurrences of bi-directional fluxes are due to an atmospheric layering where low and high-frequency eddies circle through air masses with different atmospheric stability and CO2 concentrations"

Comment: *4. The limitations you list on page 2, lines 9–10 do not include the factor of self-heating if an open-path instrument is used. Later we see that you used a Licor 7200, but since your introduction is more general I recommend adding a statement here (many use open-path instruments in the Arctic due to power constraints). This is a factor that Baldocchi (2003) was not aware off, thus you should mention this after the citation.*

Reply: OK, we propose to add the following sentence here: "Also, when open-path gas analyzers are used, a bias may be introduced due to surface heating of the instrument itself (Burba et al., 2008)."

Comment: *5. There is confusion about your argument why you focus on 2014/2015 data and less on 2013: on page 3, line 3 you write: "were only recorded as wet*

molar densities and without the cell pressure necessary to convert them to dry mixing ratios". Is this a typo or did I misunderstand this statement? It is the H2O density measurement that is needed to convert from wet to dry mixing ratios. Temperature and cell pressure are only necessary to convert from densities to mixing ratios. A similar confusion is found on page 8, lines 1–2: "since they were wet rather than dry molar densities". Before you argued because of the mixing ratios, here one wonders why the ogive method should not work with wet molar densities if it would work with dry molar densities?
Please clarify these things. In principle you could use the Webb-Pearman-Leuning correction for your 2013 fluxes. Why did you not use this method to better profit from your interesting dataset?

Reply: We acknowledge that these statements are somewhat inconsistent and misleading. The problem with the 2013 data is that a sample-by-sample conversion of recorded densities to mixing ratios would require the cell pressure, which hasn't been recorded. And while EddyPro can use the WPL correction to calculate corrected fluxes, this feature is not implemented in the current version of the ogive optimization software. We propose to clarify this issue in the "Methods/Data processing" section, writing: "We mainly focused on data collected between September 2014 and December 2015, when data quality and coverage was highest. CO2 concentrations collected in 2013 were only recorded as molar densities and without the cell pressure necessary for a sample-by-sample conversion to mixing ratios according to the Webb-Pearman-Leuning correction proposed by Sahlee et al. (2008), which is currently the only option implemented in the ogive optimization software. Hence, we only report 2013's fluxes from EddyPro as supplementary support for our findings."

The misleading sentence on page 3 line 23 would be removed. The sentence on page 8 lines 1-2 could then be simplified, reading:
"In 2013, EddyPro calculations yielded a smaller total annual CO2 balance of -79 gC m-2 (see Fig. S2d in the supporting information), whereas ogive optimization fluxes could not be calculated from 2013's raw CO2 measurements (cf. section 2.3)."

<u>Detailed technical remarks</u>

Comment: *1/10: use K instead of ◦C for temperature differences*
Reply: OK

Comment: *3/1: add "flux" in EC CO2 flux measurements*
Reply: OK

Comment: *4/21: use "s" not "sec"*
Reply: OK, we changed this throughout the entire manuscript.

Comment: *5/10: correlations between time series depend on the measurement interval; an r > 0.9 definitely does not hold for 10 Hz data, but may be seen with*

monthly data. Either specify which aggregation level you talk about, or simply remove this statement in parentheses. Giving the distance is an objective information that should be sufficient.
Reply: OK, we removed the statement in parentheses.

Comment: *6/11–12: wording reflects some inconsistency in your statistical testing. I assume you used a t-test, but if you write "on average 10 cm larger thaw depth" then this implies a one-sided t-tests (testing for "greater than"). The wording on the line below ("this difference is not statistically significant") however is the wording for a two-sided test. Please rectify.*
Reply: OK, we propose to clarify that we used a two-sided t-test by changing this part to: "The thaw depth at the centers of the polygons around the EC tower was 66cm +/-9cm (mean +/-standard deviation, sample size N=30) by the end of August. Based on the polygons in the 50% EC footprint, the drier ESE fetch area featured a thaw depth of 69 cm +/-8 cm (N=4) while the wetter NW featured 79 cm +/-4 cm (N=4), which is not a statistically significant difference (p=0.10)."

Comment: *7/9: do not specify "$p < 10{-}12$" since statistical models are not supposed to be accurate down to $p < 10{-}12$. Normally for low values it is sufficient to indicate something on the order of $p < 0.0001$ or so.*
Reply: OK

Comment: *8/13: use K instead of ◦C for temperature differences*
Reply: OK

<u>References:</u>

Hugelius, Gustaf, et al. "Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps." Biogeosciences 11.23 (2014): 6573-6593.

Burba, George G., et al. "Addressing the influence of instrument surface heat exchange on the measurements of CO2 flux from open‐path gas analyzers." *Global Change Biology* 14.8 (2008): 1854-1876.