Reply to Anonymous Referee #2 – RC1

We thank the reviewer for their constructive review of the manuscript. We have made all of the specific recommended changes. Please see below for our response to each of the comments.

General Comments

** The manuscript is extremely well written. * This paper addresses an important need in the community with a practical and well-described method for estimating emissions rapidly and on a broad scale. * While I understand that there was not an opportunity to benchmark the estimates against other methods of emissions estimation, the lack of validation remains a significant weakness. I nevertheless recommend publication, but this caveat should be recognized at key steps in the analysis. * The largest omission from the paper is the lack of any uncertainty estimate for the emissions from the region. Some effort should be made to rectify this in the final paper. * I don't understant the use of linear regressions (with variable slope and offset) for the detection rate estimates. Justification of why this analysis should be used over the simple calculation of rate = emitting sources / total sources should be provided, or the authors should revert to the simpler analysis.*

We appreciate the reviewer's general comments. The reviewer's concerns surrounding both uncertainty estimates and the linear regression plots are dealt with more explicitly in the Specific Comments section. We have addressed these comments in detail below.

Specific Comments

- P1 L17: emissions estimates for the Montney development does not have an uncertainty estimate. It is difficult to interpret the emission results without an uncertainty associated with it.

In our study we have made a minimum emissions estimate by combining the minimum detection limit of our applied method with our calculated emission frequencies for the infrastructure in the survey area. We expect that the total CH4 emission volume for the area is higher than our reported estimate.

A regulator-sponsored FLIR study was released at the same time we submitted our manuscript to ACP (GreenPath (2017)). The study was independent of ours, but took place in the Alberta portion of the Montney formation (the same play that is being developed in the field area of our study). The study by GreenPath Energy reported almost identical emission frequencies and emission volumes as we calculated for our field area. The results of our study reinforce the emission patterns of the GreenPath study across a larger sample size.

We have added the following text to section *3.4 Methane Emission Inventory Estimate* of our manuscript to address how this newly released study validates our method of volume estimation.

"Our emission frequency calculation for Active wells (0.47) was very similar to the emission frequency of 0.53 that was recently calculated in the Alberta Montney near Grande Prairie (GreenPath, 2017). Our method of calculating emission frequencies is corroborated by this recent FLIR study in the Alberta Montney, which increased our confidence in using emission frequency calculations to estimate a minimum CH4 inventory for the development."

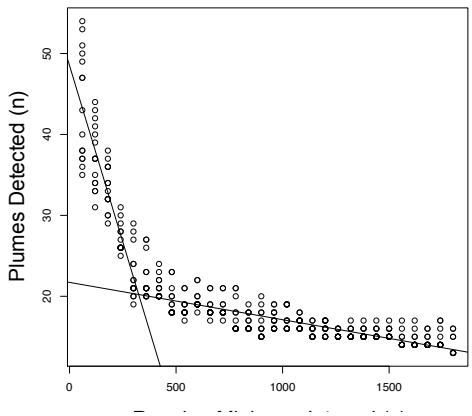
- P5 L1 - 10: The authors state that they are using excursions in the eCO2:eCH4 ratio (<150) as indications of natural gas emissions. However, I would imagine that other sources of CO2 could add noise to this ratio (especially since there are other vehicles that contribute to excess CO2). Figure 3 further indicates this issue. A fairly obvious alternative would be to use the same RMRI algorithm and use eCH4 > threshold as a criterion for when emissions are detected. It would be helpful if the authors could provide some more justification why the ratio eCO2:eCH4 is a better metric than simply eCH4.

The method of using excess ratios (particularly eCO2:eCH4) for plume source attribution in an upstream oil and gas environment is described in Hurry et al. (2016). We have added the following text to the manuscript in section *2.2 Identification of Natural Gas Emissions* to clarify that a detailed explanation of the method can be found in this paper.

"This eCO_2:eCH_4 approach has proven to be a useful fingerprinting tool in oil and gas environments because a single ratio value can help elucidate the presence of multiple emission source types. In this study, we follow a procedure similar to Hurry et al. (2016), and a detailed explanation of the method is described in that paper."

- P5 L10-12: "Our optimal RMRI was taken to be the point at which anomalies were maximized, but also where we avoided the rapid noise-associated increase associated with extremely short RMRIs": in practice, how was this optimization performed? It appears to be a subjective choice. Is this true? It would be preferable if the choice was made objectively using quantitative criteria; it would also be preferable to have the same algorithm be used for all surveys.

We did not choose the RMRI value for each survey subjectively. The optimization was performed with an algorithm that was applied to all surveys individually. We have added the following figure and associated text to the paper to clarify the quantitative process we used to determine the RMRI for each survey. Please see



the figure, caption, and revised text below.

"Figure 2: Example of a regression plot that demonstrates the optimization process we used to calculate an RMRI for each survey. The RMRI for each survey was chosen where the two linear regression lines intersect."

- P5 L 18-19: "Combustion values were also recorded along the routes when eCO2:eCH4 exceeded 1000, and were related to vehicle tail-pipe emissions and industry". What does 'combustion values' mean?

This sentence has been re-worded in the manuscript to better explain how we filtered out emissions related to combustion.

"We also detected occurrences of combustion emissions along our survey routes, and we differentiated them from other emission sources by filtering out all values where eCO2:eCH4 > 1000. Combustion-related emission sources include vehicle tailpipe emissions and industry (ex. power generation)."

- P5 L24-25: "because ratios are more conservative than concentrations in valleys and other areas where pooling of gases is common, and fewer false positives are likely" - doesn't the RMRI algorithm take care of slowly varying concentrations of CH4? It would be good to demonstrate clearly why eCO2:eCH4 is an advantage; if one were to reproduce this method at a larger scale, it would be good to provide clear understanding of why the CO2 concentration is required.

It is possible that eCH4 would have been sufficient and may well have given similar results with few false positives. However, the excess ratio technique is established to be more useful in areas of complex upstream geochemistry to partition a number of emission source types (please see answer to comment *P5 L1-10* for explanation and reference to Hurry et al. (2016)). We did not resolve multiple peaks within the excess ratio density plots (Fig. 4 in the revised manuscript), which we would expect to see if there were multiple source types throughout our surveys. The excess ratio technique provided confidence that the source types are related to the infrastructure to which we were proximal during our surveys.

- P5 L28-30: why was the value 150 selected? What is the effect of this selection on, for example, the emissions estimate, the number of emitters detected, the detection limit, etc. Similarly, what is the effective limit on detection of the system, in units of eCO2:eCH4?

The value of 150 was selected based on peaks in the eCO2:eCH4 density distributions (Fig. 3). Although there is not a clear peak on each graph, many of the routes showed leveling out of the "natural" peak (~215) near 150-175. We chose 150 to be conservative, and it acts similarly to setting a methane excess threshold. Since our survey routes were focused in areas of dense oil and gas development, the elevated density of emissions with eCO2:eCH4 values <150 were interpreted to be from oil and gas related sources. The value of 150 was also considered to be conservative enough to exclude diluted CH4 from natural

sources. Also, the exact ratio threshold often does not affect the number of plumes detected, but rather the width of the plume (duration while surveying), which is not pertinent to this study.

- P6 L7: are there any estimates of cattle emission in this region that could be included?
We were unable to retrieve this information for the fieldwork area and dates. However, our use of a 50% emission persistence threshold for identifying emitters likely rules out the possibility that we included emissions from livestock in our calculations.

- P7 L10: how is this probability defined? Per mile? Per second? For the whole route? This isn't clear.
This probability was defined for the whole route. We have now clarified in the manuscript that we calculated the probability of false plume detection for the entire Control Route.

- P7 L1-5: The kernel density plots do not have a clear knee below 215. Where is 150 on this plot? why was 150 selelected, and not 125 or 175, for example?
Please see answer to comment *P5 L28-30.*

P7 L16-20 and Fig 4. Was wind direction used to evaluate whether a plume should have been detected from the green well pads? Are the databases of well locations up to date? Was there an effort to corroborate locations with on-ground survey or satellite imagery?
The source location databases were up to date at the time we retrieved them (July, 2015). Locations of the majority of sources in the database near our surveys were verified during the on-ground survey campaigns. A section has been added to the manuscript about the uncertainty in infrastructure inventory in response to a comment from Anonymous Referee #1 *p9 1.7-8*. We have also reworded the caption of Figure 4 (now Fig. 5 in revised manuscript) for clarity.

"Figure 5: A subset of infrastructure locations that we surveyed during our field campaign in attributed form. This figure serves as an example of how we attributed wells and processing facilities to on-road plumes. Grey lines represent the survey route. In this case 31 wells or facilities were surveyed, and we used our attribution technique, which accounts for wind direction and distance to source, to determine whether or not these wells and processing facilities were probable emission sources.

P7 L32: "it had to have > 50% emission persistence." Similarly, did persistence include wind direction? In other words, did persistence include whether the potential source was upwind of the vehicle at the moments the vehicle passed by?
Yes, our calculation of emission persistence included only the sources we had sampled. And in order for a source to be considered sampled, at least three successive datapoints had to be downwind and within 500 m of the infrastructure in question. We have clarified this in the following section of the manuscript:

"In this study, emission persistence is defined as the number of surveys on which

a CH4-enriched plume was attributed to a piece of infrastructure, divided by the number of times we surveyed that infrastructure in the downwind direction. A plume was only attributed to a piece of infrastructure if we recorded three or more successive CH4-enriched measurements within 500 m in the downwind direction of the source. And in order for a piece of infrastructure to be classified as an emission source, it had to have > 50% emission persistence."

P11 L8: "concentrations will decrease exponentially away from a release source": the dependence on distance is not exponential. Gaussian plume models predict something like ~1/d to 1/dˆ2, for example.

Thank you for pointing this out. We have removed "exponentially" from this sentence in the revised manuscript.

P11 L11-18: Wouldn't nearby plumes (with faster time signatures) be diluted more than more distant plumes? And wouldn't the peak area (in time) be conserved for short pulses? This is a very big adjustment of the concentrations and therefore the emissions. Did you use peak height or peak area to estimate emissions?

Gaussian plume analysis depends on plume centerline concentrations, not widths.

P12 L9: Rather than using the MDL as the average estimate of emissions, wouldn't it be possible to actually craft an estimate of emissions given the plume dispersion model and estimated distances?

The process of calculating emission rates using Gaussian plume dispersion for each individual datapoint is computationally intensive because of the amount of measurements collected. The technique of applying volume estimates to mobile survey data was not developed at the time we processed these data. Our research group is currently developing a similar technique of volume estimation, but this will be part of a separate study and ground validation is still required.

p12 L28: It is important to include some uncertainty estimates for the emissions estimate. Even a simple low and high estimate of error is better than nothing. For example, the estimates of errors on the slope of the active wells could be used to bound the estimate.

Please see our answer to comment *P1 L17* from Anonymous Referee #2 for an explanation of added text about method validation. The linear regression plots have also been changed to bar graphs in response to comment on Fig. 5, 6, and 7.

p14 L9: It's not clear how this method identifies super emitters, since the authors do not present a clear method for quantifying emissions and identifying the largest emitters. How does this method help identify the largest emitters?

This section of the manuscript is referring to the benefits of using an on-ground detection method that surveys a large fraction of infrastructure throughout the development. In comparison to emission factor inventory estimates, we are more likely to have captured emissions from super-emitters. We have added the

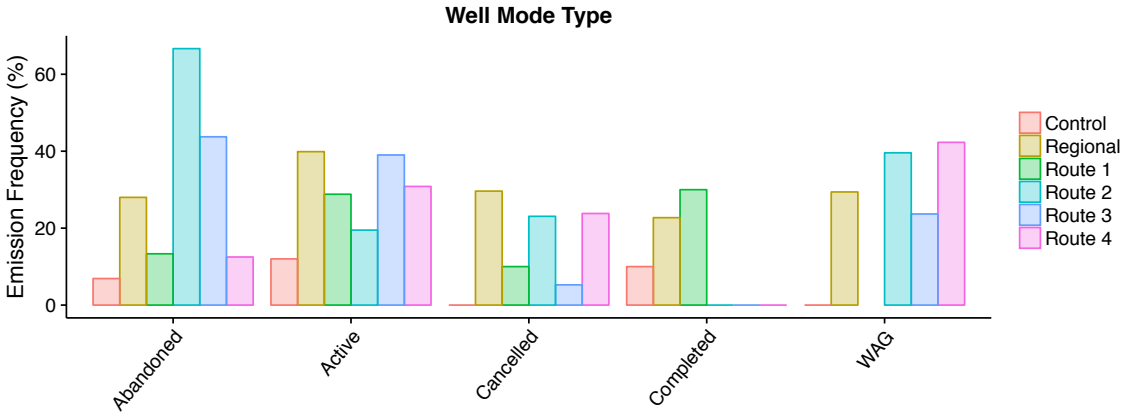
following text to section *3.1 Measured Gas Signatures* to address our results relative to what would be expected from super-emitting sites:

"We did not see any CH4-rich plumes that would be characteristic of a super-emitter. This is evident by the fact that the maximum raw CH4 value we recorded was low (8.148 ppm). These low emission magnitudes are inline with results from GreenPath Energy (2017), which used FLIR cameras to assess emission sources in the Alberta portion of the Montney formation."

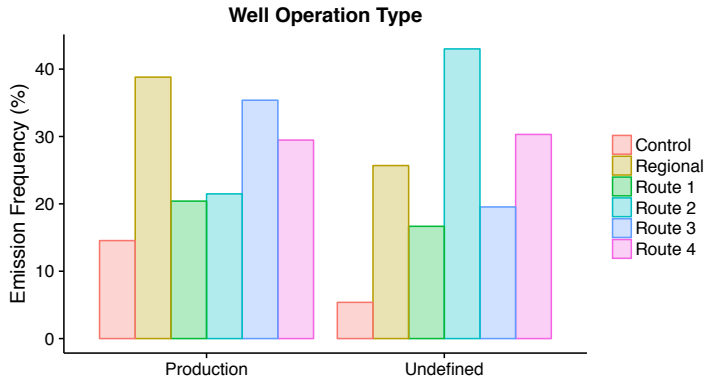
Fig 5: In some panels (e.g., the top panels), the regression lines do not pass through zero. This doesn't make any physical sense. Why should there be a threshold for number of wells surveyed below which no emissions should occur? Why would there be no emissions for surveys with fewer than 60 wells surveyed? I don't understand the rationale for a linear regression. Why not simply ratio the total number of sites with emissions / total number of sites surveyed across all surveys for each category? This would make more intuitive sense. Alternatively, the linear regressions could be forced through zero, which would be similar.

Fig 6 and 7: similar comments to above for Fig. 5.

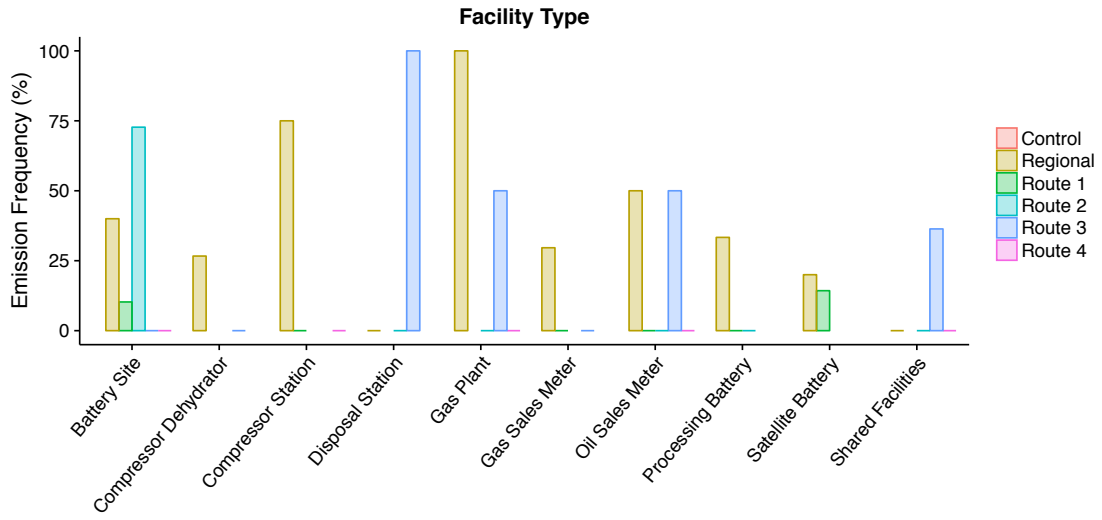
We agree and have changed the linear regression plots to bar graphs which show the percentage of infrastructure emitting for each source-type. Please see the graphs and captions below. We have also made minor changes to the manuscript text accordingly.



"Figure 6: Emission frequencies for each well mode type for all surveyed infrastructure on each route. These emission frequencies were considered in our total emissions inventory calculations."



"Figure 7: Emission frequencies for each well operation type for all surveyed infrastructure on each route. Certain operation types for which we did not have a representative sample are not included (such as Injection, Disposal, and Observation wells)."



"Figure 8: Emission frequencies for each facility type for all surveyed infrastructure on each route. These emission frequencies were considered in our total emission inventory calculations."

Fig 8: Is the occurrance structure due to the fact that some areas were surveyed only three times, which did not allow for a 50% persistence point, for example? This set of plots is a bit confusing.

(This is now Figure 9 in the revised manuscript). In this figure, "Occurrence" (y-axis) refers to the number of pieces of infrastructure emitting at each level of persistence (x-axis). The y-axis has been re-named to "Unique Wells/Facilities (n)" for simplicity. Below is the edited caption.

“Figure 9: The cumulative number of unique wells/facilities versus emission persistence (%) across all 30 mobile surveys. Persistence refers to the repeated tagging of a piece of infrastructure as a possible emission source based on the method of plume attribution we applied in this study.”

Fig 9: what do negative mean eCH4 excursions mean (gray bars of lower panels)?
(This is now Figure 10 in the revised manuscript). We have removed the grey error bars from this figure. Below is the edited caption.

“Figure 10: Effect of infrastructure age and operator size on detected emissions. The size of the dots represents the number of samples taken. Red dots are those recorded at the 100% persistence level, green dots are at 50% persistence.”

Fig 10: could you add in the survey paths on this plot for reference?
(This is now Figure 11 in the revised manuscript). We have chosen not to add the survey routes because the size of the dots already represents the sample size in each area.

Typographical error and other small comments

P1 L13-15: "older infrastructure tended to emit more often (per unit) with comparable severity in terms of measured excess concentrations on-road." - unclear; per unit? what is a unit? reword for clarity, please.
“Unit” was referring to each individual piece of infrastructure. This has been reworded in the manuscript for clarity.

“Multiple sites that pre-date the recent unconventional Montney development were found to be emitting, and we observed that the majority of these older wells were associated with emissions on all survey repeats.”